Resiliency: A Systematic Review of Adult Characterological Measures of Resilience and Reliability and Validity Generalization Studies of the Brief Resilience Scale

Teresa A. Allan

Thesis submitted to the University of Ottawa
in partial Fulfillment of the requirements for the
Doctorate in Philosophy degree in Clinical Psychology

School of Psychology
Faculty of Social Sciences
University of Ottawa

Dedication

To the resilient ones

Abstract

This dissertation consists of two projects. Project 1 is comprised of two systematic review studies in which self-report measures of adult characterological resilience were identified and summarized. In Study 1, 51 personality-based instruments across five personality-based resilience constructs (ego-strength, grit, hardiness, mental toughness, and resilience) were identified. Information for each measure inclusive of measure title, construct, item count, factors, response scale and items, scoring method, and score range was summarized. In Study 2, 1,322 articles were identified that described 1,193 studies and 1,351 participant samples who completed between one and four of these resilience measures. Measure use frequency and contextual use information related to study and sample attributes are summarized. Project 2 is a series of meta-analyses that were conducted to investigate the psychometric properties of scores on the Brief Resilience Scale (BRS) reported in 149 studies of adult participants. Reliability generalization methods are used to summarize 86 observations of internal consistency (mean Cronbach's $\alpha = .86$) and validity generalization methods are used to summarize eight observations of convergent validity ($M_r = .61$). Concurrent validity is also investigated via validity generalization methods for the most frequently observed mental health correlates of BRS scores. For each set of mental health measures, an estimated summary effect was calculated (anxiety, $M_r = -.47$; depression, $M_r = -.49$; optimism, $M_r = .44$; social support, $M_r = .24$; and stress, $M_r = -.53$). Of the significant moderators conducted for concurrent validity estimates, the largest was sample category (accounting for 100% of the variance in the observed correlations between the BRS and measures of social support). Geographic location and participant age, accounted for 55%, and 24%, of the variation in the observed correlations between the BRS and significantly moderated the relation between the BRS and perceived stress. Discussion for both

projects include how the obtained relate to resilience theory, and the potential broader implications of these findings. Also based upon these reviews and meta-analyses, the challenges of resilience measure use are discussed and observations/recommendations are provided.

Keywords:

Resilience, Hardiness, Ego Strength, Mental Toughness, Grit, Brief Resilience Scale, BRS, reliability, validity

Statement of Co-Authorship

The two manuscripts included in this dissertation were prepared in collaboration with my dissertation supervisor. I was primary author and Dr. John Hunsley was the second author for the first manuscript, “Resilience Keeps Bouncing Back: A Systematic Review of Characterological Resilience Measures in Adults”, and the second manuscript titled “The Brief Resilience Scale: Meta-Analyses of Reliability, Convergent Validity, and Concurrent Validity.” For both, the third author, William Lawrence provided inter-rater reliability coding and made contributions to the R code for some of the figures. As the primary author on all manuscripts, I consulted with Dr. Hunsley to develop the conceptualization of the research questions and methods, and carried out the data collection, coding, statistical analyses, and initial preparation of the manuscripts. Dr. Hunsley also provided guidance and assistance in all aspects of the project, especially in the refinement of the research questions, methods, and editing of the manuscripts.

Acknowledgments

I would like to express gratitude and appreciation to Dr. John Hunsley, my thesis supervisor, who provided me with invaluable guidance, feedback, and expertise throughout the development and completion of the dissertation projects to follow. In addition, across the course of my degree, John provided an 'always open door' and a safe, non-judgemental space to ask for his advice, input, and feedback. Thank you - I am in your debt.

I would also like to thank my committee members: Drs. Andrea Ashbaugh, Patrick Gaudreau, Betsy Kristjansson, and my external examiner Dr. Donald Saklofske, who provided a combination of insightful comments, challenging questions, candid feedback, and thoughtful suggestions throughout this process.

Thank you to my "data wranglers": William Lawrence, BSc., and Dr. Eric Kuebler for their hours of coding data, data management input, and tech support. Further, expressions of gratitude to my undergraduate thesis advisor, Dr. Robert Cribbie, at York University, who provided an invaluable introduction to both systematic review procedures and programming in R. Rob continues to allow me to pick his brain on 'all things R and statistical analyses'. He has been a go-to resource who kindly shares his expertise, and I appreciate this very much. Further thank you's go to resilience measure developers: Dr. Bruce Smith, Dr. Paul Bartone, Dr. Jonathan Davidson, and Dr. Gail Wagnild for their correspondence; and to the numerous researchers who responded to my requests for additional statistics, data, and information.

Finally, I would like to thank all of the individuals who comprise "my little world". Friends, allies, colleagues, and colleagues who have become friends: Your kindness and support cannot be measured. Thank you.

List of Tables and Figures

Project 1

Project 2

Table of Contents

General Introduction

Resilience, "bouncing back," returning to a homeostatic state, or failing to be perturbed from that state when presented with adversity and challenge is not a new idea. Resilience theory describes this ability to bounce back using overlapping constructs (ego-strength, hardiness, mental toughness, grit, and resilience) and has proposed differing states of permanence for resilience (static vs. dynamic, change slowly over years vs. change in response to an event vs. can be changed by participating in 'weekender' resilience workshop). Both general resilience (resilience that is pervasive across contexts) and context-specific resilience (resilience occurring in specific aspects of the individual's life or in specific settings) has been the focus of research. Inconsistent terminology use is common in the resilience literature, both in terms of differing definitions of resilience and in terms of research design, with some researchers using the term resilience to describe a predictor variable and others using the term resilient to describe a group of asymptomatic participants or to define a specific desired outcome. Even if only focusing on general, characterological resilience in a specific age group (adults) there are many resilience measures available, and these vary substantially in the amount of evidence available related to observations of reliability and validity specific to the samples of participants being measured. Given the inconsistencies and variation, it is no small task to select an appropriate evidence-based measurement instrument specific to examining general resilience in adults that has adequate evidence of prior reliability and validity.

Overview of the Dissertation

The aim of this dissertation is to explore and describe self-report instruments that are available to measure resilience in adults from the characterological perspective; it is comprised of two projects.

Project 1: “Resilience Keeps Bouncing Back: A Systematic Review of Characterological Resilience Measures in Adults” involves two systematic review studies. Study 1 in this project is a systematic review that identified 51 personality-based measurement instruments across five resilience constructs (ego-strength, hardiness, mental toughness, grit, and resilience). Information for each measure inclusive of measure title, construct, item count, factors, response scale and items, scoring method, and score range was summarized. Study 2 in this project is a large-scale systematic review that included 1,322 articles with samples of participants who completed at least one resilience measure identified in Study 1. Measure use frequency and contextual use information related to study and sample attributes were summarized. Based upon these reviews, I discuss the challenges of resilience measure use, and share observations and recommendations related to being aware of the limitations and/or overcoming challenges inherent in the measurement of resilience and the selecting measures for resilience research.

Project 2: “The Brief Resilience Scale: Meta-Analyses of Reliability, Convergent Validity, and Concurrent Validity”. This project consists of a series of meta-analyses focused on one of the most frequently used unidimensional resilience measures (Brief Resilience Scale [BRS]; Smith, Dalen, Wiggins, Tooley, Christopher, & Bernard, 2008). The second project includes the BRS data from the journal articles that were identified in Project 1 (search date range 1953-2016) and studies published since that time (date range 2016-2019), including published theses and dissertations from the date of the publication of the BRS in 2008 to 2019. Reliability generalization methods are used to summarize observations of internal consistency. Validity generalization methods are used to summarize observations of convergent validity. Concurrent validity is also investigated via validity generalization methods for the most frequently observed mental health correlates of BRS scores (anxiety, depression, optimism,

social support, and stress). For each set of mental health measures, an estimated summary effect is calculated and potential moderators are investigated. I discuss what can and cannot be reasonably inferred from these data, the challenges of resilience measure use, and share observations and recommendations related to how study design and missing information impacts the interpretation of these data. Also discussed is how the results obtained relate to the BRS in these specific contexts and relate to resilience theory, as well as the broader implications of these findings.

The overall utility of these projects is that knowledge gaps in the resilience literature are highlighted and addressed. The first gap addressed is there had been no recent, psychometrically focused, evidence-based systematic review that includes resilience and synonymous constructs that also summarized the structure of available resilience instruments, nor was there a systematic review that summarized the frequency of resilience instrument application. The psychometric information included the manuscript for Project 1 may aid resilience researchers in the selection of measures that are appropriate to the populations they wish to study. Project 2 provides a unique contribution to the resilience literature in that there have been no meta-analyses of the Brief Resilience Scale, which was identified as one of the top five most-applied resilience measures. Project 2 uses meta-analytic reliability generalization methods to examine administrations of the BRS in detail and provides summaries of available reliability evidence (internal consistency estimates) and describes missing reliability evidence (test-retest reliability information). Project 2 also uses validity generalization methods to examine convergent validity (strength of association between the BRS and other measures of general resilience) as well as concurrent validity evidence for mental health constructs that were correlated with the BRS (anxiety, depression, optimism, social support, and stress). Potential moderators of these types of

reliability and validity evidence are examined, and the results of each are discussed.

An additional utility that emerged through the process of developing these systematic and meta-analytic summaries was that further hypotheses related to resilience and resilience theory came to light as well as questions related to existing hypotheses, both of which as processes of the scientific method work toward a better evidence-based understanding of this construct. An additional contribution that this research provides is a discussion related to how studies of resilience and current reporting practices, in terms of the impact of the omission of essential statistical and methodological information (e.g., effect sizes, details on study samples) and of some of the practices observed in published studies affect the reliability and validity of resilience assessments (e.g., selection of only certain test items, alteration of items, and the assumption that the reliabilities reported for norming populations apply to the population of focus in each study).

History of Resilience: From Metals to Characterological Mettle

Resilience, "bouncing back," returning to a homeostatic state, or failing to be perturbed from that state when presented with adversity and challenge is not a new idea. Derivatives of the Latin terms *resilio, reslui, resilire,* and *resiliens* came into use in the English language in the mid-1620s (Resilience, 2018; Resilio, 2018). At that time, resilience was used to describe the propensity of an object to return to its original shape after an impact or to resist being disrupted altogether. Present day metallurgists and physicists use the term resilience interchangeably with the elasticity modulus of an alloy or composite material to describe the stress and force required before that material will break, bend, or experience permanent deformation from its original state (Young & Budynas, 2002). Essentially, examining the property of resilience is to ask: *how much can this object endure before it breaks?* Objects that are resilient withstand more stress than do objects that are not resilient.

In physics, there is an agreed upon definition for resilience as a property. Elements or compounds that are higher in resilience have a higher resistance to disruption or deformation (change) and recoil faster (bounce back or return to homeostasis) following the application of stress. Metallurgists examine alloys and other compounds to determine how well they handle stressors, much like psychologists examine the psychological and situational stressors that humans encounter and endure. The categories of stressors used in materials analysis bears an uncanny resemblance to those in used in psychology: short-time static loadings (acute, temporary, or situational stressors), long-time static loadings (chronic stressors), repeated loadings (as it sounds, repetitive stress), and dynamic loadings (stressors that come and go, are intermittent, or may be unpredictable) (Young & Budynas, 2002). Although materials science is more concrete than psychology and sets of mathematical formulae may be applied to calculate the precise resilience of an object or alloy, needless to say the "properties" of human beings are less easily quantified. For human properties such as latent variables that cannot be measured directly, researchers in the behavioural sciences must "construct" a definition or description of key elements to measure and represent the abstraction to be measured. Thus, as the term implies, *constructs* are constructed representations. Constructs may represent transient states (e.g., excitement, happiness, anxiety, or depression) or characteristics that are hypothesized to be more permanent or more pervasive across situations (e.g., openness, conscientiousness, extraversion, agreeableness, neuroticism). The closest analogue to a fixed property in the field of psychology is the personality trait (Allport, 1921; Galton, 1884; Tupes & Cristal, 1961).

Beginning in the Middle Ages with Hippocrates' four temperaments (sanguine, phlegm, cholera, and melancholy) tied to the four humours (blood, mucus, black bile, and yellow bile), many constructs have been proposed to describe, categorize, and explain the human personality.

Ranging in number and complexity from the relatively simplistic psychoanalytic tripartite model with three key interactive elements (the ego, id, and superego) to the Massachusetts Institute of Technology's online listing of the 638 primary personality traits, and spanning several centuries, it is certain that at least some redundancy, definitional overlap, and simultaneous construct development has resulted in the existence of differentially named constructs that appear to describe or measure approximately the same thing (Costa & McRae, 1992; Freud, 1927; James, 1907; Massachusetts Institute of Technology, 2018).

The earliest mention of a resilience-like construct occurred early in the last century. William James, in his classic work *The Energies of Men*, stated "Can any one of us refine upon the conceptions of mental work and mental energy, so as later to be able to throw some definitely analytic light on what we mean by 'having a more elastic moral tone' or by 'using higher levels of power and will?" (James, 1907, p. 331). Although there were no resilience constructs, nor any psychometric instruments to measure resilience in 1907, it was identified over a century ago that in instances of war, and instances of disaster that certain human beings persevered in ways they themselves never imagined possible. Anecdotal evidence, war stories, and James' own observations of unusual psychological strength demonstrated by survivors of the massive earthquake and subsequent fires of San Francisco in 1906 inspired him to put out a call for research to identify this 'thing', whatever it may be, that enabled certain individuals to persevere in a wide range of difficult and sometimes horrific circumstances, because it was of great value and was worthy investigation and measurement (James, 1907).

Human mettle – a "vigor and strength of spirit or temperament" (Mettle, 2018, p. 1) when facing adversity or challenge – was deemed a personality trait in the 1920's and was quantitatively measured as ego-strength first in 1953 (Barron, 1953). Interest in the measurement

of trait resilience has waxed and waned over the last 65 years, and it seems that at least every 20 years or so resilience constructs bounce back in the psychological literature with approximately the same key characteristics, but sometimes with new construct names.

Modern Resilience Constructs

Chronologically, the constructs used to assess and describe resiliency have been: ego-strength, hardiness, mental toughness, grit, and resilience (Barron; 1953; Duckworth, Peterson, Matthews, & Kelley, 2007; Freud, 1927; Kobasa, 1979; Strycharczyk & Clough, 2015; Wagnild & Young, 1993). Each of these constructs is described in the following sections.

Ego-Strength. In the 1920's, the psychoanalytic theory of personality was presented in a series of essays by Freud where he described an abstraction that he named the ego. The ego was hypothesized to reside in the brain above the limbic system and below the cerebral cortex (Freud 1927). The function of the ego was to act as moderator between the pleasure-seeking id and reality (Freud, 1905, 1927; Nunberg 1942). Psychoanalytic personality theorists proposed that individuals with greater ego-strength adapted more effectively to change, tolerated greater delay of gratification, exhibited greater mental endurance to both internal and external stressors, and had greater potential to cope with, or recover from, adversity and mental health difficulties than did those with less ego-strength (Barron, 1953; Freud, 1927; Gottesman, 1959). The strength of ego, as a personality trait, was hypothesized to be predominantly influenced by childhood experience and to become fully developed between the age of 12 and the beginning of adulthood, at which time ego-strength remains relatively consistent across situations and the lifespan (Freud, 1905). As it was theorized that ego-strength was associated with coping and could be measured to predict response to psychotherapy, it was of practical value to identify and measure ego-strength. (Barron, 1953). Self-report scales for this purpose were developed in the 1950s and

remain in use today. For example, Barron's Ego-Strength (Es) Scale for Minnesota Multiphasic Personality Test (MMPI) and the MMPI scales for depression, hysteria, paranoia, and schizophrenia were found to be negatively correlated (r = -.60, -.39, -.62, and -.55) (Barron, 1953; Block & Kremen, 1996; Gottesman, 1959; Klohnen, 1996).

Of note, in the current literature, the terms ego-strength and ego-resiliency have often been used interchangeably (Vecchione, Allesandri, Barbaranelli, & Gerbino, 2010). Block stated "ego-resiliency encompasses a set of traits reflecting general resourcefulness and sturdiness of character, and flexibility of functioning in response to varying environmental circumstances" and ego-resilient individuals are "engaged with the world but not subservient to it" (Block, 1969 as cited in Luthar, Cicchetti, & Becker, 2000, p. 547). It is also worth noting that there is a substantial amount of variability among researchers in defining the elements that comprise ego-strength. The variation is evident in both the range of operational definitions that have been provided for the construct as well as the content that comprises the specific factors and instrument items that have been selected to measure ego-strength (Barron, 1953; Block & Kremen, 1996; Freud, 1927; Gottesman, 1959; Klohnen, 1996; Vecchione et al., 2010).

Hardiness. In the 1970s, the construct of *hardiness*, described as an ability to withstand or adapt to stress, was introduced (Kobasa, 1979). Maddi (2002) reflected that work on the hardiness construct began in 1974 when his graduate student, Kobasa, discussed an article from a popular American women's magazine, the *Family Circle*. The article described ways to avoid stress (and advised to avoid change) with a very direct message delivered in the typical style of the pop-culture magazine: *avoid all stress: stress will kill you*. This message contradicted Maddi's research findings related to creativity that suggested that "creative-types" tended to seek out challenge and change, and that some creative individuals appeared to thrive under stress or

challenge (Maddi, 2002). Inspired by this contradiction, Maddi and Kobasa hypothesized that it was possible that some individuals were "hardy," and that hardiness moderated stress-related outcomes. (Maddi, 2002). They also hypothesized that hardiness to stressors could be a characterological construct (Kobasa, 1979).

To examine stress responses, they studied employees of the Bell Telephone Company in Chicago and recorded the number and type of stressors the employees reported (Kobasa, Maddi, & Khan, 1982). They then examined how each stressor related to self-reported mental and physical health symptoms (Kobasa, et al., 1982). A novel addition in this investigation is that the operational definition of stress was expanded to include a range of situations that required an individual to adapt. Positive and negative stressors, both chronic and temporary, were investigated (Maddi, 2002). Individuals who experienced a high load of stress and reported few symptoms were classified as *hardy* (Kobasa 1979; Kobasa et al., 1982). The five constructs that were identified in their research as relating to hardiness were: external locus of control, powerlessness, security, alienation from the self, and alienation from work (Kobasa, 1979). These five constructs were then fitted to three factors through subsequent analyses. "The Three C's" (commitment, control, and challenge) that comprise the theoretical base of many hardiness and resilience measurement instruments in use today stem from this initial research (Kobasa, 1979; Maddi & Khoshaba, 1994; Wagnild, 1991).

Mental Toughness. Athletic coaches of the 1980s coined the term *mental toughness*. Clough, Marchant, and Earle (2007) explained that, although the early definitions described mental toughness as a sporting ability related to consistent high performance, the factors that comprise mental toughness suggest it is better conceptualized as a personality trait than as an ability per se. Redefining mental toughness as a sport-related personality trait moved it into the

domain of sports psychology (Clough et al., 2007). Subsequently, mental toughness was adopted/generalized by several other areas of psychology to describe many types of performance that require mental fortitude under pressure (e.g., police training, academia, business management, clerical work) (Clough et al., 2007; St. Clair-Thompson, Giles, McGeown, Putwain, Clough, & Perry, 2017). Mental toughness researchers have found this trait to be positively related to well-being, being "comfortable in one's own skin" (Clough, 2007, p.2), openness, risk-taking, seeking challenge, optimistic challenge assessment (i.e., the likelihood to perceive challenge as an opportunity rather than a threat), and overall levels of achievement (Nicholls, Polman, Levy, & Blackhouse, 2004; Stamp, Crust, Perry, Clough, & Marchant, 2014).

The "Four C's" theorized to comprise mental toughness are control, commitment, challenge, and confidence (Clough et al., 2007). In this model, mental toughness is defined as "the personality trait which determines… how people deal effectively with challenge, stressors, and pressure… irrespective of circumstances." (Strycharczyk & Clough, 2015, p. 2). As might be expected, because greater mental toughness was found to correlate with higher performance in professional athletes, much of the mental toughness literature has focused on high performing athletes and methods to increase both mental toughness and athletic performance (Clough et al., 2007; Strycharczyk & Clough, 2015; Weinberg, Freysinger, & Mellano, 2018). Herein lies a paradox: Mental toughness is conceptualized as an enduring personality trait and is also defined as a malleable trait that which can be coached, trained, and increased. As constructs such as commitment and confidence are generally accepted to be amenable to change through training and experience, and personality traits are generally associated with that which is not amenable to change, I discuss this inconsistency found throughout much of the resilience literature in an upcoming section "a theoretical paradox."

Grit. Around the time that the United States was involved in lengthy military campaigns in Iraq and Afghanistan, and the sitting American president was heard to frequently say "stay the course" in reference to that military activity, the military psychologists of the new millennium described determination, adaptation, and perseverance in the face of difficulty, as a measurable factor called *grit* (Duckworth, Peterson, Matthews, & Kelly, 2007; Duckworth & Quinn, 2009; Matthews, 2008). Duckworth et al. (2007) defined grit as a personality trait that described "perseverance and passion for long term goals… that entails working strenuously toward challenges, maintaining effort and interest over years despite failure, adversity, and plateaus in progress" (Duckworth, et al. 2007, p. 1087). Early studies that measured grit found that successful completion vs. dropout from the U.S. West Point Military Academy could be predicted by grit scores (Duckworth et al., 2007; Kelly, Matthews, & Bartone, 2014).

As grit is a relatively new construct, there is, as one meta-analytic research team described, "much ado about grit" (Credé, Tynan, & Harms, 2017, p. 1). Maddi et al. (2013) found grit to be significantly correlated with hardiness ($r = .31$, $p < .001$) and described that most of the overlap between grit and hardiness was explained by the commitment and control factors. A recent investigation of 2,321 pairs of twins in the U.K. found a significant genetic overlap between grit and the Big Five personality characteristic of conscientiousness ($r = .86$) (Rimfeld, Dale, Kovas, & Plomin, 2016). That is, two gene pairs related to both conscientiousness and grit, and the researchers concluded that if conscientiousness is being measured, there is little incremental value to be gained from also measuring grit. The increase of explained variance was found to be $R^2 = .06$ (6%) when grit was added to their multivariate model that included conscientiousness (Rimfeld et al., 2016). Given the current debate around grit, it appears that further validation studies are needed to investigate the construct in greater

depth.

Resilience. Personality and positive psychology researchers in the 1990s used the term *resilience* to describe "personal qualities that enable one to thrive in the face of adversity" (Connor & Davidson, 2003, p. 76). A systematic review of 271 resilience-related articles noted that a common occurrence in the resilience literature is the recognition of the difficulty of precisely defining resilience and found substantial variability among the given definitions for resilience (Windle, 2011). Two distinct schools of thought are present in the resilience literature: those that define resilience characterologically (i.e., as a personality trait) and those that define resilience dynamically (i.e., as a process) (Bartone, Ursano, Wright, & Ingraham, 1989; Gillespie, Chaboyer, & Wallis, 2007; Smith-Osborne & Bolton, 2013). The two theoretical bases share the underlying concept that resilience is a property or ability related to bouncing back or thriving in conditions of adversity (Connor and Davidson, 2003; Padesky & Mooney, 2012).

Diverse operational definitions have been proposed for characterological/trait-like resilience (Canavan, 2008; Cicchetti, 2010; Fletcher & Sarkar, 2013; Wagnild, 2013; Windle, 2011). These include (a) average/normative development in the presence of a higher than average cumulative quantity of risk factors ("beating the odds"), (b) functioning at an average or above average level or maintaining personal equilibrium under stress or current adverse conditions (coping/adjustment), (c) recovery to a baseline or average level of functioning after catastrophic adversity or severe deprivation ("bouncing back"/returning to homeostasis), (d) achieving expected norms under average conditions, and, in some contradiction to (d), (e) the achievement of an anomalous positive outcome following adversity (Canavan, 2008; Garmezy, 1993; Seligman & Csikzentmihalyi, 2000; Masten, 2001; Smith-Osborne & Bolton, 2013;

Wagnild & Young, 1993; Windle, 2011).

For both characterological and dynamic definitions of resilience, it is assumed that resilience is negatively correlated with mental health problems such that those who are higher in resilience when exposed to trauma would be less likely to develop depression, anxiety, or post-traumatic stress disorder (PTSD) (Barron, 1953; Bartone et al., 1989). Thus, a portion of the resilience literature has focused on combat veterans, survivors of natural disasters, and individuals who have experienced assault or other forms of abuse (Bartone et al., 1989; Barron, 1953; Garmezy, 1974; Windle, 2011). Also, in accord with a general definition of resilience it is assumed that individuals who are higher in resilience overcome adversity and either achieve average or exceptional outcomes (depending on which theoretical base is used). Thus, a portion of the resilience literature has focused on poverty, bullying, discrimination based on sex, race/ethnicity, sexual orientation, and coping with acute or chronic serious health difficulties (e.g., cancer, organ transplant, invasive high-risk surgeries, or diabetes). Some of the resilience literature expanded the definition of adversity from situations where positive prognoses would be relatively rare to include situations where positive outcomes are not exceptional, and the adversities include typical life changes that many people experience (e.g., leaving home to attend university, moving, becoming a parent, marriage, divorce, aging, retirement) (Kobasa, 1979; Maddi, 2002). Returning to the analogy with metallurgy, adversity can also be defined as repeated stress loadings, thus resilience is also studied in high-stimulus or high-risk careers where individuals are exposed regularly to stressors such as first responders, law enforcement, hospital staff (Connor & Davidson, 2003; Wagnild & Young, 1993).

Theoretical Conceptualizations of Resilience

Resilience constructs, when investigated in the behavioural sciences, have typically been

associated with either the hypothesis that resilience in adults is consistent across the lifespan, or that resilience can change in response to life events, adversity, training, or the passage of time (Barron, 1953; Connor & Davidson, 2003; Duckworth & Quinn, 2009; Funk, 1992). The postulate that resilience remains constant is the core premise underpinning characterological models of resilience (Barron, 1953; Freud, 1927; Funk, 1992; Klohnen, 1996; Kobasa, 1979; Maddi, 2002). Conversely, the assumption that resilience is malleable or has plasticity is the tenet underlying dynamic models of resilience (Clough et al., 2007; Garmezy, 1974; King, King, Vogt, Knight & Samper, 2003; Lyons, Parker, Katz, & Schatzberg, 2009; Padesky & Mooney, 2012). Both model categories and the state of permanence vs. malleability of resilience are found throughout the resilience literature.

The lack of agreement upon a definition for resiliency is further complicated by variability in the resilience literature related to how many, and which, factors best characterize resilience (Bartone et al., 1989; Connor & Davidson, 2003; Kobasa, 1979; Maddi, 2002; Smith et al., 2008; Wagnild & Young, 1993). Both characterological and dynamic models range in the number of factors or dimensions posited to describe resilience. Some resilience researchers have proposed that resilience is unidimensional and is best measured using a single factor (Barron, 1953; Campbell-Sills & Stein, 2007; Freud, 1927; Windle, Markland, & Woods, 2008). Others have proposed that resilience is comprised of two to five factors (Bartone et al., 1989; Connor & Davidson, 2003; Duckworth & Quinn, 2009; Kobasa, 1979; Maddi, 2002; Wagnild & Young, 1993).

Single-factor measurement models function under the premise that each item represents the same amount of resilience relative to other items and that the sum score describes resilience (Campbell-Sills & Stein, 2007). Multi-factorial models function under the premise that resilience

is better explained by a number of sub-constructs called factors (Connor & Davidson, 2003). There is additional variation in that in some instruments each factor accounts for an equal amount of resilience, and in other instruments the factors comprise unequal proportions of the overall resilience score. For example, Duckworth and Quinn's (2009) two-factor model described resilience as a combination of two factors: consistency of interest and perseverance, and each factor on the Short Grit Scale is comprised of six items. Wagnild and Young (1993) also proposed a two-factor structure for resilience: Personal competence (17 items), and Acceptance of self and life (8 items). Connor and Davidson (2003) described a five-factor model of resilience that underlies the CD-RISC-25 measure: Factor 1 (8 items) describes personal competence, standards and perseverance, Factor 2 (7 items) describes trust of one's instincts and stress tolerance, Factor 3 (5 items) describes acceptance of change and social relations, Factor 4 describes control (3 items), and Factor 5 (2 items) describes spirituality (Connor & Davidson, 2003). Those listed here represent are only a few examples of the vast array of proposed resilience structures. In general, there is substantial variability in the theoretical factor structure of resilience and in the instruments used to measure it. Given the differing schools of thought that relate to the factor structure, basis, and very nature of resilience it is not surprising that measuring this construct is complex, and sometimes paradoxical.

Utility of Studying Resilience

Overall, the breadth of the literature and the inconsistency of definitions makes it difficult to provide a comprehensive summary statement about how the construct of resilience has made a difference in our understanding of personality and its relation to adjustment and/or functional improvements following adversity. The first resilience construct identified: ego-strength, was used to inform psychiatric inpatient treatment decisions; such that it was assumed that

individuals with higher scores on measures of ego-strength would better respond to therapeutic interventions (Barron, 1953). Given this occurred prior to the development of many modern statistical procedures such as meta-analysis, and also pre-dated many current statistical methods, statistical evidence of this proposed relation is limited. Hospital-based measurement of ego strength, however, declined in the decades that followed, and the assessment of ego-strength to inform psychotherapeutic treatment is not a standard clinical practice today.

Resilience and similar constructs have been examined meta-analytically in specific contexts. Eschleman, Bowling, and Alarcon (2010) summarized 180 studies of hardiness and reported that hardiness is positively related to constructs expected to reduce the impact of stress (self-esteem: M_r = .43, N = 2,610; optimism M_r = .43, N =1,290; extraversion M_r = .25, N = 2,190; and self-efficacy M_r = .26, N = 206). The authors' also reported that hardiness is negatively related to constructs known to exacerbate the effects of stress (negative affect: M_r = -.37, N = 3,115; neuroticism M_r = -.36, N = 2,848; trait anger M_r = -.34, N = 378; and emotional reactivity M_r = -.33, N = 666). Finally, this meta-analytic review summarized that hardiness was positively related to social support M_r = .29, N = 7,190, problem-focused coping M_r = .28, N = 4,786; approach-focused coping M_r = .38, N = 991, and both school (M_r = .21, N = 623) and work (M_r = .17, N = 676) performance (Eschelman et al., 2010).

Hu, Zhang, and Wang (2015) conducted a meta-analysis of 60 studies to examine the relation between trait resilience and overall mental health. In their analyses, they coded mental health constructs as positive indicators of mental health (e.g., life satisfaction, positive affect) or as negative indicators of mental health (e.g., symptoms of depression, symptoms of anxiety, negative affect). The authors' reported the summary effect size for positive indicators (M_r = .50, N = 20,092) and negative indicators (M_r = -.36, N =45,188) and described that these relations

were moderated by the percentage of male participants and adversity (coded as: yes/no). The moderation effects were that the relation between both positive and negative indicators of mental health decreased in magnitude as the percentage of male participants increased, and the observed summary effect size increased as the number of participants coded as having at least one adverse experience increased (Hu et al., 2015).

Specific mental health constructs and the relation of these to resilience have also been the foci of meta-analyses. Wermelinger-Ávila, Lucchetti, & Lucchetti (2017) summarized the results of seven studies of depression symptoms in older adults. The authors conducted a meta-analytic review and reported that the relation between scores on measures of resilience and scores on measures of depression were (M_r = -.35). However, in this report the total count of study participants was not given (Wermelinger-Ávila et al., 2017.) Another meta-analytic review summarized five studies of client's resilience and psychological distress in the context of psychotherapy and reported that the relation was M_r = -.40 for N = 1,248 study participants (Bartholomew, Robbins, Joy, Kang, & Aguiñiga, 2020). Oshio, Taku, Hirano, & Saeed (2018), conducted a meta-analysis on 30 studies that examined the relation between resilience and the big five personality traits. The authors reported that the summary effects were: neuroticism, M_r = -.46, N = 15,609; extraversion, M_r = .42, N = 15.379; openness, M_r = .34, N = 11,420; agreeableness, M_r = .31, N = 11,032; and conscientiousness, M_r = -.40, N = 11,420 (Oshio et al., 2018). Each of these meta-analyses examine correlational data, likely because the bulk of the resilience literature is comprised of cross-sectional studies. Each also describes there are definitional inconsistencies across studies, and across the range of measures used to assess adult resilience. Given the discrepancies and breadth of procedures to study resilience, it is important to also examine the instruments that are being used, and to work toward developing consistent

processes and operational definitions. Methodologically, much work needs to be done to improve the clarity and quality of how resilience is studies before the study of resilience can be used improve our understanding of functioning following adversity.

Resilience Measurement

Hunsley and Mash (2018) described that psychometric tests may be administered for a number of differing practical functions in the clinical context including: diagnosis/categorization, screening, prediction, case conceptualization, treatment planning, treatment monitoring, and treatment or intervention evaluation. The first resilience measure, the Barron's Ego-Strength (*Es*) Scale for Minnesota Multiphasic Personality Test (MMPI: Hathaway & McKinley, 1942) was frequently administered as a part of the patient intake process of many psychiatric hospitals in the 1960s and 1970s (Barron, 1953; Camara, Nathan, & Puente, 2000). In hospital settings, the function of ego-strength measurement was to predict response to psychotherapy, thus aiding psychiatrists and psychologists in the selection of interventions for their patients based on the probability that psychotherapy would produce a beneficial result (Barron, 1953; Block & Kremen, 1996; Hathaway & McKinley, 1942). Modern resilience measurement, however, typically only occurs in research settings, although some resilience test developers propose that their measure would be of utility in the clinical setting. Certain instruments (e.g., the CD-RISC, 25-item version) have been validated for change sensitivity in research settings and hypothesized to be of relevance to clinical settings (Connor & Davidson, 2015). Another context where resilience measurement has been used is in the evaluation of educational programs and workshops aimed at increasing individual resilience in the workplace, academic settings, or as a personal self-improvement goal. Thus, resilience measures can be divided into two categories: instruments designed for specific contexts or for specified populations (e.g., health-related

hardiness, academic resilience, resilience related to military deployment, LBGTQIA+ resilience, visible minority resilience, youth or geriatric resilience) and general resilience measures (i.e., measures that are not indicated by the test developers to context-specific, or for use with specified populations).

General vs. Context-Specific Resilience Measures for Adults. General resilience scales are those that can be used with language- and culture-appropriate adult population and assess an overall magnitude of resilience for each person regardless of the contextual purpose of the study (Bartone, 1989). For example, the CD-RISC scales, and Dispositional Resilience Scales (DRS) have been used to study resilience in refugees, homeless individuals, medical students, emergency healthcare professionals, cricketers, university undergraduates, and many other populations of interest (Bartone, 1989; Connor & Davidson, 2015). Similarly, the Brief Resilience Scale (BRS: Smith et al., 2008), Ego-Resilience-89 scale (ER-89: Block & Kremen, 1996), the Resilience Scale 25-item Version (RS-25: Wagnild & Young, 1993), and the Resilience Scale for Adults (RSA: Friborg et al., 2006) are general resilience measures that have been applied across contexts to examine overall resilience in a diverse range of populations. Some general resilience scales such as the CD-RISC-25 and Barron's *Es* Scale have been examined extensively for their psychometric properties and norms for various populations are available (Barron, 1953; Connor & Davidson, 2015). Other scales, particularly those developed more recently, such as the Brief Resilience Scale (BRS: Smith et al., 2008) and the Short-Grit Scale (Grit-S, 8-item version; Duckworth & Quinn, 2009) appear to be frequently administered, but have less psychometric information available.

In contrast to general resilience scales, context-specific scales have been designed to assess resilience in specific environments (e.g., the workplace, academia, professional sports), to

assess the strength of the relation of resilience scores to a particular outcome (e.g., academic performance, athletic performance, increased or decreased odds of psychological distress given specific experiences), or to assess resilience given specific antecedent criteria (e.g., post-traumatic growth). A few examples of contextual scales related to specific environments are the 25-item Suicide Resilience Inventory (SRI-25: Rutter, Freedenthal, & Osman, 2008), the Health-Related Hardiness Scale (HRHS: Pollock & Duffy, 1990), and the Resilience at Work Scale (RWS: Winwood, Colon, & McEwen, 2013). As these measures were designed for use in distinct settings under defined circumstances, these scales would not be appropriate to measure general resilience.

An example of a context-specific resilience instrument related to a particular outcome is the Deployment Risk Resilience Inventory (DRRI: King et al., 2003). The DRRI was developed as a collection of risk factors to assess resilience to developing Post-Traumatic Stress Disorder (PTSD) following military deployment to Iraq. Higher scores on the DRRI were found to correlate with the onset, severity level, and outcome prognoses of PTSD (King et al., 2003). High scores on the instrument represented the endorsed number of adverse experiences the individual encountered, mostly during active deployment and combat situations (King et al., 2003). Scales for measuring context-specific resilience like the DRRI are often comprised of lists of events and situations designed to represent personal or environmental factors related to a specific context (e.g., military combat), and these are conceptualized as risk factors for a specific outcome (e.g., PTSD). Due to the context-specificity, there is less ambiguity about the suitability of a given measure across contexts, because use across contexts is impractical (e.g., the DRRI would likely be of little use to investigate resilience in first year university students).

An example of a context-specific resilience scale related to specific antecedent criteria is

the Post-Traumatic Growth Inventory (PTGI: Tedeschi & Calhoun, 1996). Calhoun and Tedeschi pioneered the term *post-traumatic growth* in the mid-1990's and defined post-traumatic growth as "the experience of positive change that the individual experiences as a results of the struggle with a traumatic event" (Calhoun & Tedeschi, 2013, p.6.). Positive change, as measured by the PTGI is examined in five sub-domains: relating to others, new possibilities, personal strength, spiritual change, and appreciation of life (Tedeschi & Calhoun, 1996). These sub-domains can be described as measuring change both in personality characteristics as well as the individual's development of skills or new cognitive perspectives. The construct (and related measurement) of post-traumatic growth requires that the individual has experienced trauma. Thus, here also, there is less ambiguity about which research and clinical contexts and/or populations are suitable for this measure and/or for the study of post-traumatic growth. Similar to the specificity described above for the DRRI, the measurement of post-traumatic growth would not be under the umbrella of general resilience measures, because a specific antecedent experience (trauma) is required to explore and study post-traumatic growth.

There is greater ambiguity, however in the suitability of general resilience scales, as many of these have been used with a wide range of populations and situations. Although each of these scales was developed with a validation sample and most test developers report some reliability and validity information, it would be an error to assume that general resilience measures are "reliable and valid" in any sample of individuals (even those very similar to the norms given for the instrument) (Hunsley & Mash, 2018; Urbina, 2004; Vacha-Hasse, 1998). Thus, in the selection of a general resilience measure, it is necessary to evaluate the psychometric properties (i.e., evidence for reliability and validity in past administrations of the measurement instrument) to assess the potential utility of the measure for the specific purpose

and sample for which it will be used.

Psychometric Properties of Measurement Instruments

This section describes the common (and commonly misunderstood) psychometric properties of reliability and validity. The importance of understanding and assessing the psychometric features of any measurement instrument, in both the base of evidence in the literature, and at the level of the researcher's own sample, is explained. Adequate reliability and validity are both necessary (but not sufficient) conditions to ensure data flowing into, and conclusions flowing out of, a psychological research study is of a reasonable quality. This section also summarizes the Hunsley and Mash (2018) guidelines for the evidence-based selection of measures based upon psychometric properties. Although Hunsley and Mash (2018) focused on clinical assessment measures, the guidelines given for selecting quality instruments can be applied to the selection and use of any psychological measure.

Reliability. For any psychometric assessment instrument to be useful the scores it produces must be reliable (Hunsley & Mash, 2018). Reliability is the degree to which a test instrument yields consistent results (Urbina, 2004). It is important to note that a very common error made by clinicians and researchers is to assume that reported psychometric reliability estimates are properties of the measure itself – and that these apply across settings, populations, and subsequent administrations (Vacha-Hasse, 1998). All too common in the literature are statements such as "we selected the ABC Inventory of X because it has been proven to be a valid and reliable measure." In the reliability literature, the practice of making such an erroneous assumption is referred to as *reliability induction* (Vassar & Bradley, 2010). To conceptualize how reliability estimates reported for the development sample of an instrument cannot be simply applied to all subsequent uses of the instrument, Urbina (2004, p. 120) stated that such an

assumption "…would be akin to saying that a fine piano that is well-tuned will always be in tune, and will sound just as good regardless of the kind of music that is played on it, or of who plays it… the quality of the sound a piano makes is a function not only of the instrument itself, but also of variables related to the music, the piano player, and the setting (e.g., the acoustics of the room) where is it played." Similarly, the reliability of a psychological measurement instrument may also differ across populations, settings, and instrument administrations.

Despite the growing emphasis that has been placed on calculating and reporting reliabilities for each sample due to potential differences across settings, populations, and other variables, it is still common for researchers to commit reliability induction errors and provide only reliabilities from other sources (e.g., other studies, instrument administration manuals) under the assumption that these values apply to their study participant's scores and omit performing reliability analyses of their own data (Rouse, 2007; Vacha-Hasse, 1998). To avoid this error, three types of reliability, both in terms of evaluating the evidence of these reliabilities in terms of prior administrations of the instrument and evaluating current observed data, can be considered when determining the reliability of a test: internal consistency, test-retest reliability, and inter-rater reliability (Hunsley & Mash, 2018; Urbina, 2004; Youngstrom, Meter, Frazier, Hunsley, Prinstein, Ong, & Youngstrom, 2017).

Internal consistency. Internal consistency is the degree of equivalence among scores on the items that contribute to the total score on a measure, given the Classical Test Theory premise that each item contributes equally to an overall score (Cronbach, 1951). Although there are a number of ways to calculate internal consistency, the most commonly reported internal consistency statistic is Cronbach's coefficient α (Cronbach, 1951; Urbina, 2004). Coefficient α values between .70 and .79 are considered to be adequate, values from .80 to .89 are considered

to be good, and values around .90 are considered to be excellent (Hunsley & Mash, 2018). However, Cronbach's α values that are "too close" to 1.00 indicate near-perfect internal consistency, can indicate that a scale contains more items than are needed to adequately measure the construct, or that the scale has a very narrow scope (Youngstrom et al., 2017). A 'higher than excellent' internal consistency value is not problematic in and of itself, however internal consistency is worthwhile to keep in mind as it is good practice to use time for clinical assessment and research as efficiently as possible (Streiner, 2003).

Coefficient α, although the most popular measure of internal consistency, is not without limitations (Haynes, Smith, & Hunsley, 2011; Streiner, 2003). First, the reporting of the test statistic does not imply that the statistical assumptions of Cronbach's α (e.g., unidimensionality, uncorrelated error terms/independence in sampling, and fixed effects/tau equivalence) have been met (Dunn, Baguley, & Brunsden, 2014; Cronbach, 1951; Cronbach & Shavelson, 2004). Second, many individuals likely select coefficient α not because it is the most appropriate choice, or the best statistic among many options developed since the publication of Cronbach's 1951 article but select it because it is either the most familiar, or because it is relatively easy to calculate mathematically (Cronbach & Shavelson, 2004). As Cronbach stated: "the numerous citations to my article by no means indicate that the person who cites it had read it, and does not even demonstrate that he or she had even looked at it. I envision that the typical activity leading to the typical citation as beginning with the student laying out his research plans for a professor or a draft report, and it would be the professor's routine practice to say, wherever a measuring instrument was used that the student ought to check the reliability of the instrument… [and] the professor would suggest using the alpha formula because the computations are well within the reach of almost all students" (Cronbach & Shavelson, 2004, p. 392). Similar to most test

statistics which have assumptions, when these are violated the estimates produced by the test-statistic become less accurate; in the case of Cronbach's α, assumption violations produce inflated estimates of reliability (Dunn et al., 2014).

A number of alternatives to Cronbach's α have been proposed. The Kuder-Richardson 20 formula, omega (ω), and bootstrapping have been proposed to examine internal consistency (Dunn et al., 2014; Haynes et al., 2011; Urbina, 2004). Despite that these alternatives have been described to be more accurate estimates of internal consistency under certain circumstances, it is very rare to see any of these statistics used in the psychological literature (Haynes et al., 2011).

Test-retest reliability. Test-retest reliability is the property of scores on an instrument that describes their robustness to time-sampling error (Field, 2005; Urbina, 2004). More plainly stated, it is the quality of scores on the instrument in a specific context to produce consistent results across administrations assuming the true value of variable being measured remains constant (Urbina, 2004). Test-retest reliability is typically assessed using Pearson's product-moment correlation coefficient *r* or Fisher's modified Intra-class Correlation Coefficients (ICCs) based on Pearson's *r* (Fisher, 1925; Pearson, 1900). Test-retest reliability correlations greater than .70 are considered to be adequate if the time between administrations is several days, good if the time between administrations is several months, and excellent if the time between administrations exceeds a year (Hunsley & Mash, 2018).

Inter-rater reliability. Inter-rater reliability is the robustness of the scores on an instrument in a given context to inter-scorer differences (Urbina, 2004). Dependent on the attributes of the data to be evaluated, there are several ways to compute inter-rater reliability, ranging from a simple percentage of agreement to more complex computations such as Pearson's *r*, Cohen's or Fleiss' κ, Pooled Cohen's κ, Gwet's AC2, or Krippendorff's α (De Vries, Elliott,

Kanouse, & Teleki, 2008; Inter-Rater Reliability, 2018; Urbina, 2004). The most commonly reported inter-rater reliability calculation is Pearson's *r*, and values between .70 and .79 are considered to be adequate, values from .80 to .89 are considered to be good, and values exceeding .90 are considered to be excellent (Hunsley & Mash, 2018).

Validity. In addition to reliability, for any psychometric instrument to be useful there must also be evidence of validity in the scores obtained for the sample being studied. Validity refers to the degree which scores on an instrument measure what they are purported to measure (Hunsley & Mash, 2018; Messick, 1990). Historically, validity was grouped into different types. Contemporary validity theory posits that the validity of test scores is a singular concept, and validity is determined by the degree to which there are multiple forms of evidence for validity (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014; Haynes et al., 2011; Hunsley & Mash, 2018; Reeves & Marbach-Ad, 2016). Evidence for validity may be in the form of content validity, construct validity, and validity generalization[1] (Haynes et al., 2011; Hunsley & Mash, 2018). Each type of validity evidence must be considered when determining the applicability and utility of scores on a psychometric instrument (Cronbach & Meehl, 1955; Hunsley & Mash, 2018). Further, Messick (1990) explained that validity is not binary – that is, it is not an "all or none" property – and that there are degrees of validity based upon the available evidence.

Although some validity evidence may seem straightforward and can be considered to be a property of the observed scores on an instrument, validity also pertains to the judgments that relate to the scores on an instrument in a given context (AERA, APA, & NCME, 2014; Hunsley & Mash, 2018). In addition to determining if the instrument has measured the desired construct

[1] Validity generalization is described in greater detail in the section "Meta-Analytic Techniques".

accurately or precisely, acceptance of a conclusion as valid must also include the consideration of how appropriate the interpretation of the data is relative to the established norms for the measure, given the relevant variables of the administration context (e.g., culture, individual difference variables) (AERA et al., 2014; Haynes et al., 2011; Hunsley & Mash, 2018). Hierarchical structures proposed to assess evidence of validity vary in the psychometric literature (AERA et al., 2014; Haynes et al., 2011; Hunsley & Mash, 2018; Messick, 1990). For the discussion that follows, and the methods used in this dissertation, I have used the structure recommended in Hunsley and Mash (2018).

Content validity. Haynes et al. (2011) described construct validity as "the degree to which elements of an instrument are relevant to and representative of the targeted construct" (p. 19). Hunsley and Mash (2018) provided guidelines to assess content validity. When instrument developers provide a clear definition of the measured construct and have also ensured that items that comprise the inventory fully capture the breadth of the construct, the measurement instrument is considered to be *adequate*. In addition to the above criteria, if both the instructions and test items were either evaluated by expert judges or pilot tested with participants the instrument can be considered to have *good* content validity. *Excellent* content validity can only be concluded if an instrument meets all of the previously described criteria and the item evaluation has occurred across multiple judges and the evaluation/ratings given by judges were defined and assessed quantitatively.

Construct validity. Cronbach and Meehl (1955) described construct validity as the "acceptance of the universe of content as defining the variable to be measured" (p. 2). Construct validity is also property of a set of obtained scores and not an inherent and generalizable quality of the instrument (Haynes et al., 2011; Reeves & Marbach-Ad, 2016). Common forms of

evidence for construct validity are: predictive, concurrent, convergent, discriminative, and incremental validity (Cronbach & Meehl, 1955; Hunsley & Mash, 2018).

Predictive and concurrent validity are sometimes referred to as criterion validity, as these types of validity describe the relation between the observed scores in a given sample to a criterion (outcome) variable (Cronbach & Meehl, 1955). The difference between predictive and concurrent validity is merely temporal – predictive validity describes the ability of the observed set of scores on the instrument to predict a future outcome, and concurrent validity is the ability of the observed set of scores on the instrument to predict an outcome variable that occurs simultaneously with the measurement (Goodwin, 2008). Convergent validity describes the degree of concordance between the observed scores of two measures that were intended to measure the same construct (Messick, 1990). Divergent validity describes the degree of dissimilarity (lack of correlation) between sets of scores obtained on measures designed to assess differing constructs and is used to demonstrate construct distinctiveness (Urbina, 2004). Incremental validity is the amount of predictive power that is added by using a set of scores obtained on a given instrument over and above what can be found using other measures or sources of information (Hunsley & Meyer, 2003). For example, a validation study of grit found that there is substantial overlap in the inventory items commonly used to assess both grit and conscientiousness, thus the researchers questioned the incremental validity (i.e., added value) of grit as a separate (and separately measured) construct (Credé et al., 2017).

Hunsley and Mash (2018) provided a set of psychometric guidelines to assist in the selection of measures with adequate, good, and excellent content and construct validity. *Adequate* construct validity includes some independent replication of the forms of validity evidence listed above, *good* construct validity includes several independent replications across

several types of validity evidence, and a rating of *excellent* would include all of the above plus evidence of incremental validity.

Meta-Analytic Techniques

Several methods have been developed to collect, describe, summarize, and/or statistically analyze multiple studies pertaining to a given topic (Rousse, 2007). One of type of summary is the systematic literature review, which is generally descriptive in nature (Higgins & Green, 2008; Ulman, 2011). Another family of summaries are the meta-analyses, where statistical data are weighted and aggregated across studies to mathematically compare effect sizes, and may also be used to calculate summary effect sizes that can be used to examine evidence for reliability and validity (Borenstein, Hedges, Higgins, & Rothstein, 2009; Quintana; 2015; Viechtbauer, 2010).

Systematic Review. A systematic review may be a collated summary of the available evidence on a topic meeting a set of pre-defined criteria, or a summary of the available information related to a specific research question (Higgins & Green, 2008). Systematic reviews are considered to be less researcher-biased and more inclusive because article inclusion is carried out in a precise and pre-defined *systematic* procedure (Higgins & Green, 2008; Ulman, 2011). Researchers conducting systematic reviews may employ either qualitative or quantitative methods, or some combination of both. The Cochrane Collaboration developed a gold standard for systematic reviews to be carried out by large teams of researchers (Higgins & Green, 2008). Although it is not possible for an individual researcher to adhere to all of these guidelines, the central tenets outlined by the Cochrane Collaboration that describe how to conduct thorough and non-biased algorithmic searches and how to select and structure *a priori* inclusion/exclusion criteria can be scaled to any size systematic review. The Cochrane Collaboration structure is built around the assessment and summarization of data obtained in Randomized Controlled

Trials (RCTs) and is frequently used to assess and summarize pharmaceutical trials and medical outcome data. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Guidelines are another set of structured systematic review processes used by researchers to conduct systematic reviews (Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009; Moher, Shamseer, Clarke, Ghersi, Liberati, Petticrew, Shekelle, & Stewart, 2015). PRISMA Guidelines provide a flexible framework for the generation of robust systematic review protocols for non-RCT studies with greater flexibility in other areas as well, such that these reviews can be feasible carried out by smaller groups of researchers and graduate research labs (Moher et al., 2009).

Meta-Analysis. As noted previously, meta-analysis is not a singular method, nor is it a single statistical technique applied across methods but is instead a family of techniques and methods. The term meta-analysis was coined by Glass and means quite literally, 'the analysis of our analyses' (Glass, 1976; Hunter & Schmidt, 2004). Meta-analyses come in a variety of forms. The most common of these is the statistical summary of observed effect sizes across studies (Borenstein et al., 2009).

Meta-analysis outline. A general process to conduct a statistical summary meta-analysis:

1) Select an effect that would be useful to summarize and choose the model that would be appropriate for these data (i.e., fixed-effects or random effects)
2) Define the criteria for articles to be included or excluded from this summary, and design a systematic algorithmic search to collect these articles across appropriate and relevant sources
3) Extract the effect sizes from the articles utilizing a method where inter-rater reliability is assessed, and data extraction follows a clearly defined coding process

4) Calculate the summary effect size, variance, standard error of the estimated effect size mean, and confidence intervals in accordance with the selected model
5) Assess the model for variance heterogeneity (in a more complex model, moderator analyses may be used to investigate pre-defined potential sources of error variability)
6) Assess the included articles for potential publication bias (for each set and sub-set of meta-analytic investigations)
7) Report and interpret the obtained information and describe the relevance and relation to existing theories related to the examined effect size (if applicable)

Reliability generalizations (RGs) and validity generalizations (VGs) can be used to summarize the previously described forms of reliability and validity evidence across studies (Schmidt & Hunter, 1977; Vacha-Hasse, 1998). It is typical, when examining findings that span several studies, research settings, individual researchers/teams of researchers, and time, there will be substantial variation in the observed effect sizes across instrument administrations and administration contexts. Moderator analyses (both group comparison analyses and meta-regression) can be used to investigate if certain parameters such as sample characteristics, settings, methods, and/or other variables of interest explain the observed variation in reported effect sizes across studies (Borenstein et al., 2009; Pigott, 2012; Quintana, 2015; Schmidt & Hunter, 1977; Vacha-Hasse, 1998; Viechtbauer, 2010).

It must also be noted that, although meta-analytic summaries are useful to understand how an instrument has performed across contexts in the past, such summaries do not guarantee future instrument performance. Summaries can be used to inform the selection of an appropriate

measurement tool but cannot replace the need to examine reliability and validity information in each new study that uses a measure. Even for instruments with an established history of use, the evaluations of reliability and validity evidence are ongoing processes, as constructs such as mental health, adjustment, and success, are defined and re-defined with the passage of time.

Fixed-Effect vs. Random-Effects Models. As noted in the steps outlined above, meta-analytic researchers select either a fixed-effect model or a random-effects model framework in which the meta-analyses are carried out (Borenstein et al., 2009; Hunter & Schmidt, 2014). Fixed-effect models (also called models of a common-effect) are appropriate when it is theorized that there is only one true effect and that observed effect variability across studies is entirely due to error (Borenstein, Hedges, Higgins, & Rothstein, 2014). Although a few homogenous effects have been found in the social and behavioural sciences (e.g., average working memory digit-span, general motor abilities), homogenous population effects and equivalent effect distributions are fairly rare in psychology (Borenstein et al., 2009; Schmidt, Law, Hunter, Rothstein, Pearlman, & McDaniel, 1993).

For many constructs in the behavioural sciences, true effects truly vary across populations and are moderated by participant characteristics (e.g., age, sex/gender identity, years of education, geographic location, culture). In studies where is theorized that true effects vary, a random effects model is the appropriate analytic framework as the model includes accounts for the variation in the observed effect size variance across studies (Borenstein et al., 2009; Borenstein et al., 2010). More technically, a random effects model assumes that the true effect sizes across a collection of studies are sampled from an unknown distribution of effect sizes with a mean denoted by θ (theta) and a standard deviation denoted by τ^2 (tau-squared) and that variation (not assumed to be error) must mathematically be included in the analyses of effects

(Pigott, 2012). Therefore, the random-effects model uses an estimated weighted population mean, θ, and an estimation of the true population variance, τ^2 to describe the observed effect variation across studies (Pigott, 2012). Although, many researchers have described model selection to have been based upon a calculated estimate for τ^2 (T^2) that was near to zero, it is recommended that model selection should be theoretically based (Borenstein et al., 2009; Hunter & Schmidt, 2004). If a fixed effects model is applied when a random effects model is needed the result can be erroneously inflate summary effect size estimates (Schmidt et al., 1993). Often, fixed effects models are erroneously selected and favoured because there are substantially fewer calculations (Hunter & Schmidt, 2004). If, however, a random effects model is selected, and the true effect does not vary or varies minimally across observed sets of scores, there is no statistical penalty to the summary effect size to having selected a random effects model (Borenstein et al., 2009; Hunter & Schmidt, 2004). Thus, in cases where the true score variance distribution is entirely unknown, or is theoretically debated, it is better practice to select the random effects model.

Effect Size Summarization. Effect size summarization processes differ for fixed effects and random effects models. In this section, the procedure to calculate summary effects for the random effects model is described, as that model is better suited to resilience data than is a fixed effects model. Borenstein et al. (2009) and Hunter and Schmidt (2004) described that one process to calculate the effect summarization (mean weighted effect size estimate) in a random effects model is as follows: 1) Collect, compute, or convert effect sizes such that all effect sizes used in each analysis use a consistent statistic 2) Compute the weight for each effect size. The purpose of this step is to weight (balance) the contribution to the overall summary effect size for an observed effect based upon variance, which also accounts for sample size. 3) Compute the

weighted mean effect size and its variance. The standard error of the estimated mean effect size and the confidence intervals surrounding this estimate may also be calculated at this stage.

Validity Generalization. Validity generalization is the generation of a mean estimate of the observed effect sizes across studies, and subsequent analyses that specifically pertain to the investigation of validity evidence (Borenstein, 2009; Hunter & Schmidt, 2004; Quintana, 2015; Veichtbauer, 2010). Validity generalization is also called VG or Hunter-Schmidt meta-analysis (Schmidt & Hunter, 1977; Borenstein et al., 2009). Schmidt & Hunter (1977) developed validity generalization following an observation that validity evidence for a single measurement instrument varied situationally. The rationale behind conducting validity generalizations is to summarize the past evidence of validity/psychometric performance of an instrument (Rafilson, 1991). Of interest in many of these investigations is to explain observed variability across studies (i.e., determine if there are moderators that can explain the differential performance of an instrument across contexts (e.g., settings, populations, methods) (Borenstein et al., 2009; Hunter & Schmidt, 2004; Pigott, 2012; Schmidt & Hunter, 1977). Validity generalization can assist researchers by providing information relating to the stability of validity evidence across studies and may also provide information related to patterns of variability (e.g., an instrument may produce more valid scores only in certain in certain settings or samples with specific sociodemographic characteristics) (Hunsley & Mash, 2018; Hunter & Schmidt, 2004).

Validity generalization is included in the Standards for Educational and Psychological Testing as a recommended process to examine validity evidence (AERA et al., 2014). Hunsley and Mash (2018) provided the following guidelines to assess the extent of validity generalization that can be also be used when selecting resilience assessment instruments. *Adequate* evidence of validity generalization would that some evidence (adequate predictive, concurrent, convergent,

and/or discriminative validity) supports the use of a given resilience instrument with either more than one specific type of population *or* across settings. *Good* validity generalization would be that a preponderance of the available evidence that supports the use of a given resilience instrument with either more than one specific type of sample *or* across settings, and to obtain a rating of *excellent*, a preponderance of the available validity evidence would have to support the use of the resilience instrument across samples with a diverse range of sociodemographic parameters and in multiple settings.

It should also be noted that although these ratings describe validity generalization in terms of adequate, good, and excellent, this rating describes the range of previous usage of a measurement tool and not the psychometric quality of the tool. Also, a psychological inventory that is developed for use with a specific population could be rated as inadequate in terms validity generalization if there are not multiple populations or settings in which the instrument would apply (e.g., specialized tools for studying resilience in astronauts or Antarctic personnel). This is not problematic in and of itself, as long as the tool has adequate evidence of prior and present validity for the environment and sample in which is it being used (Rafilson, 1991).

Relevance to the present research. Few resilience studies investigate the convergent validity of resilience constructs. Thus, although analyses of convergent validity are included, the results of these should be interpreted with caution simply due to the limited amount of available data. Predictive validity data are nearly non-existent in the resilience literature, therefore, for present dissertation research, the concurrent validity of resilience instrument scores in relation to specific correlates (i.e., strength of association of between resilience measure scores and observed scores on another instrument or group of instruments, such as measures of anxiety) were summarized. Given sufficient heterogeneity of variance, potential moderator variable such

as age, sex/gender identity, or other sample characteristic such as group membership – healthcare workers, military personnel, university undergraduates can also be useful information for future researchers. To date no validity generalization studies have been conducted for any of the adult resilience instruments that met inclusion criteria for the systematic reviews conducted for this dissertation.

Reliability Generalization. As described previously, reliability is not a property inherent to an instrument, thus it is necessary to consider the reliability of the measure in each administration/research context (Rouse, 2007; Urbina, 2004). Reliability generalization is the generation of a mean estimate of the observed effect sizes across studies, and subsequent analyses that specifically pertain to the investigation of reliability evidence (Borenstein, 2009; Hunter & Schmidt, 2004; Vacha-Hasse, 1998). Vacha-Hasse (1998) proposed reliability generalization, an extension Schmidt et al.'s (1976, 1977) validity generalization, as one method to summarize evidence of instrument reliability across settings, populations, and across other parametric attributes of interest.

Internal consistency, which is most commonly reported using Cronbach's α, is typically included in reliability generalizations because this statistic can be calculated for the observed sets of scores on the majority of self-report instruments (Cronbach, 1951; Urbina, 2004). Other reliabilities may also be summarized using this process (test-retest reliability, inter-rater reliability) using the total number of observations and the calculated variance of the reliability evidence across studies. Similar to validity generalization variance heterogeneity can be assessed for each summary effect estimate of interest. Significant heterogeneity across reported reliability effect sizes can be examined through moderator analyses to determine if potential moderator variables (e.g., participant or study characteristics) can explain the differential performance of an

instrument across contexts (Borenstein et al., 2009; Hunter & Schmidt, 2004; Quintana, 2015; Pigott, 2012; Vacha-Hasse, 1998). Reliability generalization can be of utility to researchers as prior evidence relating to the stability of reliability across studies can elucidate patterns of variability (e.g., an instrument may produce more reliable scores in certain settings or samples with specific sociodemographic characteristics); which can be useful information when selecting research instruments and designing studies (Hunsley & Mash, 2018; Hunter & Schmidt, 2004).

Relevance to the present research. Few resilience studies have investigated the test-retest reliability of resilience instruments. Thus, the few data that were available were insufficient or unsuitable to be combined for meta-analysis. For example, where the duration of time between administrations differs, and for pre- post-intervention studies where change was expected between resilience measure administrations, data cannot be meaningfully combined. Therefore, for the Project 2 meta-analyses, internal consistency was the reliability variable of primary interest and test-retest reliability could not be assessed meta-analytically. Potential moderators of the variation in the reliability coefficients across studies was also examined to determine if the observed variance can be explained by contextual factors or by sample or participant characteristics (e.g., age, geographic location, sample category). To date there are no published reliability generalizations for any of the adult resilience instruments that met the systematic review inclusion criteria (approximately 50 scales).

Moderator Analyses. A fundamental assumption and rationale for conducting meta-analyses of all types of effect sizes (treatment effects sizes, reliability effect sizes, validity effect sizes) is that these will differ from study to study, else why summarize and calculate estimated averages for these values? All observed variation is assumed to have some degree of heterogeneity that is an aggregate of systematic and unsystematic variation (Field, 2005; Howell,

2014). Statistical artefacts such as sampling and measurement error are the predictable, random, within-group unsystematic sources of error variance (Hunter & Schmidt, 2004). Other artefacts are in some way attributable to 'the ecosystem' of all factors and variables that may impact, influence, or confound the outcome of a study (hence, these are the systematic artefacts). Ideally, error variance reflects only true variation – i.e., the variation that is not due to artefacts (Haynes et al., 2011; Hunter & Schmidt, 2004). However, this perfect research scenario (absence of artefacts) is unlikely when it comes to research in the behavioural sciences, as all psychological research data are affected by at least some artefact variables (Borenstein et al., 2009). The objective of research in general is to increase the amount of variability that is explained (Field, 2005). For example, it is of utility to attempt to identify variables that moderate relations when is it observed that scores on an instrument have greater reliability or validity in certain contexts or under certain conditions (Higgins, Thompson, Deeks, & Altman, 2009; Vacha-Hasse, 1998).

Heterogeneity of variance. As described in the previous section, it is important to assess if variation in effect sizes across studies is present, and if a practically relevant proportion of that variation can be subsequently explained by moderator variables. Three statistics are typically used to investigate variance heterogeneity: τ^2 (the standard deviation of the true effect size) estimated using T^2, the Q-statistic (used to assess the statistical significance of T^2), and I^2 (to evaluate the magnitude of heterogeneity across observed effects) (Dersimonian & Laird, 1986; Higgins et al., 2009; Huedo-Medina et al., 2006).

Tau-squared (τ^2). τ^2 is the standard deviation of the true effect size. T^2 is the notation used to denote that τ^2 is an estimate, as τ^2 cannot be known without sampling the entire population, and even then there would be some degree of error. The statistics τ^2/T^2 are used in meta-analyses to represent the between-studies variance for the effect of interest (Borenstein, et

al., 2009). When $\tau^2 = 0$, effects across studies are homogenous (Hunter & Schmidt, 2004).

The Q-statistic. The *Q*-statistic is the sum of squares of the effect of interest where the j^{th} square is weighted by the reciprocal of the estimated effect size variance (Dersimonian & Laird, 1986). The *Q*-statistic follows the Chi-square (χ^2) distribution, therefore can be tested for significance (DerSimonian & Laird, 1986; Huedo-Medina et al., 2006). Rejecting the null hypothesis (H_0: homogeneity of variance) suggests heterogeneity of variance across study effect sizes is present but does not describe the magnitude of the variance heterogeneity (Huedo-Medina et al., 2006). For each set of effect sizes where *Q* is found to be statistically significant (suggesting the presence of heterogeneity of variance for that set of observed effects), the magnitude of heterogeneity is then calculated using the I^2 statistic. For strengths of association where significant effect size variability across studies is found moderator analyses can be conducted to investigate potential sources for the observed variation.

I-squared (I^2). I^2 was proposed to be a proportion (ranging from 0 to 100) to reflect how much of the difference across observed effect sizes reflects true differences as "a signal-to-noise ratio" (Borenstein et al., 2009, p. 117; Higgins et al., 2003). The I^2 statistic is used in meta-analyses to estimate the magnitude of between-studies effect size heterogeneity (Dersimonian & Laird, 1986; Huedo-Medina et al., 2006). Higgins, Thompson, Deeks, and Altman (2003) provided the following interpretation for I^2 values (range 0 – 100%): I^2 of approximately 25% represents low heterogeneity of variance among effect sizes, $I^2 \approx 50\%$ moderate, and $I^2 \approx 75\%$ represents high heterogeneity of variance.

Heterogeneous Variance Result. When significant (and theoretically relevant) variability is found among sets of reliability coefficients, validity evidence, or effect sizes that are observed across studies or across instruments, this variation can sometimes be explained by moderator

variables (Borenstein et al., 2009). Moderator analyses, including both group comparisons and meta-regression, can be used to examine the relation of potential moderators to criteria (Hunter & Schmidt, 2004; Vacha-Hasse, 1998). Group comparison is suitable for potential source variables that are categorical in nature and meta-regression is used where potential moderator variables are continuous in nature (Quintana, 2015; Thompson & Higgins, 2002; Veichtbauer, 2010).

Investigation of publication bias. As the inclusion criteria limit data collection to peer-reviewed journal articles, dissertations, and theses, it is necessary to assess for and provide a discussion of any effects found for publication bias (i.e., the hypothetical impact of non-published studies being excluded from the analyses) (Borenstein et al., 2009). In addition to visually inspecting the distribution of the standard error of the effect sizes that are included in each analysis using a funnel plot, three statistical methods can be used to evaluate potential publication bias in effect sizes such as Cohen's *d* and Pearson's *r*: Rosenthal's Classic Fail-Safe *N*, and Orwin's Fail-Safe *N*, and Egger's *Z* (Egger et al., 1997; Orwin, 1983; Rosenthal, 1979). Examination using all three methods is recommended (Borenstein, 2009; Orwin, 1983; Rosenthal, 1979). Howell and Shield's Fail-Safe *N* was used in place of Orwin's and Rosenthal's Fail-Safe *N* calculations to assess the File Drawer Effect for estimates of internal consistency, as the aforementioned methods are not recommended to assess potential publication bias on reported values of Cronbach's α (Howell and Shields, 2008; Piedmont, 2014).

Rosenthal's Fail-Safe N. In Rosenthal's classic method (Rosenthal, 1979), it is assumed that missing data are missing due to the 'file drawer' effect – i.e., that the effect sizes of unpublished studies were zero. Rosenthal's Fail-Safe *N* yields the number of studies with a zero/null effect size that would be needed for the mean effect size of the meta-analytic summary

to become non-significant. A small number of studies relative to the count of studies included suggests that missing data with detrimental impact to the present analysis, and a large number of studies relative to the number of included studies suggests that unpublished null effect sizes of zero would be inconsequential to the current analysis (i.e., if 50 studies are included, and the result is that 500 unpublished studies with null effect sizes would need to exist for the present observed summary effect size to be impacted by publication bias, then it is assumed that evidence of publication bias to the degree that the present results are impacted is insufficient). Detailed computational formulae for Rosenthal's Fail-Safe *N* are given in Rosenthal, 1979. For the meta-analyses in Project 2 these computations were made using the metafor package for R (Quintana, 2015; Veightbauer, 2010).

Orwin's Fail-Safe N. Orwin's (1983) Fail-Safe Method is another way to examine publication bias. This calculation differs from Rosenthal's (1979) in that it can be defined to be a more conservative estimate. The reason it *can* be, but could also yield very similar results, is that, instead of calculating the number of missing/unpublished studies needed to render the observed summary effect size to be insignificant, the formulae yield the number of missing/unpublished studies needed to reduce the calculated summary effect size to a value that the researcher has defined as inconsequential. This allows for flexibility in the evaluation of publication bias (Orwin, 1983). Detailed computational formulae for Orwin's Fail-Safe *N* are given in Orwin, 1983. For the meta-analyses in Project 2 these computations were made using the metafor package for R (Quintana, 2015; Veightbauer, 2010).

Egger's Z. Further, funnel plots (i.e., figures that illustrate the dispersion of points around a mean) can be used to visually display sources of potential publication biases. The general convention for these plots is that each included effect size for a given meta-analysis is assigned

to the x-axis, and the standard error of the effect size is assigned to the y-axis. When it is estimated that there is little or no publication bias for a given set of effect sizes, data are symmetrically distributed about the mean (the appearance of which is similar to an inverted funnel). When publication bias is likely, data are asymmetrically distributed or non-centrally clustered (Egger, Smith, Schneider, & Minder, 1997). As visual inspection is often misleading and insufficient, funnel plots can be statistically investigated using Egger's *Z*, a regression test for significant distributional asymmetry (Egger et al., 1997). This test, based in linear regression, generates a regression line where potential bias is interpreted as the degree by which the slope of that line significantly deviates from zero (Egger et al., 1997). For the meta-analyses in Project 2 these computations were made using the metafor package for R (Quintana, 2015; Veightbauer, 2010).

Howell and Shields Fail-Safe N. As reliability coefficients (e.g., Cronbach's α) are estimates of variance (the amount of true score variance when items on a measure are correlated with the overall measure score) and in this regard effect sizes such as Pearson's *r* which is a strength of association calculation (and is the square root of the estimated explained variation), different computations are needed to assess publication bias for Reliability Generalization analyses (Howell & Shields, 2008; Piedmont, 2014). Differing in specific computation, but similar in function to other Fail-Safe *N*, this formula estimated the possible influence that "file-drawer" effect studies may have on the mean reliability values calculated in meta-analyses of the reliability coefficient Cronbach's α. Similar to Orwin's (1989) formulae, a theoretically-based acceptable target threshold can be selected such that the Fail-Safe *N* statistic yields the number of unpublished studies that would be needed reduce what Howell and Shields refer to as "a plausible worst-case reliability coefficient of .80 standard deviations below the reporting

samples' mean" (Howell & Shields, 2008, p. 123).

Previous Systematic Review of Resilience Measures

Although resilience measures have been in use since 1953, only a few systematic reviews of resilience measures have been published. This section provides a summary of the available reviews of resilience measurement instruments.

Review 1: Wagnild and Young (1991). The earliest measure review was by Wagnild and Young (1991) and described five measures. Three were measures of adult general resilience were described: The Hardiness Inventory 60-item battery of five measures (Kobasa 1979), Hardiness Inventory 36-item version (citation not given, measure currently n/a) and the Hardiness Inventory, 50-item version (citation not given, measure currently n/a).

The versions of Hardiness Inventory, including the number of items and theoretical base for each instrument and three studies that applied the Hardiness Inventory, were described. Observed scores on the Hardiness Inventory were reported to correlate with positively with social support (r = .13), challenge (r = .23), and health status (r = .43), the latter representing a relation of higher hardiness relating to better overall physical health. The original Hardiness Inventory was reported to positively correlate with burnout (r = .43), and a revised version was reported to negatively correlate with the same construct (r = -.39).

Review 2: O'Neal (1999). The second review, conducted by O'Neal (1999), described seven scales. Five of these were adult general resilience measures, the Cognitive Hardiness Scale (CHS: Nowack, 1989), Psychological Hardiness Scale (PHS: Younkin, 1992); Personal Views Survey III (PVS-III: Hardiness Institute, 1985, as cited in O'Neal 1999), Resiliency Scale (RS: Jew, Green, & Kroger, 1999) and the Resilience Scale, 25-item version (RS-25: Wagnild & Young, 1993).The remaining instruments measured context specific resilience scale, and group

resilience, one each. O'Neal reported that a search of the PsycINFO database from 1996 to 1999 found 600 references to the word resilience, and that for the prior 12 years, 1984 to 1995, 68 references were found, indicative of the growth of interest and publications related to the construct of resilience in 1999. It is unclear how the measures included in the review were selected from the 68 identified references. The author described that the review did not provide a complete listing of all instruments available for assessing the resilience and was not inclusive of all of the available reliability and validity data for each instrument. It was stated that the objective of the review was to provide a general overview of resilience and how it can be measured. The review provided a conceptual and theoretical overview of each of the seven instruments, the factors that comprise each instrument, examples of instrument administrations, and author-selected psychometric information available for each scale.

Review 3: Ahern, Kiehl, Sole, and Byers (2006). The third review of instruments to measure resilience was conducted by Ahern, Kiehl, Sole, and Byers (2006). The review included six resilience measures. Five of these instruments were for the measurement of resilience in adults, the Baruth Protective Factors Inventory (BPFI: Baruth & Carroll, 2002), Brief-Resilient Coping Scale (BRCS: Sinclair & Wallston, 2004), Connor & Davidson Resilience Scale (CD-RISC-25: Connor & Davidson, 2003) Resilience Scale 25-item version (RS-25: Wagnild & Young, 1993), and the Resilience Scale for Adults 37-item version (RSA-37: Friborg et al., 2003).

The remaining instrument measured resilience in adolescents. A systematic search of six databases (CINAHL, Pre-CINAHL, Academic Search Premier, PsycINFO, PsycARTICLES and MEDLINE) was conducted. The keyword search used the terms *Resilience* and *Scales* or *Instruments*, with a limiter for the English language and humans. The search yielded 370 articles

and 32 were selected for detailed review, and 23 were retained, and six resilience measurement instruments were found.

The review included an overview of each instrument, including the theoretical base, factor structure, and scoring procedure for the measure. Frequency data were given and provided a count of how many times each instrument was used across the 23 selected articles. The review included validation information available from the instrument developers. Psychometric quality was given descriptively (e.g., *content validity was described as established by a panel of experts*) and psychometric quality ratings were described in a binary format (e.g., convergent validity and discriminant validity were reported as *present* vs. *not present*). Known population applications were listed for each instrument. Finally, each instrument was rated on a scale of 1 to 3 "based on their credibility to study resilience in adolescents" (Ahern, et al., 2006, p. 121) using an author-derived evaluation sale. The RS received a score of 3 (best), and the RSA, ARS, and CD-RISC received a score of 2 (acceptable), and the remaining measures received a score of 1 and it was noted that a score on 1 indicated there was insufficient data to assess these measures.

Review 4: Windle, Bennett, and Noyes (2011). The fourth review of resilience measurement scales was conducted by Windle, Bennet, and Noyes (2011). Nineteen resilience measures were included. Twelve of these measures were general measures of adult resilience, and seven were for use with adolescents. The adult measures that were included were: the Brief Resilience Scale (BRS: Smith et al., 2008), Connor & Davidson Resilience Scale 10-item version (CD-RISC-10: Campbell-Sills & Stein, 2007), Connor & Davidson Resilience Scale 25-item version (CD-RISC-25: Connor & Davidson, 2003), Dispositional Resilience Scale 15-item version (DRS-15: Bartone 1995), Dispositional Resilience Scale 30-item version (DRS-30: Bartone 1991), Dispositional Resilience Scale 45-item version (DRS-45: Bartone 1989), Ego-

Resiliency (ER: Klohnen, 1996), Ego-Resilience 14-item measure (ER-89: Block & Kremen, 1996), Psychological Resilience (PR: Windle, Markland, & Woods, 2008), Resilience Scale 25-item version (RS-25: Wagnild & Young, 1993) Resilience Scale for Adults 33-item version (RSA-33: Friborg, Barlaug, Martinussen, Rosenvinge, & Hjemdal, 2005), and the Resilience Scale for Adults 37-item version (RSA-37: Friborg et al., 2003).

Eight databases (Applied Social Sciences Index and Abstracts, Medline, PsycINFO, Social Sciences Citation Index, Science Citation Index, Arts and Humanities Citation Index, Greenfile, and the Cochrane Database of Systematic Reviews) were searched. A combined definition field and keyword search strategy was used. The definition search term used was resilienc* and this was paired with using an AND operator with keywords biol*, geog*, community, Interven*, determin*, relat*, predict*, review, questionnaire, assess*, scale, and instrument. A Google search was also used with terms *resilience scale*, *resilience questionnaire*, *resilience assessment* and *resilience instrument*. The search yielded 2,979 articles and the 271 included articles spanned two decades (1989 – 2009). Seventeen measurement scales were identified, and fifteen (plus four revised versions) were retained, thus 19 measures reviewed. No detail was given related to the two excluded instruments.

The review provided a general overview of each measure and the theoretical base underlying each scale. Where data were available, the authors described how the test developers created and selected items for each scale. The number of items and number factor structure of each scale was given. Frequency of use and measure application were not described. Windle et al. (2011) rated each measure on content validity, internal consistency, criterion validity, construct validity, reproducibility agreement, test-retest reliability, responsiveness, floor/ceiling effects, and interpretability, using an author-derived rating scale. The authors used this scale to

assign quality ratings (range 0-2 per property) for each of the properties listed above. A score of 0 indicated that no information was available and a rating of two indicated the highest quality mark. The operational definitions varied per dimension. For example, for internal consistency, a rating of 0 = no information was found, or that Cronbach's α was either below 0.70 or exceeded 0.95; a score of 1 = "No factor analysis or doubtful design or method" (Windle et al., 2011, p. 3), and a score of 2 indicated "Factor analyses were performed on adequate sample size (7 times the number of items) and [sample size was] >100, and Cronbach's alpha(s) are between .70 and .95" (Windle et al., 2011, p. 3). For construct validity, 0 = no information was available, 1 = less than 75% of hypotheses were confirmed, and 2 = "specific hypotheses were formulated and at least 75% of results support those hypotheses" (Windle et al., 2011, p. 3). The remainder of the item definitions are similar where 0 = not available and 2 = an author-defined benchmark for the highest quality. Across the nine dimensions the highest score given to any instrument was 7, and this was given to the two versions of the RSA, the CD-RISC-25, and the BRS.

Review 5: Smith-Osborne and Bolton (2013). The fifth review of resilience measures was an investigation of measures that could be used across the lifespan and was conducted by Smith-Osborne and Bolton (2013). Their review included 10 measures. Four of the instruments were developed for use with children and adolescents. The general measures of resilience for use with adults that were included were the Baruth Protective Factors Inventory (BPFI: Baruth & Carroll, 2002), Brief-Resilient Coping Scale (BRCS: Sinclair & Wallston, 2004), Connor & Davidson Resilience Scale 25-item version (CD-RISC-25: Connor & Davidson, 2003), Resilience Scale (RS: Wagnild & Young, 1993), and the Resilience Scale for Adults 33-item version (RSA-33: Friborg et al., 2005).

The remaining measure was a context-specific adult measure to assess only mid-life

resilience. A search was conducted using 19 databases (a full listing is available in Smith-Osborne & Bolton, 1993); using the keywords *resilience instruments, resilience, resiliency, resilience measures, protective mechanisms,* and *scale validation*. The search yielded 874 articles after duplicates were removed – 106 instruments were identified and 96 were excluded due to lacking validation, lacking peer reviewed, the authors' categorization as not useful for practice, or where operational definition of resilience differed from conceptualization as a personality trait related to anomalous positive outcomes. Thus, 10 instruments were included. The definitions for the categories above were not provided.

The review provided an overall description of each instrument, including validation information given by the instrument developers. Frequency of measure application is not described, however known uses and a few administration examples are given for each measure. The review provided information related to the cost of each scale and where to obtain each measure. The authors described that quality assessments were conducted, and each instrument was assigned "+", "–", or "n/a" across 12 dimensions (representative sample spectrum, reference standard (norms), time for adversity change during testing limited, partial verification, differential verification, incorporation, test review, diagnostic review, clinical review, withdrawals, and sponsoring precluded. Operational definitions of these dimensions are not provided in the review: it is not entirely clear which of the methods described was used or how the properties were assessed. It is, however, described that the 10 selected instruments are those that the authors concluded sufficiently meet an adequate quantity of these criteria. Details and procedures related to instruments that were not selected are not given.

Review 6: Pangallo, Zibarras, Lewis, and Flaxman (2015). The sixth review of resilience instruments was conducted by Pangallo, Zibarras, Lewis, and Flaxman (2015) and

reviewed 17 resilience measures they included. The described that the purpose of the review was to view "resilience through the lens of interactionism" (Pangallo et al., 2015, p. 1). Eleven of the measures were instruments assess general resilience in adults: the Baruth Protective Factors Inventory (BPFI: Baruth & Carroll, 2002) Brief-Resilient Coping Scale (BRCS: Sinclair & Wallston, 2004), Brief Resilience Scale (BRS: Smith et al., 2008), Connor & Davidson Resilience Scale 10-item version (CD-RISC-10: Campbell-Sills & Stein, 2007), Connor & Davidson Resilience Scale 25-item version (CD-RISC-25: Connor & Davidson, 2003), Ego-Resilience, 14-item version (ER-89: Block & Kremen, 1996), Ego-Resilience 10-item version (ER-89-R: Allesandri, Vecchione, Caprara, & Letzring, 2012), Psychological Resilience (PR: Windle, Markland, & Woods, 2008), Personal Views Survey-III-R (PVS-III-R: Maddi, Harvey, Khoshaba, Lu, Persico, & Brow, 2006), Resilience Scale 25-item version (RS-25: Wagnild & Young, 1993), and the Resilience Scale for Adults 33-item version (RSA-33: Friborg et al., 2005).

The remaining measures included in the review assessed context-specific resilience. A systematic search of article titles across of ten databases used the following Boolean terms: *resilien** AND *questionnaire* OR *scale* OR *instrument* OR *measure*, NOT *youth* OR *child* or *adolescent*, and yielded 263 articles. Pangallo et al. described that hardiness measured were excluded, as they conceptualized hardiness as a personality trait and resilience as a dynamic variable, thus measures of trait resilience were excluded. These criteria appear to be contradictory to the inclusion some instruments, as Maddi's (2006) PVS-III-R hardiness scale and other measures that defined resilience as a personality trait were included.

The review described the theoretical base for each measure and the samples that were used by the measure developers to initially test each instrument. Internal consistency estimates

reported by the instrument developers were also provided, as well as the number of items and the factors that comprised each inventory. This review also included a thematic analysis and quality assessment scores for each instrument. Quality was defined as concordance with the author-defined themes of interactionism. A value of one was given for each theme that was present in the instrument. Themes that were assessed were: adaptability, self-efficacy, active coping, positive emotions, mastery, hardiness, supportive relationships, and structured environment. An additional theme, conceptual adequacy was rated as: *minimal*, *partial*, and *yes*. The operational definitions for these categories are not provided.

The second study in the review used the same selection of measures and a second novel assessment of quality rating was used. Pangallo et al. used some of the definitions and items given in Windle et al. (2011) and altered other items and their rating criteria. Similar to Windle et al. (2011), scores ranged from 0-2. The dimensions evaluated were: Evidence based on test content, internal consistency, stability (test-retest reliability), replicability (defined as exploratory analysis followed by confirmatory factor analysis), discriminant evidence, convergent evidence, and the extent of measurement application (defined as >12 studies, 5-12 studies, and <5 studies). The authors did not describe how the extent of measure application was assessed (no additional systematic searches were described, and data was not given for frequency of use by measure). The measure receiving the highest total rating for psychometric properties was the PCQ: Psychological Capital Questionnaire (Luthans, Youssef, & Avolio, 2007). The PCQ measures psychological capital using four factors: hope, optimism, self-efficacy, and resilience. The second, third, and fourth place measures were the RSA, BRS, and CD-RISC-25, respectively.

Review 7: Salisu and Hashim (2017). The most recent review of resilience measures was conducted by Salisu and Hashim (2017). The review described 26 resilience instruments

(inclusive of child, adolescent, and context-specific measures). The English-language general adult resilience measures that were identified were: the Baruth Protective Factors Inventory (BPFI: Baruth & Carroll, 2002), Brief-Resilient Coping Scale (BRCS: Sinclair & Wallston, 2004), Brief Resilience Scale (BRS: Smith et al., 2008), Connor & Davidson Resilience Scale, 2-item version (CD-RISC-2: Vaishnavi, Connor, & Davidson, 2007), Connor & Davidson Resilience Scale 10-item version (CD-RISC-10: Campbell-Sills & Stein, 2007), Connor & Davidson Resilience Scale 25-item version (CD-RISC-25: Connor & Davidson, 2003), Ego-Resiliency (ER: Klohnen, 1996). Ego-Resilience 14-item version (ER-89: Block & Kremen, 1996), Ego-Resilience 10-item version (ER-89-R: Allesandri et al., 2012), Psychological Resilience (PR: Windle et al., 2008), Resiliency Scale (RS: Jew et al., 1999), Resilience Scale (RS-25: Wagnild & Young, 1993), and the Resilience Scale for Adults 33-item version (RSA-33: Friborg et al., 2005).

Four instruments were selected for review: The Brief Resilience Scale (BRS: Smith et al., 2008), Connor & Davidson Resilience Scale, 10-item version (CD-RISC-10: Campbell-Sills & Stein, 2007), Connor & Davidson Resilience Scale (CD-RISC-25: Connor & Davidson, 2003), Resilience Scale for Adults 33-item version and the (RSA-33: Friborg et al., 2005).

The method/procedure that was used to develop the list of 26 measures was not described. The selection process to choose the four measures that were reviewed was unclear. It was stated that the choice was based on the Windle et al. (2011) review, using their top ranked instruments (although the selections were inconsistent with the top ranked instruments in the referenced review), and based upon frequency of instrument use. A process to determine frequency of use and use data was not provided or described for the measures.

The review included a brief description of each of the four measures. The number of

items, factor structure, and a few examples of test use were given for each measure. Internal consistency was described for two of the measures. The authors concluded that the CD-RISC-10 items scale was the most widely used instrument to study resilience. Salisu and Hashim also concluded: "This paper also discovered that [the] majority of studies that used [the] CD-RISC were conducted in medical and/or disaster studies. It is therefore imperative to use this scale to assess the resilience of individuals in other fields such as entrepreneurship and general management" (Salisu & Hashim, 2017, p. 30). A systematic review procedure and quantitative data were not provided in support of this concluding statement.

Summarization and Need for the Present Review. As outlined in the previous paragraphs, to date, seven reviews of resilience measures have been conducted. The largest of these identified 13 instruments to measure general resilience in adults and discussed four of these in detail (Salisu & Hashim, 2017). The second largest, included 11 measures of adult general resilience (Pangallo et al., 2015). The most recent review that used systematic article methods and that also provided the frequency of resilience measure use in the collected articles was in 2006. Ahern et al. (2006) provided frequency of use information for 23 instrument applications and included five adult general resilience that met the inclusion criteria used in the systematic review conducted for this dissertation. This review was also the largest in terms of examining measure use.

As the number of available instruments to assess adult characterological resilience exceeds the number discussed in any prior review, the planned systematic reviews that comprise Project 1 provide more information related to measure use than the existing reviews on the topic. Further, the plan to include a summary of the frequency of measure use based upon a large-scale systematic review of five resilience constructs sets this project apart from previous reviews, as

this has not been an objective of the previous reviews and the provision this information is likely to be useful to others who wish to study and measure resilience.

Overview of the Current Studies

In the previous sections I outlined that (a) a systematic review inclusive of measure use frequency, detail related to the populations being studied with resilience measures is needed, and (b) there are no published reliability or validity generalization studies on adult general resilience measures, inclusive of the investigation of potential moderator variables. Therefore, the dissertation includes two projects as outlined in the Overview of the Dissertation section.

Project 1

Resilience Keeps Bouncing Back:

A Systematic Review of Characterological Resilience Measures in Adults

Corresponding author:

Teresa A. Allan

Co-Authors:

John Hunsley

William Lawrence

University of Ottawa

May 17, 2021

Abstract

In this pair of systematic review studies, self-report measures of adult characterological resilience were identified and summarized. In Study 1, 51 personality-based instruments across five personality-based resilience constructs (ego-strength, grit, hardiness, mental toughness, and resilience) were identified. Information for each measure inclusive of measure title, construct, item count, factors, response scale and items, scoring method, and score range was summarized. In Study 2, 1,322 articles were identified that described 1,193 studies and 1,351 participant samples who completed between one and four of these resilience measures. Measure use frequency and contextual use information related to study and sample attributes were summarized. Based upon these reviews, the challenges of resilience measure use are discussed and observations/recommendations are provided.

Keywords:

Resilience, Hardiness, Ego Strength, Mental Toughness, Grit

Introduction

Resilience, "bouncing back," returning to a homeostatic state, or failing to be perturbed from that state when presented with adversity and challenge is not a new idea. Derivatives of the Latin terms *resilio, reslui, resilire,* and *resiliens* came into use in the English language in the mid-1620s (Resilience, 2018; Resilio, 2018). At that time, resilience was used to describe the propensity of an object to return to its original shape after an impact or to resist being disrupted altogether. Metallurgists and physicists use the term resilience interchangeably with the elasticity modulus of an alloy or composite material to describe the stress and force required before that material will break, bend, or experience permanent deformation from its original state (Young & Budynas, 2002). Essentially, examining the property of resilience is to ask: *how much stress can this object endure?* As stressors underlie many psychological symptoms and difficulties, it became of interest to the psychological scientists of the early 20th century to ask: *how much stress can this human endure before becoming substantially distressed?* and importantly, *how can we identify individuals with characteristics that will be more resistant to, or more tolerant of stress?* Given the value of identifying individuals who would be most likely to withstand stressors related to a wide range of professions, circumstances, or events, it is not surprising that a number of resilience constructs, models, and related measures for resilience were developed.

The first mention of a resilience-like personality construct occurred early in the last century. Human mettle – a "vigor and strength of spirit or temperament" (Mettle, 2018, p. 1) when facing adversity or challenge – was deemed a personality trait in the 1920's and was first quantitatively measured as ego-strength (Barron, 1953). Although interest in the measurement of trait resilience has waxed and waned over the last 70 years, it seems that novel resilience constructs emerge in the psychological literature every couple of decades with many of the same

key characteristics. Chronologically, the constructs used to assess and describe resiliency have been: ego-strength, hardiness, mental toughness, grit, and resilience (Barron; 1953; Connor & Davidson, 2003; Duckworth, Peterson, Matthews, & Kelly, 2007; Epstein, 1983; Freud, 1927; Kobasa, 1977; Strycharczyk & Clough, 2015).

Psychodynamic theorists, beginning in the 1920s, described a human personality characteristic called *ego-strength* which was hypothesized to be higher in individuals who tended to adapt, solve problems, and live relatively average lives after experiencing events that were expected to be psychologically disruptive (Freud, 1927; Nunberg, 1942). Thus, it was deemed to be of value to identify and measure ego-strength, and scales to assess ego-strength were first developed in the 1950s (Barron, 1953; Block & Kremen, 1996; Gottesman, 1959; Klohnen, 1996). In the 1970s, the construct of *hardiness*, an ability to withstand stressors, "weather the storm," and/or achieve successful adaptation to change, was introduced (Kobasa, 1977) and, subsequently, a number of quantitative measures to assess hardiness were developed (Funk, 1992; Kobasa, 1977; Maddi & Khoshaba, 1994). Sports psychologists of the 1980s and 1990s described *mental toughness* and the military psychologists of the new millennium described *grit*; both constructs are described as determination, adaptation, and perseverance in the face of difficulty (Duckworth et al., 2007; Duckworth & Quinn, 2009; Matthews, 2008; Sheard, Golby, & van Wersch, 2009; Strycharczyk, 2017). Personality researchers and scholars leading the positive psychology movement in the 1990s used the term *resilient* to describe individuals who "bounce back" from adversity or thrive exceptionally in distressing or challenging contexts (Bartone, Ursano, Wright, & Ingraham, 1989; Connor & Davidson, 2003; Seligman & Csikszentmihalyi, 2000; Wagnild & Young, 1993).

All of the aforementioned characterological constructs and the instruments used to

measure them share common elements and overlapping definitional themes. The individual high in ego-strength, who is resilient, hardy, mentally tough, or gritty is typically conceptualized as adaptive, resourceful, unshaken, confident in his/her ability to persevere in order to obtain an unexpected positive outcome "against the odds." Therefore, in order to be succinct, these instruments to assess any quantify any of these overlapping constructs are referred to *resilience measures*.

Although it is beyond the scope of the present systematic reviews of characterological resilience measures to go into detail, it is relevant to note that generalized resilience/personality models are not the only conceptualizations of psychological resilience found in the research literature. Resilience has also been conceptualized as a contextual strength, in terms of an individual's resistance to the development of specific psychological symptoms or specific undesirable outcomes (Benishek, Feldman, Shipon, & Mecham, 2005; King, King, Vogt, Knight, & Samper, 2003). Such models conceptualize resilience as a differential function or ratio comparing the sum of an accumulation of events or risk factors (e.g., exposure to active military combat, exposure to a natural disaster, personal injury) to the sum representing a set of protective factors (e.g., strong social supports, financial resources, years of formal education) (Windle, 2011). Although risk factors and protective factors can include personality traits, most also include specific experiences that place an individual "at risk" and include external resources that protect the individual from the undesired outcome. For example, the Deployment Risk and Resilience Inventory (DRRI) is used to calculate a resilience value by comparing a list of pre-defined potentially stressful combat experiences with a list of pre-deployment risk and protective factors to understand how resources and supports help to reduce post-traumatic stress symptoms under the conditions and stressors related to combat and other military operations (King et al.,

2003). Other examples of the measurement of contextually specific resilience include the Suicide Resilience Inventory (Rutter, Freedenthal, & Osman, 2008), and the Resilience at Work Scale (Winwood, Colon, & McEwen, 2013). A number of domain-specific resilience scales have also been developed, including measures of academic hardiness, health-related hardiness, mental toughness in professional athletes, resilience to LGBTQIA+ discrimination, and other context- or population-specific domains (e.g., Benishek, Feldman, Shipon, & Mecham, 2005; Pollock & Duffy, 1990; Sheard et al., 2009; Smith & Gray, 2009).

The inherent difficulties related to generating operational definitions for resilience constructs, the nature of resilience measurement challenges, and the merits and concerns of adopting various resilience models have been presented in depth (e.g., Canavan, 2008; Funk, 1992; O'Neal, 1999; Windle, 2011). Similarly, issues around the differentiation between resilience constructs have also been described (e.g., Maddi, Erwin, Carmody, Villarreal, White & Gunderson, 2013; Wagnild & Young, 1991). Given the numerous potentially overlapping constructs and the proliferation of instruments to assess general resilience from a characterological perspective, it can be a challenge for researchers to select appropriate, evidence-based measures for their specific research questions, resilience-building program evaluations, or clinical assessment applications.

Reviews of resilience measure reviews have examined from six to 19 different measures, including general resilience measures for use across the lifespan, domain-specific measures, and measures in which the resilience construct is viewed as involving risk factors, protective factors, and exposure to a specific set of events (Ahern, Kiehl, Sole, & Byers, 2006; O'Neal, 1999; Pangallo, Zibarras, Lewis, & Flaxman, 2015; Smith-Osborne & Bolton, 2013; Windle, et al., 2011). In the largest review, Windle et al. (2011) examined 19 resilience measures, 12 of which

were measures for adults that had some validation evidence and one that had not yet been psychometrically evaluated. However, with the recent surge of interest in positive psychology concepts and the introduction of new constructs that seem to share definitional attributes with resiliency (i.e., grit and mental toughness), there are far more measures than those reviewed by Windle et al., many of which have at least some psychometric evidence to support their use. Accordingly, as there is no comprehensive review of these numerous characterological resilience measures, we undertook a systematic review of the published literature on these measures in order to provide summary information about the frequency of each measure's use and details on the target populations and methodological characteristics associated with each measure.

The Present Studies

In this research, an answer was sought for the following research questions: what self-report instruments are available for the measurement of general (i.e., not context-specific) characterological resilience in adults? (Study 1), and what is the frequency and context of use for each identified measure? (Study 2). The purpose of Study 1 was to identify characterological measures of resilience in adults that met an *a priori* defined set of inclusion criteria. The results of Study 1 describe for each identified measure the instruments associated resilience construct, number of items, the factor structure and list of factor(s) that comprise each instrument, number of items loading to each factor, response scale type, response items, scoring range, and key references. The purpose of Study 2 was to determine the contexts and frequency with which the measures identified in Study 1 have been applied in original empirical studies and to generate a contextual use summary for a subset of the most frequently applied measures of adult characterologically-based resilience. The results of Study 2 include the frequency of resilience measure(s) administered, geographical location of study participants, participant population (e.g.,

university students, military veterans), details of the research methods (number of populations, number of conditions), and the chronometric research methodology that was used.

Our focus is on filling a gap in the resilience literature by summarizing how frequently the measures of general characterological resilience (i.e., non-contextually specific measures) have been used, where, with whom, and with what research design. Fundamental tenets of research and scientific investigation are to work from what is known to what has not yet been investigated and to understand the evidence that informs our current hypotheses and assumptions. The overall utility of this project is to add to the larger picture of adult characterologically based resilience measurement by creating a summary of what personality-based measures are available, how the measures are structured, and how frequently and with whom each has been used. A discussion of considerations related to generalization from these data and the psychometric pitfalls of resilience measurement (and how to avoid some of these) is also provided.

Study 1

The purpose of Study 1 was to investigate what self-report instruments are available for the measurement of general (not context-specific) characterological resilience in adults and to generate a list of these measures. A secondary purpose was to summarize for each measure: (a) the instrument name and abbreviation, (b) the construct measured, (c) the number of items comprising the measure, (d) the nature of factors/instrument subscales, (e) the type of response scale, (f) response scale items, (g) scoring method, and (h) instrument development references.

Materials and Methods.

Literature search. A systematic database search consistent with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Guidelines was conducted (Moher,

Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009; Moher, Shamseer, Clarke, Ghersi, Liberati, Petticrew, Shekelle, & Stewart, 2015). Five databases (PsycINFO, Medline, Eric, CINAHL, and Health and Psychosocial Instruments) were used to identify adult personality-based measures of resilience. The following Boolean algorithm search of article titles was used:

```
ti("Resilienc*" OR "Hardiness OR Hardy" OR "Grit" OR "Ego
strength OR Ego-strength OR Ego resilienc* OR Ego-resilience" OR
"Psychological Strength OR Psychological Fortitude" OR "Mental
Toughness")

AND

ti("Measur*" OR "Inventor*" OR "Scale" OR "Questionnaire" OR
"Assess*" OR "Quanti*" OR "Test" OR "Psychomet*" OR "Survey"
```

The rationale for use of a search algorithm was to streamline the search process and ensure consistency across databases. The use of the `AND` separator between the algorithmic statements served to focus the search toward relevant articles (e.g., an article titled "Resilience Scale" or "Measuring Hardiness" would be selected by this search) and reduce the yield of irrelevant articles (e.g., using this method titles such as "Resilience of Memory," "Hardiness of Lima Beans in Peru," "Hardy-Weinberg equilibrium" would not be included). Further search limiters were applied, where possible, in each database to limit articles to publication in English, with human participants, and in peer-reviewed journal articles. No date limiters were applied to this search.

Inclusion/exclusion criteria. Resilience measures located by this search were included if the following eight criteria operationally defined in this section were met:

(1) Generalized characterological resilience: Only measures of pervasive/generalized and trait-based resilience were included. Thus context-specific, non-characterological, and

non-generalized resilience measures were excluded.

(2) Individual resilience: Only measures of individual resilience were included. Thus, measures for group resilience (i.e., family, hospital, or city) resilience were excluded.

(3) Adults: Only instruments designed for use with individuals 18+ years of age were included.

(4) Language: Only measures available in the English language were included.

(5) Self-Report: Only self-report measures were included. Thus, observer-report resilience instruments were excluded.

(6) Number of items: Measures comprised of fewer than three items were excluded. The rationale for this criterion is that it is a common recommendation for factor analyses that a minimum of three items should load onto a factor in order for that factor to be a viable descriptor or predictor variable (Costello & Osbourne, 2005; Credé, Harms, Niehorster, & Gaye-Valentine, 2012; Tabachnick & Fidell, 2007).

(7) Objective quantitative scoring: Only measures with clearly defined quantitative scoring procedures were included. Projective measures with multiple or individualized scoring procedures (e.g., Rorschach Inkblot Test, Thematic Apperception Test) and qualitative measures (content analyses, non-structured and semi-structured interview processes to assess resilience) were excluded.

(8) Supplementary information: The search yielded a sizeable list of resilience instruments but, in many cases, the retrieved articles did not provide all of the details required to determine if a measure met inclusion criteria. In these cases, additional literature searches were conducted to obtain the instrument development and/or validation evidence documents. In some cases, where this information was not freely available, test

developers were contacted for details and test administration manuals.

Study 1 Results. The set of five searches yielded 731 documents (PsycINFO, *n* = 187; MEDLINE, *n* = 215; ERIC, *n* = 55; CINAHL, *n* = 85; Health and Psychosocial Instruments, *n* = 189) (Figure 1). Following the removal of 283 duplicates, 478 articles were reviewed. Of these, 182 articles described, reviewed, or empirically applied English-language, quantitatively summative, self-report measures of adult resilience derived from various models of personality theory. Contextual measures were used or described in 42 articles (health recovery/health habits, *n* = 11; mental toughness specific to sport performance, *n* = 10; academic hardiness or resilience, *n* = 6; suicide resilience, *n* = 4; vocational resilience; *n* = 3; acculturation / immigration resilience, *n* = 2; LGBTQIA+ population resilience, *n* = 2; college student resilience *n* = 1; depression resilience, *n* = 1; event-specific resilience, *n* = 1; and hardiness in divorce, *n* = 1). General resilience measures were described or applied in 140 articles. From these peer-reviewed empirical studies, conceptual/theoretical formulation articles, and review articles, a list of 47 adult resilience measures meeting the inclusion criteria outlined in the previous section. Table 1 is a list of these measures and 4 additional relevant instruments found in Study 2 (described in section 4.2). The table provides the additional details outlined above: measure name, abbreviation, construct that is measured, number of items, list of factors or dimensionality of the instrument, response scale type and response items, scoring method, and instrument development references. The frequency of use to measure resilience in adults and application contexts of these measures were the focus of Study 2.

Of the 51 measures, based on the descriptions in the measure development documents, the constructs measured were: hardiness (*n* = 20 instruments), resilience (*n* = 13), ego strength (*n* = 7), ego resilience (*n* = 3), both hardiness and resilience in a single instrument (*n* = 3), and grit

(n = 2). The instrument with the fewest items was the Brief Resilient Coping Scale (BRCS-4: Sinclair & Wallston, 2004), with four items to measure resilience; the measure with the greatest number of items was the Hardisurvey III-R, with 106 items to measure hardiness (Hardisurvey III-R: Maddi & Khoshaba, 2001; The Hardiness Institute, 2020). The mean number of items per instrument was 30, and the modal number of items was 20. The majority of instrument development documents (96%, n = 49) described the proposed or hypothesized structure of instrument factors. Of these, 17 instruments were single-factor/unidimensional measures, and 32 instruments were comprised of two or more factors that could be scored in addition to an overall composite score. Of the multi-factorial instruments, the number of factors proposed by the test developers were: 2 factors (n = 5), 3 factors (n = 19), 4 factors (n = 2), 5 factors (n = 4), 6 factors (n = 1), and 8 factors (n =1). The most frequent factor structure among this set of instruments was the "3C's" model of hardiness which posits that hardiness has three orthogonal factors of commitment, control, and challenge.

Study 2

To investigate the frequency and context of use for the measures identified in Study 1, a series of systematic citation searches was conducted to assemble a collection of peer-reviewed journal articles spanning the time period from the earliest resilience measure's publication date (Barron Ego-Strength scale for the Minnesota Multiphasic Personality Inventory; Barron, 1953), in 1953 to June 30, 2019. Two raters (the first and third authors) participated in the coding of each measure that met inclusion criteria (outlined in greater detail in the following sections). The contextual information for each study of each included article was recorded: (a) resilience measure(s) administered, (b) geographical location of study participants (coded by country), (c) participant population (e.g., university students, military veterans), (d) details of the research

methods (number of sampled populations, number of conditions), and (e) chronometric methodology that was used (e.g., number of times resilience data was collected from participants in the context of the selected research method).

Materials and Methods.

Literature search. Citation searching has been found to be an effective and efficient method to obtain a collection of documents that cite a single root document or collection of root documents (Jascó, 2004). Therefore, a Web of Science citation search was conducted for each source publication or developer's document where the resilience measure was first discussed or made available for use. For example, Connor and Davidson (2003) first described the 25-item Connor-Davidson Resilience Scale in the journal article titled *Development of a New Resilience Scale: The Connor-Davidson Resilience Scale (CD-RISC)* in 2003.

For resilience measures where no single measure development document was available (*n* = 2) or could not be identified due to ongoing development, multiple instrument versions, or instrument title changes, an "all-fields" search strategy was used and applied to a combined search of three databases (PsycINFO, Medline, and Health and Psychosocial Instruments). The syntax for each search included variants of the measure name found in Study 1 combined with the first author's surname to capture as many uses of the resilience measure as possible. For example, the syntax of the search that was used to capture applications of the original and revised versions of Maddi's Personal Views Survey (PVS) was:

```
af(Personal Views Survey) AND af(Maddi)) OR (af(PVS*) AND
af(Maddi)) OR af(PVS-II) OR af(PVS-III) or af(PVS-IIIR OR PVS-
III-R)
```

with added database limiters to select for peer-reviewed journal articles and English-language publications. This strategy was necessary for two sets of instruments: the PVS series of instruments by Maddi and colleagues and the hardiness instruments originally described by

Kobasa (1977) that underwent instrument title changes and subsequent item revision as the newly identified construct of hardiness was being explored and further refined.

Article inclusion and exclusion criteria. For inclusion, at least one resilience instrument meeting inclusion criteria for Study 1 must have been administered to adult participants in at least one study described in the article. The measure must have been administered in full to study participants (i.e., all items, with no item omission/addition), unaltered (i.e., no rewording of items or response options, no changes in response scales or scoring procedures). Thus, articles were excluded if resilience instruments were not used with participants (e.g., theoretical/conceptual papers) or if no analyses of primary data were conducted (e.g., systematic reviews, meta-analyses). In addition, the following four inclusion criteria were applied to screen the articles collected in Study 2; a flowchart of the criteria application is presented in Figure 2.

(1) Adult participants: Only articles where the mean age of study participants was greater than or equal to 18 years were included. Articles with studies that did not report the mean age of participants were excluded if the participants were described as "children," "adolescents," or primary/middle/high school students. Articles with studies that did not report the mean age of participants, but an age of 18+ years could be reasonably inferred from clear semantic identifiers were included. For example, participants described as "American military combat veterans" (given age requirements for military enrollment in the United States) were included. Similarly, studies that described participants as "older adults," "successfully aging," or "retirees" were included as these descriptors suggest it is reasonable to assume participants were 18 years or older. Studies that did not report the mean age of participants and did not use clear semantic identifiers of adulthood were excluded.

(2) Language: Only articles where resilience instruments were administered to all participants in the English language were included, as this was the language in which most of the included measures were originally developed. Two measures, the Resilience Scale for Adults 37-item version (RSA-37: Friborg, Hjemdal, Rosenvinge, & Martinussen, 2003) and the 33-item version (RSA-33: Friborg, Barlaug, Martinussen, Rosenvinge, & Hjemdal, 2005) were initially developed in Norwegian; the Tachikawa Resilience Scale (TRS-10: Nishi, Uehara, Yoshikawa, Sato, Ito, & Matsuoka, 2013) was initially developed in Japanese. The original test developers of these measures also provided versions of these instruments in English (Friborg et al., 2003; Friborg et al., 2005; Nishi et al., 2013); thus, only administrations of the instrument in the English language were included. If the language of administration was not clearly stated, we coded the study as administered in the official language of the region of the participants. For example, if a study that met all other inclusion criteria described the collection of data from university students in the midwestern United States, and did not state the language of administration, the study was assumed to have been administered in the official language of that region (English) and was included. Similarly, if a study that was conducted in Madrid, Spain did not state the language of administration, it was assumed to have been conducted in the official language of that country (Spanish) and excluded.

(3) Self-Report: Only articles where study participants completed self-report measures resilience were included. Thus, articles containing observer-report applications of resilience instruments that were designed to be self-report measure were excluded.

(4) Sample size $n > 1$: Articles reporting single case studies were excluded.

Coding procedure. Data collected for this systematic review were organized into three levels: (a) article level attributes (such as title, year of publication, etc.), (b) study level attributes, as each article may describe more than one study, and (c) sample level attributes, as each study may include data from more than one sample (e.g., military veterans, undergraduate students, individuals living in areas impacted by natural disaster). Further, some articles had data and participant overlap wherein a study had been described in more than one included article. This occurred in a number of large-scale studies (e.g., studies of U.S. military veterans, longitudinal studies). Identifying overlap was critical to avoid redundancy and prevent inflated measure use counts. For example, information from a single large study described in 10 articles should only be included once because it was, in fact, one study and treating this as 10 separate studies would yield inflated and inaccurate measure use frequencies. Therefore, when a given study was described in several articles, only the article containing the largest sample was used for data extraction purposes.

Variables for Study 2 that were imported directly from the search results from the Web of Science database to the article spreadsheet were article title, year of publication, article author(s), and journal title. The remainder of the variables were generated by the first author making a "first pass" of coding on the first 500 included articles. From this, 11 sets of binary categorical variables were generated and used to describe study participant populations (Appendix A: Coding Manual). A second pass of coding was completed by the first author which consisted of coding these variables for both the first 500 articles and the remaining 822 articles. Every rater-coded attribute was coded in binary, where 1 = the variable applies to the article, and 0 = the variable does not apply to the article. For example, if a study collected data from a sample of

university psychology students, the variable was coded as `UNI_Psychology_Students = 1`. Similarly, if the same study also selected only female students the variable `SGSO_FemalesONLY = 1`. All variables were coded into two spreadsheets (studies and samples) by the two raters in accordance with an *a priori* established coding manual (Appendix A: Coding Manual). Rater-coded variable sets for each study were resilience measure(s) used, count of resilience instruments administered in each study, study location (country) and the study method structure. To summarize the study method structure, we coded the number of resilience measure data collection timepoints and the arrangement of these (e.g. $Time1_{BASELINE}$-Intervention-$Time2_{POST}$), number of unique populations from which sample data were collected, and number of study conditions. For example, a cross-sectional correlational study with one group of participants was coded as one condition, and a study utilizing a between groups design where one group received Treatment A and one group received Treatment B was coded as two conditions.

Inter-Rater Reliability. Rater 1 (the first author) coded all variables for 100% of n = 1,322 articles, and Rater 2 (the third author) coded 10% of the articles (n = 132 articles). Each article was assigned an identification number and a random number generator (in R) was used to select the articles to be coded by Rater 2 using the article identification numbers. These data were coded using two spreadsheets: attributes applicable to each study (n = 115 binary variables) and attributes applicable to each sample (n = 73 binary variables). Inter-rater reliability was assessed for each set of coded variables. Cohen's κ was calculated for each variable, and these values were summarized for each set of variables with κ_{POOLED} (Cohen, 1960; De Vries, Elliott, Kanouse, & Teleki, 2008; Sim & Wright, 2008). Although it is fairly common practice that κ is summarized across variables by averaging Cohen's κ (thus, $\bar{\kappa}$), Devries et al. found that $\bar{\kappa}$ produces an inflated estimate of inter-rater reliability when there is variability across Cohen's κ

values, and a modified calculation κ_{POOLED} produced an attenuated (i.e., less biased) estimate of inter-rater reliability (De Vries et al., 2008; McHugh, 2012). Devries et al. (2008) also found no detriment to the application of κ_{POOLED} in data with little variability. Thus, this more flexible statistic was selected, as it would be suitable for both data with little variation and data with substantial variation.

Study 2 Results.

In the process of coding the included articles, two additional versions of two of the instruments that were included in Study 1 were found. These were a 30-item version of the Cognitive Hardiness Scale (*n* = 12 articles) (CHS-30: Nowack, 1989) and both a 24- and 27-item version of the Thomas-Zander Ego Strength Scale (*n* = 1 article each) (TZ-ESS-24: Crandall, as cited in Sadowski, Woodward, Davis & Elsbury, 1983; TS-ESS-27: Zander & Thomas, 1960, unpublished paper, as cited in Martin & Grater, 1993). It was also noticed that there are two instruments with the same name (20-item Short Hardiness Scale). One was developed by Kobasa, Maddi, & Kahn (1982) and the other was developed by McNeill, Kozma, Stones, and Hannah. (1986); therefore the latter instrument (SHS-20-McNeil: McNeill et al., 1986) was included as a separate measure. As described previously, these four measures were added to the instruments included in Study 1. Citation searches were conducted for these measures using the same date range (1953 to June 2019) as the searches for each of the other instruments identified in Study 1, and these additional instrument uses have been included in the data summaries and analyses that follow.

The citation search procedure for Study 2 yielded 7,349 full-text peer-reviewed journal articles. Duplicates (*n* = 1,757) were removed and (*n* =5,596) unique articles were screened for inclusion (Figure 2). Abstract screening was not sufficient to determine if inclusion criteria were

met as abstracts typically do not contain the required level of detail, thus complete articles were screened. Screening for inclusion was conducted primarily by the first author as the inclusion/exclusion criteria were objective and defined *a priori* (e.g., there is little ambiguity in assessing if data were collected, if a specific resilience instrument was used in the study, if the mean age of participants was greater than or equal to 18 years where a mean age was given). However, in instances where the first author concluded that was unclear if the article met inclusion criteria, the article was reviewed by the second rater and a consensus was reached. The number of unique articles that were retained after this screening was n = 1,322. Three hundred articles contained an identified data repetition/overlap across one or more papers; thus the 1,322 included articles were comprised of 1,193 unique studies in which data were collected with one or more measures of characterological general resilience. The 1,193 unique studies examined 1,351 participant samples (section 4.2.4). Measure use across studies by construct, country, population category, study design and method, instrument, and contextual details for the most frequently used resilience instruments are described in the sections to follow.

Inter-rater reliability. Inter-rater reliability was assessed for the data from 132 articles (188 variables / 24,606 binary data points) using the procedure and test statistics previously described. Inter-rater reliability for the variables related to study attributes ranged from κ = .992 to 1.000 for each variable, κ_{POOLED} = .999. Inter-rater reliability for the variables related to sample attributes ranged from κ = .971 to 1.000, κ_{POOLED} = .998. (Appendix: Coding Manual contains a list of these variables).

Resilience measure use by construct. The 51 resiliency measures were used in 1,239 instances in 1,193 studies to collect data across 5 resilience constructs (Figure 3). The most frequently measured construct was resilience (n = 650 studies, 11 instruments). The second most

measured construct was ego-strength/ego-resilience (*n* = 291 studies, 9 instruments), followed by grit (*n* = 130 studies, 2 instruments), hardiness (*n* = 77 studies, 18 instruments), hardiness and resilience together on a single instrument (*n* = 72 studies, 3 instruments), and mental toughness (*n* = 31 studies, 2 instruments). The two constructs measured most frequently in the past decade (2010 – 2019) were resilience, measured in *n* = 601 studies; and grit, measured in 128 studies (Figure 4).

Resilience measure use by country. Resilience measures were used in studies conducted in 14 different countries, and in studies conducted internationally and online. The countries in which most of these studies were conducted were the United States (*n* = 826), Australia (*n* = 105), the United Kingdom (*n* = 82), and Canada (*n* = 74). One study was conducted across 59 countries, and 15 included participants in two countries. The greatest frequency of cross-country collaboration, (*n* = 8) were between Canada and the Unites States. The use frequency for all other countries included in this sample was fewer than 20 resilience studies per location (Figure 5). Twenty-three studies collected resilience measure data from online participants using platforms such as Amazon's Mechanical Turk, Qualtrics, and Facebook, and did not report the location of participants or whether participation was limited to a specified region. There were also nine studies where data was collected in person, and the location of participants was not stated and could not be determined.

Resilience measure use by sample category. Sample categories were coded based upon the selection/recruitment criteria of study participants as described by the authors of each study. In the sample of 1,193 studies described in previous sections, there were 1,351 populations (unique samples) of participants.

Based upon authors' descriptions of study participants the following participant categories were established: university students, (n = 380 samples of students currently attending university, and 13 samples that included both current students and other members of their campus community such as alumni, faculty, other campus staff), individuals selected specifically for their identification as a specific sex, gender identity, or sexual orientation (n = 217 samples), individuals selected on the basis of mental health characteristics (e.g., treatments, diagnoses, or symptoms; n = 201 samples), individuals selected for their specific vocation/profession (n = 195 samples), individuals selected on the basis of physical health characteristics (e.g., medical treatments, diagnoses, or symptoms; n = 186 samples), military personnel (n = 132 samples), individuals who reported experiencing events often associated with trauma (e.g., hurricane or military combat; n = 77) , caregivers (n = 63 samples), individuals selected for their self-identification with a specific ethnicity or cultural group (n = 57 samples), individuals selected for other specified attributes (samples of individuals identified as senior citizens, sports participants, couples, forensic settings, twins, etc., n = 149 samples), and individuals not selected for a specific parameter, experience, or attribute/convenience samples not collected in any of the other contexts (n = 98 samples) (Table 2).

Most participant samples (n = 977, 81.2%) were coded to a single category (e.g., university students, military veterans, mental health hospital inpatients, sample of participants identifying with a specific ethnicity). There were 374 unique samples that were coded to two categories (e.g., university students + individuals identifying as female; residents of a region impacted by a natural disaster + mental health characteristics). Seventy-four samples were coded to three categories, 17 samples were coded to four categories, and one sample was coded to five categories).

Resilience measure use by design and method. Most of the studies (*n* = 998, 83.7%) collected data from participants at a single time point. Of these, the majority (*n* = 875) used a single group correlational design (one group, in one condition, at one time). The remainder applied a between-groups design and collected data at a single time from two samples (*n* = 69), three samples (*n* = 68), four samples (*n* = 2), or five samples (*n* = 2). Forty-one studies collected resilience data from participants at a single point in time across two or more conditions. The next most frequently used study design was to collect resilience data at two timepoints, where the first measurement ($T1_{BASE}$) is taken at a pre-intervention or baseline time point, followed by an intervention, with the subsequent measurement of resilience occurring directly following the intervention ($T2_{POST}$) (*n* = 69). The next two most frequently applied design methodologies (both used in *n* = 35 studies) were the repeated measures design (T1 followed by T2 with no experimental or manipulated intervention, e.g., test-retest reliability studies, measures of student's resilience at the beginning and end of a given academic period) and the $T1_{BASE}$→intervention→$T2_{POST}$→$T3_{FOLLOW\text{-}UP}$ design. All other studies in our sample that measured resilience at multiple timepoints (*n* = 42) measured participants a minimum of two times and a maximum of six times. Across these studies there were 24 measurement time variations in which of an initial measurement of resilience was followed by repeated measurement at additional times during the course of an event/intervention and follow-up time points post-event or post-intervention.

Resilience measure use by instrument. It was found that 1,193 studies collected resilience data using the measures identified in Study 1 (Figure 6). The earliest use found was in 1957, with the Barron Ego-Strength Scale 68-item version. Thus, the timeframe covered in this review is January 1957 through the end date for our data collection in June 2019. The majority,

91.7% of studies, used one resilience instrument (*n* = 1,145). Two resilience instruments were used in 41 studies (3%). Three resilience instruments were used in four studies, and four resilience instruments were used in three studies.

There were six instruments that were not used in any of the studies in our retained sample of studies (Ego Strength Scale 40-item version; Milner & Wimberly, 1979; HardiSurvey III-R 65-item version: Maddi, 2002; Lang and Goulet Hardiness Scale 45-item version: Lang, Goulet, & Amsel, 2003; Me-B-Tough 43-item version; Mack, Ragan, & Mack, 2008; Resilience Scale for Adults 37-item semantic differential response version; Friborg et al., 2005, Friborg, Martinussen, Rosenvinge, 2006; Tachikawa Resilience Scale 10-item English version: Nishi, et al., 2013). Although articles that cited these six measures were found, in many instances, articles were non-empirical, or the measure was not used in the research reported in the article. In other instances, the measures were used in the research, but the studies were not included in our sample because the measure administration context did not meet inclusion criteria.

Frequently used resilience measures and use contexts. Although 51 characterological resilience measures were identified in Study 1, 64.5% of the retained studies (769 of 1,193 studies) involved the use of one or more of the following five measures: the Connor-Davidson Resilience Scale, 25-item version (CDRISC-25: Connor & Davidson, 2003) *n* = 236; the Barron Ego-Strength Scale, 68-item version (Barron-68: Barron, 1953), *n* = 198); the Brief Resilience Scale, 6-item version (BRS-6: Smith, Dalen, Wiggins, Tooley, Christopher, & Bernard, 2008), *n* = 135); the Connor-Davidson Resilience Scale, 10-item version (CDRISC-10: Campbell-Sills & Stein, 2007), *n* = 117); the Short Grit Scale, 8-item version (Grit-8: Duckworth & Quinn, 2009), *n* = 83) (Figure 6). Each of these, excepting the Barron-68, increased in use in the more recent

decades (Figure 7). Given the apparent "popularity" of these 5 measures, we next provide summary information on each of the measures.

Connor-Davidson Resilience Scale, 25-item version. Connor and Davidson (2003) developed this 25-item scale to assess resilience. Participants rate each item (e.g., *I am able to adapt when changes occur, I can deal with whatever comes my way*) on a 5-point scale (0 = not true at all, 1 = rarely true, 2 = sometimes true, 3 = often true, 4 = true nearly all of the time) in accord with their feelings over the past month. Total resilience scores on this instrument are the summation of all individual item scores across the five factors, and can range from 0 to 100; with higher scores interpreted as representing greater individual resilience. Further aspects of the individual's resilience profile can be obtained by examining the five factors that comprise the scale (Factor 1: personal competence, standards, and tenacity, Factor 2: instincts, tolerance of negative affect, and effect of stress, Factor 3: attitudes toward change and security of relationships, Factor 4: control, Factor 5: Spirituality. These factors consist of 8, 7, 5, 3, and 2 items, respectively (Connor & Davidson, 2003).

The CDRISC-25 was used in 236 studies published between 2005 and 2019; 157 studies involved data from participants in the United States, 33 studies with participants in Australia, 12 studies with participants in Canada, 12 studies with participants in the United Kingdom, seven with participants in South Africa, four studies in with participants New Zealand, and one study each with participants in Malaysia and Ireland. Two studies were international collaborations (Australia and the United Kingdom, Australia and New Zealand). The location of participants was not described in 5 studies. The CDRISC-25 was administered most frequently to university students (n = 50 samples), individuals selected on the basis of mental health characteristics (e.g., treatments, diagnoses, or symptoms; n = 45 samples), individuals selected on the basis of

physical health characteristics (e.g., medical treatments, diagnoses, or symptoms; *n* = 42 samples), military personnel (*n* = 38 samples), and individuals who report experiencing events often associated with high levels of stress or trauma (*n* = 36 samples).

Barron Ego-Strength Scale, 68-item version. Barron's 68-item ego-strength scale (Barron-68: Barron, 1953) is both a subscale of the original Minnesota Multiphasic Personality Test (MMPI: Hathaway & McKinley, 1942) and a stand-alone measure for ego strength using 68 of the MMPI's (true/false) items (e.g., *I have often been frightened in the middle of the night, I do many things which I regret afterwards*). Scoring is the same as with the other MMPI subscales, as raw scores are compared to a set of standardized norms for a subset of similar individuals and scaled using the standard *T* distribution (where the distribution mean = 50 and scores in the clinical range are those at or above 70). The raw score range for this instrument is from 0 to 68; higher scores on this scale are interpreted as indicating less ego strength (Barron, 1953; Hathaway & McKinley, 1942).

The Barron-68 was used in 198 studies published between 1957 and 2007 with 145 of these studies published between and 1957 and 1980. Most of these studies involved data from participants in the United States (*n* = 177 studies); there were 15 studies with participants in Canada, four studies with participants in the United Kingdom, one study with participants in Australia, and the region of data collection could not be determined for one study. It was not possible to identify the region of the participants in the remaining study. As might be expected, the most frequent applications of this scale occurred in mental health settings. Data were collected from (*n* = 116 samples) individuals selected on the basis of mental health characteristics (e.g., treatments, diagnoses, or symptoms) and 56 of these samples were hospital inpatients. The second most frequent sample type were current university students (*n* = 72

samples). This instrument was also used to measure ego strength in individuals selected on the basis of physical health characteristics (*n* = 31), military personnel (*n* = 17), characteristics related to profession/vocation (*n* = 17), couples (*n* = 7), caregivers (*n* = 5), convenience samples (*n* = 6) or samples selected for other characteristics. Approximately half (51.7%) of all studies that used this measure assessed participants who identified as male only.

Brief Resilience Scale. The Brief Resilience Scale was developed by Smith et al. in 2008. This measure has six items, three of which are positively worded, and three of which are negatively worded and reverse-scored (e.g., *I tend to bounce back quickly after hard times, I have a hard time making it through stressful events*). Participants completing this measure rate each item on a 5-point Likert scale where 1= strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, and 5 = strongly agree. Smith et al. (2008), using principle component analyses on the data yielded from four participant samples, found that in each sample all items loaded onto a single factor with loadings ranging from .67 to .91. The scale instructions do not indicate a specific time frame that the study participant is to reflect upon in or refer to when responding. Scores are summed and the range of possible scores for this instrument is 5 to 30. Higher scores on this scale are interpreted to indicate greater individual resilience (Smith et al., 2008).

Use of the BRS-6 in our sample (*n* = 135 studies) occurred from the publication date of the BRS-6 by Smith et al., in 2008 to 2019. Most of these studies involved data from participants in the United States (*n* = 82 studies); there were 22 studies with participants in Australia, 16 studies with participants in the United Kingdom or Ireland, four studies with participants in Canada, four studies with participants in New Zealand, and one study each in Jamaica and South Africa. There was one study that was a multinational collaboration, and four studies where the region of the participant could not be identified as these were either not stated, on conducted

online and not limited to a specific region. Data were most frequently collected from university students (*n* = 30 samples), individuals that were selected for their specific vocation/profession (*n* = 27 samples), individuals selected on the basis of physical health characteristics (e.g., medical treatments, diagnoses, or symptoms; *n* = 27 samples), convenience samples of participants who were not selected for any specific parameter or attribute (*n* = 15 samples), and individuals who self-identify as members of the LGBTQIA+ community (*n* = 8 samples).

Connor-Davidson Resilience Scale, 10-item version. Campbell-Sills and Stein (2007) used exploratory and confirmatory factor analyses to develop this 10-item, unidimensional/single-factor version of the 25-item Connor-Davidson Resilience Scale. Response options, sample items, and scoring are similar to process described above for the CDRISC-25, except that the score range for the instrument is 0-40, given fewer items (Campbell-Sills & Stein, 2007).

Use of the CDRISC-10 in this sample (*n* = 117) occurred from the publication date of the CDRISC-10 by Campbell-Sills and Stein, in 2007 to 2019. Most of these studies involved data from participants in the United States (*n* = 79 studies), there were 19 studies with participants Australia, seven with participants in the United Kingdom, four with participants in Canada, two with participants in Singapore, and one each in Antarctica, New Zealand, and Nigeria. There were two studies that were international collaborations (Canada and the United States, Australia and Canada). The region of the participants was not given and could not be identified in one study. Data were collected most frequently from individuals selected on the basis of physical health characteristics (e.g., medical treatments, diagnoses, or symptoms; *n* = 36 samples), university students (*n* = 25 samples), individuals who were specifically for their vocation/profession (*n* = 19 samples), and military personnel (*n* = 13 samples).

Short Grit Scale, 8-item version. The Short Grit Scale, developed by Duckworth and Quinn (2009) is comprised of a subset of eight items presented in The Grit Scale also referenced as the Grit-O or Original 12-item Grit Scale (Duckworth et al., 2007). This instrument is comprised of two factors (Consistency of Interest and Perseverance of Effort) that each contain four items. Consistency of Interest items are negatively worded and reverse-scored (e.g., *I often set a goal but later choose to pursue a different one.*) Perseverance of Effort items are positively worded (e.g., *Setbacks don't discourage me*). Participants rate each item on this scale on a 5-point scale 1 = Not at all like me to 5 = Very much like me. Participant scores are summed and scores range between 8 and 40, with higher scores to be indicative of higher levels of individual grit (Duckworth et al., 2007; Duckworth & Quinn, 2009).

Use of the Grit-8 in this sample (n = 83) occurred from the publication date of the measure Duckworth & Quinn in 2009 to 2019. There were 65 studies with participants in United States, eight with participants in Canada, two with participants in the United Kingdom, one with participants in Australia, and one study that was a collaboration between researchers in Canada and the United States. There were six studies conducted online through services such as Amazon's Mechanical Turk or Qualtrics where authors did not identify the location of the participants or indicate that participation was limited to any specific regions. Data were most frequently collected from university students (n = 43 samples), from convenience samples of participants who were not selected for any specific parameter or attribute (n = 26 samples), individuals who were selected for their specific vocation/profession (n = 7 samples), and individuals who self-identify as belonging to a specific ethnicity, cultural group, or race (n = 5 samples).

It should be noted, however, that in 2021, Duckworth, Quinn, and Tsukayama published a corrigendum indicating that "It has since been pointed out, correctly, that the specified model cannot be distinguished from a model with two correlated subfactors and no higher-order factor" and that "we acknowledge the mathematical equivalence of these specifications and our error in interpretation. However, we also take the opportunity to admit a more profound correction. It is now clear to us that statistical answers cannot definitively settle certain theoretical riddles, and our argument for grit as a compound of related but distinct dispositions should not have relied so heavily on the optimal factor solution for a questionnaire devised to assess it" (Duckworth, Quinn, & Tsukayama, 2021, p. 573). The authors conclude that their previously proposed structure has mathematical limitations, that statistics were relied upon too heavily, and that the true factorial structure of grit warrants further study. This commentary paper was published after articles were collected for this sample; thus, for the purposes of this study, the proposed two-factor structure of grit was used (as the new information would not have been available between 2009 and 2016).

The use of altered versions of resilience measures. A number of the excluded articles (n = 190) used resilience measures that were (a) abbreviations of the original version (n = 140), (b) versions with altered items (n = 39), or (c) abbreviated versions that also contained altered items (n = 11). Of these articles, the majority (n = 140) were published between 2009 and 2019.

With respect to abbreviated versions, 15 studies involved the use of one or more instrument subscale, with all of the items included from the subscale(s). Conversely, in the majority of the studies that involved the use of abbreviated scales (n = 125), items included in the measure did not reflect the content of complete subscales from the original instrument. The main rationales provided to justify the removal or alteration of item were the need to (a) reduce

instrument administration time and/or burden to participants or (b) eliminate items that did not fit the researchers' research focus, seemed outdated or poorly worded, could be confusing to participants, or did not fit the cultural context of the investigation; in many instances, no specific rationale was given. The use of abbreviated measures occurred with 21 instruments that, in their original versions, had between 4 and 68 items. The scales that were most frequently abbreviated were CD-RISC-25 (n = 29), RS-25 (n = 26), Barron-68 (n = 17), Grit-8 (n = 12), and the BRS-6 (n = 8).

Discussion

The measurement of resilience and the conceptually similar constructs of ego-strength, hardiness, grit, and mental toughness has occurred in the psychological literature for more than six decades. As described earlier, a number of differing approaches to measuring resilience have been proposed. We chose to focus this study on general adult characterological measures as these can be used in a broad range of settings and with many different participant samples. For resilience researchers is it essential to understand the theoretical and structural underpinnings of the available instruments (i.e., which resilience construct is measured, with how many items, with what response options, loading onto which of how many subfactors). It is also important to know the characteristics of study participants with whom the measures have been used, as the data from these samples have contributed to the current knowledge base on resilience. Further, for resilience researchers designing new studies, it is crucial to understand the extent to which a given measure has been used in their research context of interest—is the measure "tried and true" or is the use of this measure with the intended population "breaking new ground?"

Thus, the primary foci of our two systematic review studies were to generate a list of self-report instruments available for the measurement of general (i.e., not context-specific)

characterological resilience in English-speaking adults (Study 1) and to summarize the frequency and context of use for each identified measure (Study 2). The results of these studies are discussed in the sections to follow. In these, we also highlight some cautions to be considered related to generalization from studies concentrated in specific contexts, sample types, or study designs. Incidental observations related to studies that used partial measures or altered resilience instruments are also discussed. Finally, we conclude with a description of the limitations of the present research and closing thoughts.

Study 1. Hundreds of resilience measures, both specific and general, with and without a characterological focus, have been developed. In Study 1, limiting our search to only English-language measures of general (i.e., noncontextually specific), characterologically based resilience measures for adults yielded a set of 51 resilience instruments. The constructs measured by these instruments were hardiness (n = 20 instruments), resilience (n = 13), ego strength (n = 7), ego resilience (n = 3), both hardiness and resilience (n = 3), and grit (n = 2), with measures ranging in length from four items to 106 items. More recently developed and freely available measures, compared to earlier measures, tended to be shorter in length and unidimensional. Developers of measures with fewer items frequently indicated that shorter measures are useful to reduce participant burden and the time needed to collect data. Table 1 provides information on (a) the instrument name and abbreviation, (b) the construct measured, (c) the number of items comprising the measure, (d) the nature of factors/instrument subscales, (e) the type of response scale, (f) response scale items, (g) scoring method, and (h) instrument development reference.

Study 2. Although valuable as an overview of the available measures, the information in Table 1 does not describe the broader contexts in which these measures have been used; this was the focus of Study 2. Details on the frequency of measure use in published research and the

application contexts of these measures were drawn from 1,322 empirical articles containing 1,193 unique studies and comprised of 1,351 samples that met our *a priori* inclusion criteria. The following sections describe the results of our analyses for the variables we extracted and coded from these articles.

Resilience measure use by construct. The greatest growth in the study of this set of resilience constructs has occurred in the last decade, with the greatest increase occurring for the constructs of resilience and grit (Figures 3 and 4). Across 1,193 studies resiliency measures were used in 1,239 instances. The most frequently measured construct was resilience (n = 650 studies, 11 instruments), followed by grit (n = 130 studies, 2 instruments), hardiness (n = 77 studies, 18 instruments), hardiness and resilience together on a single instrument (n = 72 studies, 3 instruments), and mental toughness (n = 31 studies, 2 instruments). Thus, the number of measures available to assess a given construct is unrelated to the number of published studies that have measured the construct.

Resilience measure use by country. Awareness of the location in which studies occurred is a necessary consideration for interpreting the body of published research and constructing research hypotheses based on results of previous studies. Of the 1,193 retained studies, most identified the region where the study was conducted (Figure 5). Measures of resilience and similar constructs were administered in the English language in 14 countries, with the research use of resilience measures concentrated in four countries: the United States (75.6%), Australia (9.1%), the United Kingdom (7.7%), and Canada (7.5%). All other countries in our sample comprised less than 2% each of the total geographic distribution of study location.

There were 32 studies where the origin of the participants was not stated and could not be inferred from the information given. Of these, 23 were described as conducted online using data

collection services such as Amazon's Mechanical Turk, Qualtrics, or Facebook. Online studies are becoming more prevalent and have certain benefits (low cost, access to participants in regions the researcher may not be able to access in person, etc.). Opening participation to all persons with Internet access can be an effective way to increase sample diversity and reach rural and understudied populations; however, for these designs the collection and reporting of demographic information inclusive of where the study took place is necessary for the results of study to be meaningfully interpreted in terms of generalizability and representativeness of the sample.

Resilience measure use by population category. Measure use by sample characteristics was summarized using 11 categories and 71 subcategories derived from the information study authors used to describe their research participants (Table 2). From these data we can conclude that resilience measures are being used with participants with a breadth of characteristics. The most frequent participant category were studies that assessed resilience in university students (29.1% of studies; of these studies, 96.7% included only students, and the remainder were samples that included both students and other members of campus communities such as faculty, administrators, and other employees). The next most common categories involved participants selected for characteristics related to mental health (symptoms, diagnoses, participation in treatment; 14.8%), specific vocations (14.4%; with the most common professions selected for study to be nurses and other healthcare workers in 75 studies and first responders in 32 studies), characteristics related to physical health (symptoms, medical diagnoses, participation in treatment; 13.4%), and affiliation with the military (9.8%; approximately half of these—66 studies—measured resilience in veterans). There were also 98 studies where the only criterion for participation was being at least 18 years of age.

The finding that almost one-third of studies used university student samples is particularly noteworthy, especially as the majority of these studies (n = 284) had as participants current students attending American universities. In other words, the results of nearly one-third of published research on adult characterological resilience is based on data from participants who are young adults, enrolled in an American college or university, and in sufficiently good health to take part in higher education and participate in research. It is not inherently problematic to study resilience in university students, but it is essential to consider the generalizability of these findings given that student samples are not representative of the general American population in terms of age and socioeconomic status. For example, 37% of Americans have completed undergraduate degrees and approximately 9% have completed graduate degrees (US Census Bureau, 2021). Thus, the results of studies where data were collected from undergraduate students may generalize to this subset of the population, but is not likely to generalize to the remaining 63% of the American population. Moreover, the extent to which these data are generalizable to other countries is far from clear. It is particularly relevant to consider the results of such studies in terms of resilience theory and how resilience is defined. If researchers define resilience as thriving "against all odds" or "bouncing back from adversity," the assumption is that resilient individuals have experienced extraordinary difficulty or challenge, or have failed to become distressed in situations that would typically cause high levels of stress. It is possible that the results of resilience research with, for example, combat veterans or survivors of natural disasters may have similar correlates to what is found in research on undergraduate students. On the other hand, it is entirely possible that the extent of resilience and the pattern of associations with psychological variables may differ considerably across this range of participant samples.

In addition to noting that some types of samples were highly prevalent and over-represented in comparison to reported population demographics, we found some participant categories to be under-represented. Very few studies included participants who were in treatment for substance use or reported substance-use related difficulties (2.8%); fewer yet included forensic/justice-involved participants (.01%), participants dealing with homelessness (.01%), socioeconomically disadvantaged populations/individuals living in poverty (.01%), or individuals who reported having experienced assault or abuse (.01%). In comparison to demographic data available for Americans, each of these sample categories is more prevalent in the population than is represented here. For example, it is estimated that ongoing homelessness is experienced by 0.17% , and temporary homelessness is experienced by 2.24% of Americans (HUD: Housing and Urban Development Exchange, 2020; NLCHP: National Law Center on Poverty and Homelessness, 2018). Similarly it is estimated that violent assaults are experienced by 2.4% of the American Population (Department of Justice, 2019). This victimization estimate does not include non-reported crimes and crimes that may traumatic that did not include a component of violence (e.g., a robbery without injury, or home invasion without use of force). Each of these sample categories is a context where research could be beneficial to the planning of programs, services, and supports for underserved populations.

Resilience measure use by design and method. Of the 1,193 studies in our sample, 83.7% collected data from participants at a single time point, and 87.7% of these used a single group correlational design (i.e., one group in one condition at one time). Given that personality models of resiliency are based upon the hypothesis that resilience is a relatively stable trait, is seems reasonable that resilience researchers would choose to measure resilience at a single timepoint when using instruments with previous evidence of reliability and validity in the

population of interest. As with all cross-sectional research, however, it is important to recognize that we cannot form conclusions related to causality based on associations found at a single point in time.

The remaining proportion (16.3%) of studies measured resilience between two and six times. In studies where interventions were provided to participants, in most instances the aim of the intervention was to enhance participants' resilience, or to determine if there was a change in resilience that occurred over the course of a period of time (e.g., before and after military training, before and after attending a specific university course). However, in multiple timepoint designs, the relation between resilience theory and resilience measurement is complex and, in some instances, potentially paradoxical. For example, in studies where a resilience-building intervention is evaluated, researchers often use a pre- and post-intervention design. It is common that a personality-based resilience measure is administered at $T1_{BASE}$ (e.g., before a week-long resilience-building workshop begins) and then a second time at $T2_{POST}$ and/or $T3_{FOLLOW\text{-}UP}$ (e.g., at the end of the workshop or in a time-delayed follow-up). However, if the researcher has defined resilience *a priori* as a personality trait, which is conceptualized as a relatively consistent and unchanging quantity, a hypothesis that states a relatively brief resilience-building intervention will produce a meaningful and enduring change in personality would appear to contradict the theoretical conceptualization that an individual's resiliency is relatively static. Moreover, if a personality-based measure of resilience is used, how likely is the measure to (a) have sensitivity to detect the effects of a training course or intervention and (b) be able to detect changes in trait resilience shortly after the intervention? Approaching this from a different perspective, if resilience change scores are evident after brief or time-limited workshops and interventions, and sustained for some time, what does this mean for resilience theories and the

conceptualization of resilience? Finally, if sustained resilience change is possible from brief interventions, does this have implications for the stability of findings obtained from cross-sectional studies, especially those that examine resilience in one group in one condition at one time?

In time, as more studies with varying measurement structures are conducted, researchers may be able to investigate some of these questions meta-analytically. However, in the meantime, it may be worthwhile for researchers to consider repeated-measures/longitudinal designs that allow for an evaluation of personality-based resilience over time to observe changes. Such designs could add much to our knowledge of resilience and alterations in resilience, as they could include both instances where it is assumed that resilience will not change and instances where it is assumed that it will.

Resilience measure use by instrument. In our review of publications from 1957 to mid-2019, limited to adults, the most frequently used characterological resilience instrument for adults in English was the resilience measure Connor-Davidson Resilience Scale (CD-RICS-25), with 25 items and five factors (n = 236). The next was the Barron Ego-Strength scale for the MMPI, a unidimensional measure with 68 items (n = 198). Following this were the Brief Resilience Scale (BRS-6; n = 136) and a shortened version of the Connor-Davidson Resilience Scale (CD-RISC-10; n = 117), both unidimensional measures of resilience with six and 10 items, respectively. The fifth most applied instrument was the Short Grit Scale (Grit-8), a measure of grit with eight items and two factors. The 6th, 7th, and 8th most frequently used measures were: the Resilience Scale, 25-item version (RS-25: Wagnild, 2009), a resilience scale comprised of 25 items and two factors; the ER-89, a unidimensional ego-resiliency measure with 14 items; and the Original Grit Scale (Grit-12) which has two factors and 12 items. The remaining 39 measures

in our review were each used less than 50 times in the research literature, and 28 of these measures (i.e., 54.9% of the measures identified in Study 1) were found to have been used in five or fewer studies. From this, we can conclude there is great disparity in the amount of psychometric evidence that are available for each measure – ranging from no additional uses that met our inclusion criteria following the original publication containing the instrument (n = 6 instruments), to instruments that have been used in hundreds of studies and have normative data available for multiple samples across numerous contexts.

More than half of the studies included in our review (64.5%, n = 769) used one or more of the five most frequently applied measures to collect resilience data from participants. Four of the five most frequently applied measures (CD-RISC-25, BRS-6, CD-RISC-10, and Grit-8) were developed between the years 2000 and 2009 and the use of all of these measures increased substantially in the 2010 to 2019 period (Figure 7). As noted above, the number of measures available to assess a resilience construct is not related to the prevalence of the assessment of that construct in the literature. Of the five most used measures (CD-RISC-25, Barron-68, BRS-6, CD-RISC-10, and Grit-8), three were designed to measure resilience (CD-RISC-25, BRS-6, CD-RISC-10), one to measure ego-strength (Barron-68), and one to measure grit (Grit-8). Similarly, the most prevalent instrument factor structure across all measures included in our review was one that emphasized commitment, challenge, and control; however, three of the five most used instruments are unidimensional (Barron-68, BRS-6, CD-RISC-10), the Grit-8/Short Grit Scale is comprised of two factors consistency of interest and perseverance of effort, and the CD-RISC-25 uses a 5-factor structure (Factor 1: personal competence, standards, and tenacity, Factor 2: instincts, tolerance of negative affect, and effect of stress, Factor 3: attitudes toward change and security of relationships, Factor 4: control, Factor 5: Spirituality). The Barron-68, as noted is a

subset of original MMPI items that may be administered as a standalone measure or evaluated as a clinical subscale of the MMPI. Use of the Barron-68 has declined in recent decades (likely due to the development of the MMPI-2) and no uses of this measure were found after 2007. Thus, most of the recent frequently used instruments tend to be short in length (6 to 10 items) and require 5 – 10 minutes for participants to complete (BRS-6, CD-RISC-10, and Grit-8).

Norms for the CD-RISC-10 and CD-RISC-25 are available for general population samples, students and young adults, participants with PTSD symptoms and diagnoses, participants with symptoms or diagnoses of depression, participants with other psychiatric diagnoses, participants with medical problems, participants in non-treatment seeking individuals who have experiences trauma, and participants under increased levels of stress (Davidson, 2018). Although studies have been conducted to examine reliability and validity evidence for the BRS-6 and Grit-8 scores, such extensive normative data are not presently available for these measures.

The use of altered versions of instruments. As described in previously, 190 articles were excluded from our summary, as we included only studies where researchers had used complete resilience instruments (i.e., all items from an established instrument), did not alter the wording of items or response options, did not change the response options to a different scale, and did not developed a novel scoring procedure. Administration and/or analysis of partial resilience measures (n = 140), resilience measures that were altered (n = 39), or both (partial measures that were altered, n = 11) in the present sample was concentrated in the most recent decade (2010-2019, n = 140).

Although the rationale for researchers making changes to measures is often related to ethical considerations (such as shorter measures being a more efficient use of the participants time or reducing participant fatigue) or is attributed to item relevance (i.e., authors removed

items, or altered items they believed were not an appropriate fit for their participants), measure alteration can be problematic. It is a common misconception that altered, shortened, or revised instruments retain evidence of reliability and validity that have been found in studies that used an original or previous version of the instrument. Data providing evidence of validity do not apply to subsequently revised versions of the instrument (Haynes, Smith, & Hunsley, 2019). When items or responses are altered in any way—be it wording, context, or the addition or removal of responses or items—the instrument effectively becomes a novel scale with no psychometric evidence that the new measure captures the construct of interest reliably, accurately, and completely (Smith, McCarthy, & Anderson, 2000). Unfortunately, it is not uncommon for researchers to describe an altered version or partial instrument as "reliable and valid" on the basis that original instrument had evidence of reliability and validity.

Limitations. A limitation of the current research is that results sets were limited to publications in peer-reviewed journal articles. Thus, information in the grey literature and theses and dissertations were not included. Another limitation of the current research is that, for practical and scope-management purposes, we limited our sample to resilience instruments administered in the English language and excluded translated versions of English-language measures. Therefore, the summary offered here is representative of the use of English-language measures administered in English, limited to participants who were able to complete the measures in English, also limited to articles published in the English language. Therefore, these findings cannot be assumed to generalize to any non-English administrations of these instruments, nor to English-language administrations that were published in any non-English language. Also related to practical and scope-management purposes, only instruments where the full instrument was designed to measure resilience were included (i.e., broad-scope instruments

that included one or more subscales that measured resilience were not included). Thus, the summary offered here also does not include information related to the total number of resilience instruments available and the absolute frequency of use for all characterological resilience measures.

Conclusions

Resilience as a character trait has been of interest to researchers in psychology and psychiatry for over a century. Research on the topic began with no clearly defined constructs, no measures, and only questions about how some individuals managed to withstand or "bounce back" from war, earthquakes, and a massive fire that consumed much of San Francisco in 1906. The human condition is rife with adversity, and it always will be – be it daily stressors from a high intensity workplace, systemic problems such as discrimination and poverty, or unexpected conditions such as natural disaster, incidents, and accidents.

At present, a substantial proportion of the resilience literature is focused upon university students in the United States, and very little resilience research is conducted with individuals in treatment for substance use/substance-use related difficulties, forensic/justice-involved populations, participants dealing with homelessness, socioeconomically disadvantaged populations/individuals living in poverty, and individuals who reported having experienced assault or abuse. Resilience research is also lacking in minority populations where individuals often experience inequality and differential treatment due to ethnicity, sexual orientation, and gender identity. For researchers, clinicians, social program developers, and law and policy makers, learning more about contextual resilience, resilience in terms of protective factors and risk factors, and general characterological resilience is useful in the development of programs and the provision of equal and inclusive services to underserviced populations.

The essence and foundation of the scientific method is working from what we know toward what we do not know using ethically acceptable, systematic, evidence-based research methods. A part of this process is the design of quality research, which includes understanding how the instruments to be used in a study have been used previously, and the types of reliability and validity evidence that are available (or unavailable) for that instrument in the population to be studied. Our contribution to this process a set of summaries of evidence related to structure and use of instruments to measure resilience, ego-strength, hardiness, grit, and mental toughness spanning the time period from 1953 to 2019. We hope that our findings will serve as an informative and useful starting point for the researcher who is interested in the tools available to investigate resilience in adults using a characterologically based instrument.

Acknowledgments

We would like to thank Drs. Paul Bartone, Jonathan Davidson, and Gail Wagnild for providing supplemental psychometric details, sample measures, and administration instructions for the instruments they developed. We would also like to thank Dr. Robert Cribbie for sharing his statistical expertise.

References

Ahern, N. R., Kiehl, E. M., Sole, M. L., & Byers, J. (2006). A review of instruments measuring resilience. *Issues in Comprehensive Pediatric Nursing, 29*, 103–125. https://doi.org/10.1080/01460860600677643

Alessandri, G., Vecchione, M., Caprara, G., & Letzring, T. D. (2012). The Ego Resiliency scale revised: A cross-cultural study in Italy, Spain, and the United States. *European Journal of Psychological Assessment, 28*, 139-146. https://doi.org/10.1027/1015-5759/a000102

Baltzell, A., & Zaichkowsky, L. (1999). *Psychological factors and resources related to rowers' coping in elite competition* (ProQuest Dissertations Publishing). Retrieved from http://search.proquest.com/docview/304512229/

Baron, L., Reznikoff, M., & Glenwick, D. (1993). Narcissism, interpersonal adjustment, and coping in children of holocaust survivors. *The Journal of Psychology*, *127*, 257–269. https://doi.org/10.1080/00223980.1993.9915560

Barron, F. (1953). An ego-strength scale which predicts response to psychotherapy. *Journal of Consulting Psychology, 17*, 327-333. https://doi.org/10.1037/h0061962

Bartone, P. T. (1991). Development and validation of a short hardiness measure. Paper presented at the annual convention of the American Psychological Society Annual Convention, 1991, June, Washington, DC.

Bartone, P. T. (2007). *DRS-15 (v.3) - Dispositional Resilience Scale-15 normative data, adults and college students.* (Unpublished manual). Personal communication September 27, 2014)

Bartone, P. T., Ursano, R. J., Wright, K. M., & Ingraham, L. H. (1989). The impact of a military air disaster on the health of assistance workers. *The Journal of Nervous and Mental*

Disease, 177, 317-328. https://doi.org/10.1097/00005053-198906000-00001

Benishek, L. (1996). Evaluation of the factor structure underlying two measures of hardiness. *Assessment, 3,* 423–435. https://doi.org/10.1177/107319119600300408

Benishek, L. A., Feldman, J. M., Shipon, R. W., Mecham, S. D. (2005). Development and evaluation of the Revised Academic Hardiness Scale. *Journal of Career Assessment, 13*, 59-76. https://doi.org/10.1177/1069072704270274

Bergeron, D., & Tylka, T. L. (2007). Support for the uniqueness of body dissatisfactions from the drive for muscularity among men. *Body Image, 4*, 288-295. https://doi.org/10.1016/j.bodyim.2007.05.002

Bernard, L., Belinsky, D., & Bernard, L. (1993). Hardiness, stress, and maladjustment: Effects on self-reported retrospective health problems and prospective health center visits. *Journal of Social Behavior and Personality, 8*, 97–110. https://doi.org/1761710113

Betz, N. E., & Campbell, C. (2003). Self-efficacy and personality correlates of instrumentality. Unpublished manuscript.

Block, J., & Kremen, A. M. (1996). IQ and ego-resiliency: Conceptual and empirical connections and separateness. *Journal of Personality and Social Psychology, 70*, 349-361. https://doi.org/10.1037/0022-3514.70.2.349

Campbell-Sills, L., & Stein, M. B. (2007). Psychometric analysis and refinement of the Connor-Davidson Resilience Scale (CD-RISC): Validation of a 10-item measure of resilience. *Journal of Traumatic Stress, 6*, 1019-1028. https://doi.org/10.1002/jts.20271

Canavan, J. (2008). Resilience: Cautiously welcoming a contested concept. *Child Care in Practice, 14*, 1-7. https://doi.org/10.1080/13575270701733633

Clough, P. J., Earle, K., & Sewell, D. (2002). Mental toughness: The concept and its

measurement. In I. Cockerill (Ed.), *Solutions in sport psychology* (pp. 32– 43). London: Thomson Publishing.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*. https://doi.org/10.1177/001316446002000104

Connor K. M., & Davidson, J. R. T. (2003). Development of a new resilience scale: The Connor-Davidson Resilience Scale (CD-RISC). *Depression and Anxiety, 18*, 76-82. https://doi.org/10.1002/da.10113

Costello, A. B., & Osbourne, J. W. (2005). Best practices in exploratory factor analysis: Recommendations for getting the most from your analysis. *Practical Assessment, Research, & Evaluation, 10*, 1-5. Retrieved May 17, 2017 from https://doaj.org/article/48b25a79cac54ce782cbd85252bcc009

Crandall, R. (1973). The measurement of self-esteem and related constructs. In J. P. Robinson, & P. R. Shaver, *Measures of social psychological attitudes* (Revised ed., pp.45-168). Ann Arbor, MI: Institute for Social Research.

Credé, M., Harms, P., Niehorster, S., & Gaye-Valentine, A. (2012). An evaluation of the consequences of using short measures of the big five personality traits. *Journal of Personality and Social Psychology, 102*, 874-888. https://doi.org/10.1037/a0027403

Davidson, J. R. T. (2018). *Connor-Davidson Resilience Scale (CD- RISC) Manual*. Retrieved November 18, 2019, from http://www.cd-risc.com.

Department of Justice. (2019). Bureau of Justice statistics releases results from 2018 National Crime Victimization Survey. Retrieced from: https://bjs.ojp.gov/content/pub/press/cv18pr.pdf

De Vries, H., Elliott, M., Kanouse, D., & Teleki, S. (2008). Using Pooled Kappa to summarize

interrater agreement across many items. *Field Methods*, *20*, 272–282. https://doi.org/10.1177/1525822X08317166

Diaz, A., Knighton, E., Dennehy, V., & Ross, S. (2013). *The relationship between body awareness and mental toughness in collegiate athletes* (ProQuest Dissertations Publishing). Retrieved from http://search.proquest.com/docview/1417764724/

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology, 92*, 1087-1101. https://doi.org/10.1037/0022-3514.92.6.1087

Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (Grit-S). *Journal of Personality Assessment, 91*, 166-174. https://doi.org/10.1080/00223890802634290

Duckworth, Quinn, P. D., & Tsukayama, E. (2021). Revisiting the factor structure of grit: A commentary on Duckworth and Quinn (2009). *Journal of Personality Assessment, 103*, 573–575. https://doi.org/10.1080/00223891.2021.1942022

Epstein, S. (1983). *Ego Strength Scale*. Princeton, N.J: Educational Testing Service.

Freud, S. (1927). *The ego and the id.* Honolulu, HI: Hogarth Press.

Friborg, O., Barlaug, D., Martinussen, M., Rosenvinge, J. H., & Hjemdal, O. (2005). Resilience in relation to personality and intelligence. *International Journal of Methods in Psychiatric Research, 14*, 29–42. https://doi.org/10.1002/mpr.15

Friborg, O., Hjemdal, O., Rosenvinge, J. H., & Martinussen, M. (2003). A new rating scale for adult resilience: What are the central protective resources behind healthy adjustment? *International Journal of Methods in Psychiatric Research, 12*, 65-76. https://doi.org/10.1002/mpr.143

Friborg, O., Martinussen, M. & Rosenvinge, J. H. (2006). Likert-based vs. semantic differential-based scorings of positive psychological constructs: A psychometric comparison of two versions of a scale measuring resilience. *Personality and Individual Differences, 40*, 873-884. https://doi.org/10.1016/j.paid.2005.08.015

Funk, S. C. (1992). Hardiness: A review of theory and research. *Health Psychology, 11*, 335-345.

Gottesman, I. I. (1959). More construct validation of the ego-strength scale. *Journal of Consulting Psychology, 4*, 342-346. https://doi.org/10.1037/h0038904

Greene, R., & Nowack, K. (1995). Hassles, hardiness and absenteeism: Results of a 3-year longitudinal study. *Work & Stress*, *9*, 448–462. https://doi.org/10.1080/02678379508256892

Hathaway, S. R., & McKinley J. C. (1942). *Manual for the Minnesota Multiphasic Personality Inventory.* Minneapolis, MN: University of Minnesota Press.

Haynes, S. N., Smith, G. T., & Hunsley, J. D. (2019). *Scientific foundations of clinical assessment* (2nd ed.). New York, NY: Routledge.

Housing and Urban Development. (2020). HUD 2020 Continuum of Care homeless assistance programs: Homeless populations and subpopulations. Retrieved from: https://files.hudexchange.info/reports/published/CoC_PopSub_NatlTerrDC_2020.pdf

Hunsley, J. & Mash E. J. (2018). *A guide to assessment that work*. (2nd ed.). New York, NY: Oxford University Press.

Jascó, P. (2004). Savvy searching: Citation searching. *Online Information Review, 28*, 454-460. Retrieved from http://search.proquest.com/docview/194484842/

Johnson, D. C., Pilusny, M. A., Erbes, C. R., King, D., King, L., Litz, B. T., Schnurr, P. P., Friedman, M., Pietrzak, R. H., Southwick, S. M. (2011). Development and initial

validation of the Response to Stressful Experiences Scale. *Military Medicine, 176,* 161-169. https://doi.org/10.7205/MILMED-D-10-00258

Keyworth, C., & Cookfair, J. (1998). *Responses of visiting nurses to sexual harassment by clients* (ProQuest Dissertations Publishing). Retrieved from http://search.proquest.com/docview/304491247/

King, L. A., King, D. W., Vogt, D. S., Knight, J., & Samper, R. E. (2003). Deployment Risk and Resilience Inventory (DRRI): A collection of measures for studying deployment-related experiences in military personnel and veterans. *Military Psychology, 18*, 89-120. https://doi.org/10.1207/s15327876mp1802_1

Klohnen, E. C. (1996). Conceptual analysis and measurement of the construct of ego-resiliency. *Journal of Personality and Social Psychology, 70*, 1067-1079. https://doi.org/10.1037/0022-3514.70.5.1067

Kobasa, S., Maddi, S., & Zola, M. (1983). Type A and hardiness. *Journal of Behavioral Medicine*, *6*, 41–51. https://doi.org/10.1007/BF00845275

Kobasa, S. C., Maddi, S. R., & Kahn, S. (1982). Hardiness and health: A prospective study. *Journal of Personality and Social Psychology, 42*, 168-177. https://doi.org/10.1037/0022-3514.42.1.168

Kobasa, S. C., & Puccetti, M. C. (1983). Personality and social resources in stress resistance. *Journal of Personality and Social Psychology, 45*, 839-850. https://doi.org/10.1037/0022-3514.45.4.839

Kobasa, S. C. O. (1977). *Stress, personality, and health: An overlooked possibility.* (Unpublished doctoral dissertation). University of Chicago, Chicago, Illinois, USA.

Lang, A., Goulet, C., & Amsel, R. (2003). Lang and Goulet Hardiness Scale: Development and

testing on bereaved parents following the death of their fetus/infant. *Death Studies, 27*, 851-880. https://doi.org/10.1080/074811180390241813

Langemo, D. (1990). Impact of work stress on female nurse educators. *The Journal of Nursing Scholarship*, 159–162. https://doi.org/10.1111/j.1547-5069.1990.tb00200.x

Lee, H. (1991). Relationship of hardiness and current life events to perceived health in rural adults. *Research in Nursing & Health*, *14*, 351–359. doi:10.1002/nur.4770140506

Mack, M., Ragan, B., & Mack, M. (2008). Development of the mental, emotional, and bodily toughness inventory in collegiate athletes and nonathletes. *Journal of Athletic Training*, *43*, 125–132. https://doi.org/10.4085/1062-6050-43.2.125

Maddi, S. (2002). The Story of Hardiness: Twenty years of theorizing, research, and practice. *Consulting Psychology Journal: Practice and Research, 54*, 173–185. https://doi.org/10.1037/1061-4087.54.3.173

Maddi, S. R. (2013). *Turning stressful circumstances into resilient growth*. Irvine, CA: Springer.

Maddi, S. R., Erwin, L. M., Carmody, C. L., Villarreal, B. J., White, M., & Gundersen, K. K. (2013). Relationship of hardiness, grit, and emotional intelligence to internet addiction, excessive consumer spending, and gambling. *The Journal of Positive Psychology, 8*, 128-134. https://doi.org/10.1080/17439760.2012.758306

Maddi, S. R., Harvey, R. H., Khoshaba, D. M., Lu, J. L., Persico, M., & Brow, M. (2006). The personality construct of hardiness, III: Relationships with repression, innovativeness, authoritarianism, and performance. *Journal of Personality, 74,* 575–597. https://doi.org/10.1111/j.1467-6494.2006.00385.x

Maddi, S. R., & Khoshaba, D. M. (1994). Hardiness and mental health. *Journal of Personality Assessment, 63*, 265-274. https://doi.org/10.1207/s15327752jpa6302_6

Markstrom, C. A., & Marshall, S. K. (2007). The Psychosocial Inventory of Ego Strengths: Examination of theory and psychometric properties. *Journal of Adolescence, 30*, 63-79. https://doi.org/10.1016/j.adolescence.2005.11.003

Markstrom, C. A., Sabino, V. M., Turner, B. J., & Berman, R. C. (1997). The Psychosocial Inventory of Ego Strengths: Development and validation of a new Eriksonian measure. *Journal of Youth and Adolescence, 26,* 705-732. https://doi.org/10.1023/A:1022348709532

Martin, C., & Grater, H. (1993). Mystical experience and mental health: Three perspectives (ProQuest Dissertations Publishing). Retrieved from http://search.proquest.com/docview/304078152/

Matthews, M. D. (2008). Toward a positive military psychology. *Military Psychology, 20*, 289-298. https://doi.org/10.1080/08995600802345246

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, *22*, 276–282. https://doi.org/10.11613/BM.2012.031

McNeil, K., Kozma, A., Stones, M. J., & Hannah E. (1986). Measurement of psychological hardiness in older adults. *Canadian Journal on Aging / La Revue canadienne du viellissement, 5*, 43-48. https://doi.org/10.1017/S0714980800005006

Mettle (2018). In *Merriam-Webster Online Dictionary.* Retrieved April 27, 2018, from https://www.merriam-webster.com/dictionary/mettle

Milner, J. (1988). An ego-strength scale for the Child Abuse Potential Inventory. *Journal of Family Violence, 3*, 151–162. https://doi.org/10.1007/BF00994032

Milner, J. S., & Wimberley, R. C. (1979). An inventory for the identification of child abusers. *Journal of Clinical Psychology, 35*, 95–100. https://doi.org/10.1002/1097-4679(197901)35:1%3C95::AID-JCLP2270350112%3E3.0.CO;2-K

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-Analyses: The PRISMA Statement. *PLoS Med, 6*. https://doi.org/10.1371/journal.pmed1000097

Moher, D., Shamseer, L., Clarke, M, Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., & Stewart, L. A. (2015). Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement. *Systematic Review, 4*. 1-9. https://doi.org/: 10.1186/2046-4053-4-1

National Law Center on Poverty and Homelessness. (2018). Homelessness in America: Overview of data and causes. Retrieved from: https://nlchp.org/wp-content/uploads/2018/10/Homeless_Stats_Fact_Sheet.pdf

Neill, J. T., & Dias, K. L. (2001). Adventure education and resilience: The double-edged sword. *Adventure Education and Outdoor Learning, 2*, 35-42. https://doi.org/10.1080/14729670185200061

Nishi, D., Uehara, R., Yoshikawa, E., Sato, G., Ito, M., & Matsuoka. (2013). Culturally sensitive and universal measure of resilience for Japanese populations: Tachikawa Resilience Scale in comparison with Resilience Scale 14-item version. *Psychiatry and Clinical Neurosciences, 2013, 67*, 174-181. https://doi.org/10.1111/pcn.12028

Nowack, K. (1990). Initial development of an inventory to asses stress and health. *American Journal of Health Promotion, 4*, 173-180. https://doi.org/10.4278/0890-1171-4.3.173.

Nowack, K. M. (1989). Coping style, cognitive hardiness, and health status. *Journal of Behavioral Medicine, 12*, 163-178. https://doi.org/10.1007/BF00846548

Nunberg, H. (1942). Ego strength and ego weakness. *Psychoanalytic Journal for the Arts and Sciences, 3*, 25-40. Retrieved from http://search.proquest.com/docview/1289639745/

O'Neal, M. R. (1999). *Measuring resilience*. (Unpublished doctoral dissertation). Retrieved from https://search-proquest-com.proxy.bib.uottawa.ca/docview/62411455?accountid=14701

Pangallo, A., Zibarras, L., Lewis, R., & Flaxman, P. (2015). Resilience through the lens of interactionism: A systematic review. *Psychological Assessment, 27,* 1-20. https://doi.org/10.1037/pas0000024

Pollock, S. E., & Duffy, M. E. (1990). The Health-Related Hardiness Scale: Development and psychometric analysis. *Nursing Research, 39*, 218-222. https://doi.org/10.1097/00006199-199007000-00008

Resilience. (2018). In *Online Etymology Dictionary*. Retrieved April 26, 2018, from https://www.etymonline.com/word/resilience

Resilio. (2018). In *Latin-Dictionary.com*. Retrieved April 26, 2018, from http://www.latin-dictionary.net

Revenson, T., & Cassel, J. (1991). An exploration of leadership in a medical mutual help organization. American Journal of Community Psychology, 19, 683–698. https://doi.org/10.1007/BF00938039

Rhodewalt, F., Zone, J., & Rhodewalt, F. (1989). Appraisal of life change, depression, and illness in hardy and nonhardy women. *Journal of Personality and Social Psychology*, *56*, 81–88. https://doi.org/10.1037/0022-3514.56.1.81

Rutter, P.A., Freedenthal, S., & Osman, A. (2008). Assessing protection from suicidal risk: Psychometric properties of the suicide resilience inventory. *Death Studies, 33*, 142-153. https://doi.org/10.1080.07481180701801295

Sadowski, C. J., Woodward, H. R., Davis, S. F., & Elsbury, D. L. (1983). Sex differences in adjustment correlates of locus of control dimensions. *Journal of Personality Assessment,*

47, 627-631. https://doi.org/10.1207/s15327752jpa4706_7

Sanders, G., & Suls, J. (1982). *Social psychology of health and illness*. Hillsdale, N.J: L. Erlbaum Associates.

Schwartz, S. J., Zamboanga, B. L., Weisskirch, R. S., Rodriguez, L. (2009). The relationships of personal and ethnic identity exploration to indices of adaptive and maladaptive psychosocial functioning. *The International Society for the Study f Behavioural Development, 33*, 131-144. https://doi.org/10.1177/0165025408098018

Scott, C., & Thomas, V. (2003). *The relationship among self -acceptance, ego strength, and perception of marital interaction within African American married couples* (ProQuest Dissertations Publishing). Retrieved from http://search.proquest.com/docview/305329768/

Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist, 55*, 5-14. https://doi.org/10.1037/0003-066X.55.1.5

Sheard, M., Golby, J., & van Wersch, A. (2009). Progress toward construct validation of the Sports Mental Toughness Questionnaire (SMTQ). *European Journal of Psychological Assessment, 25*, 186-193. https://doi.org/10.1027/1015-5759.25.3.186

Sim, J., & Wright, C. (2005). The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical Therapy, 85*, 257–268. http://dx.doi.org/10.1093/ptj/85.3.257

Sinclair, V. G., & Wallston, K. A. (2004). The development and psychometric evaluation of the Brief Resilient Coping Scale. *Assessment, 11*, 94-101. https://doi.org/10.1177/1073191103258144

Smith, B. W., Dalen, J., Wiggins, K., Tooley, E., Christopher, P., & Bernard, J. (2008). The Brief Resilience Scale: Assessing the ability to bounce back. *International Journal of*

Behavioral Medicine, 15, 194-200. https://doi.org/10.1080/10705500802222972

Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment, 12*, 102-111. https:doi.org/10.1037/1040-3590.12.1.102

Smith, M. S. & Gray, S. W. (2009). The Courage to Challenge: A new measure of hardiness in LGBT adults. *Journal of Gay & Lesbian Social Services, 21*, 73-89. https://doi.org/10.1080/10538720802494776

Smith-Osborne, A., & Bolton, K. W. (2013). Assessing resilience: A review of measures across the life course. *Journal of Evidence-Based Social Work, 10*, 111-126. https://doi.org/10.1080/15433714.2011.597305

Strycharczyk, D. (2017). *Mental Tough Questionnaire: A User Guide*. (version 2.00). Rossett, United Kingdom: AQR International.

Strycharczyk, D., & Clough, A. (2015). *Developing mental toughness: Coaching strategies to improve performance, resilience, and wellbeing.* London, UK: KoganPage.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics.* (5th ed.). New York, NY: Pearson.

The Hardiness Institute. (2020). Frequently asked Personal Views Survey questions. Retrieved April 18, 2020, from http://hardisurvey.com/Research_FAQs.htm

US Census Bureau. (2021). CPS historical timeline. Retrieved August 26, 2021 from: https://www.census.gov/library/visualizations/time-series/demo/cps-historical-time-series.html

Wagnild, G. (2009). *The Resilience Scale user's guide version 3.33.* Worden, MT: Resilience Center.

Wagnild, G., & Young, H. M. (1991). Another look at hardiness. *Journal of Nursing Scholarship, 23*, 257-260. https://doi.org/10.1111/j.1547-5069.1991.tb00682.x

Wagnild, G. M., & Young, H. M. (1993). Development and psychometric evaluation of the Resilience Scale. *Journal of Nursing Measurement, 1*, 165-178.

Windle, G., Benett, K. M., & Noyes, J. (2011). A methodological review of resilience measurement scales. *Health and Quality of Life Outcomes, 9*, 1-18. https://doi.org/10.1186/1477-7525-9-8

Winwood, P. C., Colon, R., & McEwen, K. (2013). A practical measure of workplace resilience: Developing the Resilience at Work Scale. *Journal of Occupational and Environmental Medicine, 55*, 1205-1212. https://doi.org/10.1097/JOM.0b013e3182a2a60a

Young, W. C., & Budynas, R. G. (2002). *Roark's formulas for stress and strain*. New York, NY: McGraw-Hill.

Younkin, S. L. (1992). *The development and validation of the Psychological Hardiness Scale.* (Unpublished doctoral dissertation). Ohio State University, Columbus Ohio, USA.

Zalaquett, C., & Wood, R. (1998). *Evaluating stress: A book of resources*. Lanham, MD: Scarecrow Press.

Zander, A., & Thomas, E. (1960). *The validity of a measure of ego strength.* Unpublished manuscript. Department of psychology, University of Michigan. Ann Arbor, United States

Project 1 Tables and Figures

Table 1.

Characterological Measures Resilience Constructs Identified in Study 1

Measure Name	*Construct*	*# of items*	*Factor(s)*	*Response Scale*	*Scoring Method*	*Reference*
Abridged Hardiness Scale / Short Hardiness Scale (AHS/SHS-20)	Hardiness	20	3 Hardiness Factors: • Commitment (6 items) • Control (9 items) • Challenge (5 items) per Rhodewalk & Zone, 1989	Not reported	Not reported	Funk (1992), Kobasa (1977), Rhodewalt & Zone (1989)
Barron's Ego Strength Scale, original version for MMPI (Barron-68)	Ego Strength	68	Unidimensional	*true / false*	MMPI scaled scoring system using standardized *T* scores (Raw score range 0-68)	Barron (1953)
Barron's Ego Strength Scale, revised version for MMPI-2 (Barron-52)	Ego Strength	52	Unidimensional	*true / false*	MMPI-2 scaled scoring system using standardized *T* scores (Raw score range 0-52)	Barron (1953)
Brief Resilience Scale (BRS-6)	Resilience	6	Unidimensional	5-point Likert scale. 1= strongly agree 2 = disagree 3 = neutral 4 = agree 5 = strongly agree	Sum of item Scores (Range 5 - 30)	Smith et al. (2008)
Brief Resilient Coping Scale (BRCS-4)	Resilience	4	Unidimensional	5-point Likert-like scale, with semantic anchors at values 1 and 5 1 = does not describe you at all 5 = describes you very well	Sum of item Scores (Range 4 - 20)	Sinclair & Wallston (2004)
Cognitive Hardiness Scale (CHS-18)	Hardiness	18	3 Hardiness Factors: • Involvement • Control • Challenge (items per factor not specified)	Not reported	Not reported	Nowack (1990)
Cognitive Hardiness Scale (CHS-30)	Hardiness	30	3 Hardiness Factors: • Involvement • Control • Challenge (items per factor not specified)	5-point Likert scale: 1 = strongly agree 2 = agree 3 = neither agree nor disagree 4 = disagree 5 = strongly disagree	Sum of item Scores (Range 30 - 150)	Nowack (1989), Nowack (1990)
Connor-Davidson Resilience Scale (CDRISC-10)	Resilience	10	Unidimensional	see CDRISC-25	Sum of item Scores (Range 0 - 40)	Campbell-Sills & Stein (2007)
Connor-Davidson Resilience Scale (CDRISC-25)	Resilience	25	5 Resilience factors: • Competence (8 items) •Tolerance of stress (7 items) • Response to change (5 items) • Control (3 items) • Spirituality (1 item)	5-point Likert scale. 0 = not true at all 1 = rarely true 2 = sometimes true 3 = often true 4 = true nearly all of the time	Sum of item Scores (Range 0 - 100)	Connor & Davidson (2003)

Measure Name	*Construct*	*# of items*	*Factor(s)*	*Response Scale*	*Scoring Method*	*Reference*
Abridged Hardiness Scale / Short Hardiness Scale (AHS/SHS-20)	Hardiness	20	3 Hardiness Factors: • Commitment (6 items) • Control (9 items) • Challenge (5 items) per Rhodewalk & Zone, 1989	Not reported	Not reported	Funk (1992), Kobasa (1977), Rhodewalt & Zone (1989)
Barron's Ego Strength Scale, original version for MMPI (Barron-68)	Ego Strength	68	Unidimensional	*true / false*	MMPI scaled scoring system using standardized *T* scores (Raw score range 0-68)	Barron (1953)
Barron's Ego Strength Scale, revised version for MMPI-2 (Barron-52)	Ego Strength	52	Unidimensional	*true / false*	MMPI-2 scaled scoring system using standardized *T* scores (Raw score range 0-52)	Barron (1953)
Brief Resilience Scale (BRS-6)	Resilience	6	Unidimensional	5-point Likert scale. 1= strongly agree 2 = disagree 3 = neutral 4 = agree 5 = strongly agree	Sum of item Scores (Range 5 - 30)	Smith et al. (2008)
Brief Resilient Coping Scale (BRCS-4)	Resilience	4	Unidimensional	5-point Likert-like scale, with semantic anchors at values 1 and 5 1 = does not describe you at all 5 = describes you very well	Sum of item Scores (Range 4 - 20)	Sinclair & Wallston (2004)
Cognitive Hardiness Scale (CHS-18)	Hardiness	18	3 Hardiness Factors: • Involvement • Control • Challenge (items per factor not specified)	Not reported	Not reported	Nowack (1990)
Cognitive Hardiness Scale (CHS-30)	Hardiness	30	3 Hardiness Factors: • Involvement • Control • Challenge (items per factor not specified)	5-point Likert scale: 1 = strongly agree 2 = agree 3 = neither agree nor disagree 4 = disagree 5 = strongly disagree	Sum of item Scores (Range 30 - 150)	Nowack (1989), Nowack (1990)
Connor-Davidson Resilience Scale (CDRISC-10)	Resilience	10	Unidimensional	see CDRISC-25	Sum of item Scores (Range 0 - 40)	Campbell-Sills & Stein (2007)
Connor-Davidson Resilience Scale (CDRISC-25)	Resilience	25	5 Resilience factors: • Competence (8 items) •Tolerance of stress (7 items) • Response to change (5 items) • Control (3 items) • Spirituality (1 item)	5-point Likert scale. 0 = not true at all 1 = rarely true 2 = sometimes true 3 = often true 4 = true nearly all of the time	Sum of item Scores (Range 0 - 100)	Connor & Davidson (2003)
Connor-Davidson Resilience Scale (CDRISC-10)	Resilience	10	Unidimensional	see CDRISC-25	Sum of item Scores (Range 0 - 40)	Campbell-Sills & Stein (2007)
Connor-Davidson Resilience Scale (CDRISC-25)	Resilience	25	5 Resilience factors: • Competence (8 items) •Tolerance of stress (7 items) • Response to change (5 items) • Control (3 items) • Spirituality (1 item)	5-point Likert scale. 0 = not true at all 1 = rarely true 2 = sometimes true 3 = often true 4 = true nearly all of the time	Sum of item Scores (Range 0 - 100)	Connor & Davidson (2003)
Dispositional Resilience Scale (DRS-15)	Hardiness & Resilience	15	3 Hardiness Factors: • Commitment (5 items) • Control (5 items) • Challenge (5 items)	see DRS-45	Sum of item Scores (Range 0 - 45)	Bartone (2007)
Dispositional Resilience Scale (DRS-30)	Hardiness & Resilience	30	3 Hardiness Factors: • Commitment (10 items) • Control (10 items) • Challenge (10 items)	see DRS-45	Sum of item Scores (Range 0 - 90)	Bartone (1991)
Dispositional Resilience Scale (DRS-45)	Hardiness & Resilience	45	3 Hardiness Factors: • Commitment (15 items) • Control (15 items) • Challenge (15 items)	4-point Likert scale. 0 = none 1 = a little 2 = often 3 = very often	Sum of item Scores (Range 0 - 135)	Bartone et al. (1989)
Ego Resilience Scale for California Personality Inventory (CPI) (Klohnen-ER-29)	Ego Resilience	29	Unidimensional	*true / false*	CPI scale scoring system using standard *T* distribution (Raw score range 0-29)	Klohnen (1996)

Ego Strength Scale (Epstein-ESS-26)	Ego Strength	26	Unidimensional	5-point Likert-like scale with strongle disagree and strongly agree as anchor points (values for each anchor point not given in Schwartz, et al., 2009)	Sum of item Scores (Range not reported)	Epstein (1983) as cited in Schwartz et al. (2008)
Ego Strength Scale (ESS-40)	Hardiness	40	Unidimensional	*true / false*	Sum of item Scores (Range 0 - 40)	Milner (1988), Milner & Wimberly (1979)
Grit Scale / Original Grit Scale (Grit-12)	Grit	12	2 Grit factors: • Consistency of Interest (6 items) • Perseverance of Effort (6 items)	5-point Likert-like scale 1 = not at all like me 5 = very much like me values 2-4 not semantically defined	Sum of item Scores (Range 12 - 60)	Duckworth et al. (2007)
Hardiness Scale (HS-15)	Hardiness	15	Not reported	4-point Likert-like scale with semantic anchors: 0 = strongly disagree 3 = strongly agree values 1-2 not semantically defined	Sum of item Scores (Range 0 - 60)	Kobasa & Puccetti (1983) additional detail from Baron, Reznikoff, & Glenwick (1993)
Hardiness Scale (HS-20)	Hardiness	20	3 Hardiness Factors: • Commitment (7 items) • Control (9 items) • Challenge (4 items)	Not reported	Challenge item score doubled and summed with other subscale scores, then converted to standardized score (Lee, 1991)	Kobasa, Maddi, & Zola (1983) additional detail from Lee (1991)
Hardiness Scale (HS-13)	Hardiness	13	Not reported	Not reported	Not reported	Kobasa in Sanders & Suls (1982), Revenson & Cassel (1991)
Hardiness Scale 50-item version / Revised Hardiness Scale (HS-50/RHS-50)	Hardiness	50	3 Hardiness Factors: • Commitment (16 items) • Control (17 items) • Challenge (17 items) per Bernard et al., 1993	Not reported	Not reported	Authors cited communication from Kobasa in 1989 as source of scale additional detail from Bernard, Belinsky, & Bernard (1993)
Hardisurvery III-R (Hardisurvery III-R-65)	Hardiness	65	3 Hardiness Factors: • Commitment • Control • Challenge (items per factor not specified)	Scoring performed by The Hardiness Institute, response option details unavailable	Hardiness Institute Scoring algorithm unavailable	Maddi (2002), The Hardiness Institute (2020)
HardiSurveyIII-R	Hardiness	106	3 Primary Hardiness Factors: • Commitment • Control • Challenge (items per factor not specified)	5-point scale with semantic anchors: 0 = complete disagreement 4 = complete agreement values 1-3 not semantically defined	Sum of item Scores (Range 0 - 424)	Maddi & Khoshaba (2001), The Hardiness Institute (2020) additional detail from: Funk, 1992; O'Neal 1999
Hardy Personality Inventory (HPI-50)	Hardiness	50	3 Primary Hardiness Factors: • Commitment • Control • Challenge (items per factor not specified)	Not reported	Sum of item Scores (Range 0 - 100)	Kobasa, Maddi, & Kahn (1982) additional detail from: Langemo (1990)
Lang and Goulet Hardiness Scale (LGHS-45)	Hardiness	45	3 Hardiness Factors: • Sense of Personal Control (16 items) • Active Orientation (13 items) • Sense of Meaning (16 items)	5-point Likert scale 1 = strongly disagree 2 = disgree 3 = neither agree nor disagree 4 = agree 5 = strongly agree	Sum of item Scores (Range 45 - 225)	Lang et al. (2003)
MeBTough Assessment (Me-B-Tough-43)	Mental Toughness	43	3 Mental Toughess Factors: • Perceived mental toughness • Perceived emotional toughness • Perceived physical toughness (items per factor not specified)	7-point Likert-like scale with semantic anchors 1 = almost never 4 = sometimes 7 = always values 2, 3, 5, 6 not semantically defined	Sum of item Scores (Range 43 - 329)	Mack & Ragan (2008)
Mental Toughness Questionnaire (MTQ-48)	Mental Toughness	48	4 Mental Toughness Factors: • Commitment • Control • Challenge • Confidence items per factor not described	5-point Likert-like scale with semantic anchors 1 = strongly disagree 5 = strongly agree values 2-4 not semantically defined	Sum of item Scores (Range 48 - 240) Scoring and report services available from copyright holder AQR International	Clough et al. (2002), Diaz et al. (2013), Strycharczyk (2017)

Mental Toughness Questionnaire (MTQ-18)	Mental Toughness	18	see MTQ 48	see MTQ 48	Sum of item Scores (Range 18 - 90)	Clough et al. (2002)
Personal Views Survey (PVS) (PVS-50)	Hardiness	50	3 Hardiness Factors: • Commitment (16 items) • Control (17 items) • Challenge (17 items)	4-point Likert scale 0 = not at all true 1 = a little true 2 = quite a bit true 3 = completely true	Scoring performed by The Hardiness Institute Score range & scoring algorithm unavailable	The Hardiness Institute (2020) additional detail from: Keyworth & Cookfair (1998)
Personal Views Survey II (PVS-II-50)	Hardiness	50	3 Hardiness Factors: • Commitment • Control • Challenge (items per factor not specified)	4-point Likert scale 0 = not at all true 1 = a little true 2 = quite a bit true 3 = completely true	Scoring performed by The Hardiness Institute Score range & scoring algorithm unavailable	The Hardiness Institute (2020) additional detail from: Zalaquett & Wood (1997) as cited in Baltzell (1999)
Personal Views Survey III (PVS-III-30)	Hardiness	30	3 Hardiness Factors: • Commitment • Control • Challenge (items per factor not specified)	4-point Likert scale 0 = not at all true 1 = a little true 2 = quite a bit true 3 = completely true	Scoring performed by The Hardiness Institute Score range & scoring algorithm unavailable	The Hardiness Institute (2020)
Personal Views Survey III-R (PVS-III-R-18)	Hardiness	18	3 Hardiness Factors: • Commitment (6 items) • Control (6 items) • Challenge (6 items) items per factor in Atkinson, 2013	4-point scale with semantic anchors 0 = not at all true 3 = completely true values 1-2 not semantically defined	Scoring performed by The Hardiness Institute Score range & scoring algorithm unavailable	Maddi (2013), Maddi, Harvey, Khoshaba, Lu, Persico, & Brow (2006), The Hardiness Institute (2020)
Psychological Hardiness Scale (PHS-20)	Hardiness	20	Unidimensional	5-point Likert scale 1 = strongly disagree 2 = disgree 3 = aren't sure are or neutral 4 = agree 5 = strongly agree	Sum of item Scores (Range 40 - 200)	Betz & Campbell (2003) as cited in Bergeron & Tylka (2007)
Psychological Hardiness Scale (PHS-40)	Hardiness	40	Unidimensional	5-point Likert scale 1 = strongly disagree 2 = disgree 3 = aren't sure are or neutral 4 = agree 5 = strongly agree	Sum of item Scores (Range 40 - 200)	Younkin (1992)
Psychosocial Inventory of Ego Strength (PIES-64)	Ego Strength	64	8 Ego Strength Factors: • Hope (8 items) • Will (8 items) • Purpose (8 items) • Competence (8 items) • Fidelity (8 items) • Love (8 items) • Care (8 items) • Wisdom (8 items)	5-point Likert-like scale 1 = does not describe me 5 = describes me very values 2-4 not semantically defined	Sum of item Scores (Range 64 - 320)	Markstrom et al. (1997), Markstrom & Marshall (2007)
Resilience Scale (RS-14)	Resilience	14	Unidimensional	see RS-25	Sum of item Scores (Range 14 - 98)	Wagnild & Young (1993), Wagnild (2009)
Resilience Scale (RS-15)	Resilience	15	Unidimensional	see RS-25	Sum of item Scores (Range 15 - 105)	Neill & Dias (2001)
Resilience Scale (RS-25)	Resilience	25	2 Resilience Factors: • Personal Competence (17 items) • Acceptance of Self and Life (8 items)	7-point Likert-like scale with semantic anchors 1 = disagree 7 = agree values 2-6 not semantically defined	Sum of item Scores (Range 25 - 175)	Wagnild & Young (1993), Wagnild (2009)
Resilience Scale (RS-26)	Resilience	26	Uses resilience factors of RS-25 with one overall resilience item not mapped to a specific factor	see RS-25	Sum of item Scores (Range 26 - 182)	Wagnild & Young (1993), Wagnild (2009)
Resilience Scale for Adults, 33-item Likert response English version (RSA-33)	Resilience	33	6 ResilienceFactors: • Personal Competence (6 items) • Personal Strength (4 items) • Structured Style (4 items) • Social Competence (6 items) • Family Cohesion (6 items) • Social Resources (7 items)	5-point Semantic Differential Interval Scale	Sum of item Scores (Range 33 - 165)	Friborg et al. (2005)

Resilience Scale for Adults, 37-item Likert response English version (RSA-37)	Resilience	37	5 Resilience Factors: • Personal Competence (10 items) • Social Competence (7 items) • Family Coherence (7 items) • Social Support (8 items) • Personal Structure (5 items)	7-point Likert-like scale with semantic anchors 1 = not true at all 7 = very true values 2-6 not semantically defined	Sum of item Scores (Range 37 - 259)	Friborg et al. (2003), Friborg et al. (2006)
Resilience Scale for Adults, 37-item semantic differential response version, English) (RSA-37sd)	Resilience	37	5 Resilience Factors: • Personal Competence (10 items) • Social Competence (7 items) • Family Coherence (7 items) • Social Support (8 items) • Personal Structure (5 items)	7-point Semantic Differential Interval Scale	Sum of item Scores (Range 37 - 259)	Friborg et al. (2005), Friborg et al. (2006)
Response to Stressful Experiences Scale (RSES-22)	Resilience	22	5 Resilience Factors: • Meaning-making & Restoration (9 items) • Active Coping (6 items) • Cognitive Flexibility (1 items) • Spirituality (1 items) • Self-efficacy (2 items) Note: one item has 2 factors and 2 items have no factor	5-point Likert-like scale with semantic anchors 0 = not at all like me 4 = exactly like me values 1-3 not semantically defined	Sum of item Scores (Range 0 - 88)	Johnson et al. (2011)
Revised Hardiness Scale (RHS-36)	Hardiness	36	3 Hardiness Factors: • Commitment (12 items) • Control (16 items) • Challenge (8 Items)	4-point Likert-like scale with semantic anchors 0 = not at all true 3 = completely true values 1-2 not semantically defined	Sum of item Scores (Range 0 - 108)	Benishek (1996), Funk (1992) additional detail from Greene & Nowack (1995)
Short Grit Scale (Grit-8)	Grit	8	2 Grit factors: • Consistency of Interest (4 items) • Perseverance of Effort (4 items)	5-point Likert-like scale 1 = not at all like me 5 = very much like me values 2-4 not semantically defined	Sum of item Scores (Range 8 - 40)	Duckworth et al. (2007), Duckworth & Quinn (2009)
Short Hardiness Scale (SHS-20-McNeill)	Hardiness	20	3 Hardiness Factors: • Commitment (4 items) • Control (5 items) • Challenge (4 items) Note: items 2, 4, 11, 13, 15, 16 & 18 not assigned to these factors	Items 1-14 scored on a 4-point scale with 1 semantic anchor: 0 = not at all true andvalues 1-3 not semantically defined Items 15-20 are scored as 0 or 2	Sum of item Scores (Range 0 - 54)	McNeill et al. (1986)
Tachikawa Resilience Scale, English version) (TRS-10)	Resilience	10	Unidimensional	7-point Likert-like scale with semantic anchors 1 = strongly disagree 7 = strongly agree values 2-6 not semantically defined	Sum of item Scores (Range 10 - 70)	Nishi, Uehara, Yoshikawa, Sato, Ito, & Matsuoka (2013)
The Ego Resilience Scale / ER89 (ER-89-14)	Ego Resilience	14	Unidimensional	4-point Likert scale. 1 = does not apply at all 2 = applies slightly 3 = applies somewhat 4 = applies very strongly	Sum of item Scores (Range 14 - 56)	Block & Kremen (1996)
The Ego Resilience Scale-Revised (ER-89-R-10)	Ego Resilience	10	2 Ego-Resilience Factors: • Optimal Regulation (6 items) • Openness to Life Experience (4 items)	7-point Likert-like scale with semantic anchors,response options 2-6 not semantically defined. 1 = does not apply at all 7 = applies very strongly	Sum of item Scores (Range 10 - 70)	Alessandri et al. (2012)
Thomas-Zander Ego Strength Scale (TZ-ESS-20)	Ego Strength	20	Unidimensional	*true / false*	Sum of item Scores (Range 0 - 20)	Scott & Thomas (2003), Zander & Thomas (1960)
Thomas-Zander Ego Strength Scale (TZ-ESS-24)	Ego Strength	24	Unidimensional	*true / false*	Sum of item Scores (Range 0 - 24)	Crandall (1973) as cited in Sadowski et al. (1983)
Thomas-Zander Ego Strength Scale (TZ-ESS-27)	Ego Strength	27	Unidimensional	*true / false*	Sum of item Scores (Range 0 - 27)	Zander & Thomas (1960) as cited in Martin (1993)

Table 2.

Measure Use by Population Category

Population Category & Subcategories	*n* of Studies
University	393
Medical/Nursing Students	35
Psychology Students	98
Other University Students	247
University related - Other	13
Military	132
Active Duty	36
Veterans	66
Cadets	9
Military Personnel - Other	21
Vocation / Profession	195*
First Responders	22
Healthcare Professsions - other	32
Mental Health Professionals	10
Nurses	25
Physicians / Medical Doctors	8
Selected for Work / Profession - Other	81
Teachers (non-university)	8
University Faculty/Other University Employees	10
Caregivers	63*
Parenting or Pregnancy	34
Of Family with Mental Health or Medical Diagnosis	36
Caregiver - other	2
Medical Patients	181
Brain Injury (ABI, TBI, Organic)	11
Arthritis	8
Asthma	3
Cancer	24
Chronic Pain, and/or Fibromyalgia	16
Chronic-Obstructive Pulmonary Disease (COPD)	2
Diabetes	2
HIV and/or AIDS	13
Heart Condition (Myocardial Infarction, Heart surgery patients, and/or Other heart failure)	9
Multiple Sclerosis (MS)	8
Parkinson's Disease	2
Physical Disability (incl. hearing or vision loss)	6
Spinal Cord Injury (SCI)	6
Medical Pts. - Various / Other Specific Condition	57
Control Group Participants w/o Medical Condition	14
Mental Health (MH)	201*
MH Services - Inpatients	63
MH Services - Outpatients	36
MH Services - Inpatient/Outpatient status not stated	20
Individuals recruited for specific diagnoses or symptom clusters	
Anxiety (generalized, social, and worry)	5
Depression and/or Suicide-related	12
Post-Traumatic Stress Disorder Diagnosis/Symptoms	15
Schizophrenia and/or other psychotic symptoms	10
Substance Use Disorder Pts./Symptoms	39
Asymptomatic comparison group / control group	18
MH Symptoms - Other	30
Specific events known or hypothesized to cause distress	77
Bereavement	9
Combat (participant, or expsure to)	5
Holocaust or other Genocide	3
Impacted by Earthquake	7
Impacted by Hurricane	5
Impacted by Industrial Explosion	3
Impacted by Oil Spill	5
Motor Vehicle Accident	2
Reported Experience of Assault or Abuse	14
Impacted by other event / various events	23
Control Group - did not experience specific event	2
Selected for Sex, Gender, or Sexual Orientation	217*
Sample limited to individuals who identify as Female	98
Sample limited to individuals who identify as Male	103
Sample limited to individuals who identify as heterosexual	5
Sample limited to individuals who identify as LGBTQIA+	23
Sample limited to cisgendered individuals only	1
Sample limited to transgender-identified individuals only	5
Ethnicity (selected for identification with specific ethnicity)	57
Not selected for a specific parameter or attribute /Convenience Sample	98
Other	149*
Sample was limited to individuals age 65+	14
Athletics, Sport or Exercise Participation	46
Sample was limited to Couples only	14
Forensic or Prison Setting	11
Homeless Individuals or Shelter Residents	11
Immigrants	2
Impacted by Poverty, low SES, or impoverished region	10
Twins or Triplets	5
Other (e.g., specific conference, retreat attendees, etc.)	37

*Overall count does not equal sum of subcategories because samples were coded with more than one subcategory

Figure 1. Flowchart of Inclusion/Exclusion Criteria for Study 1

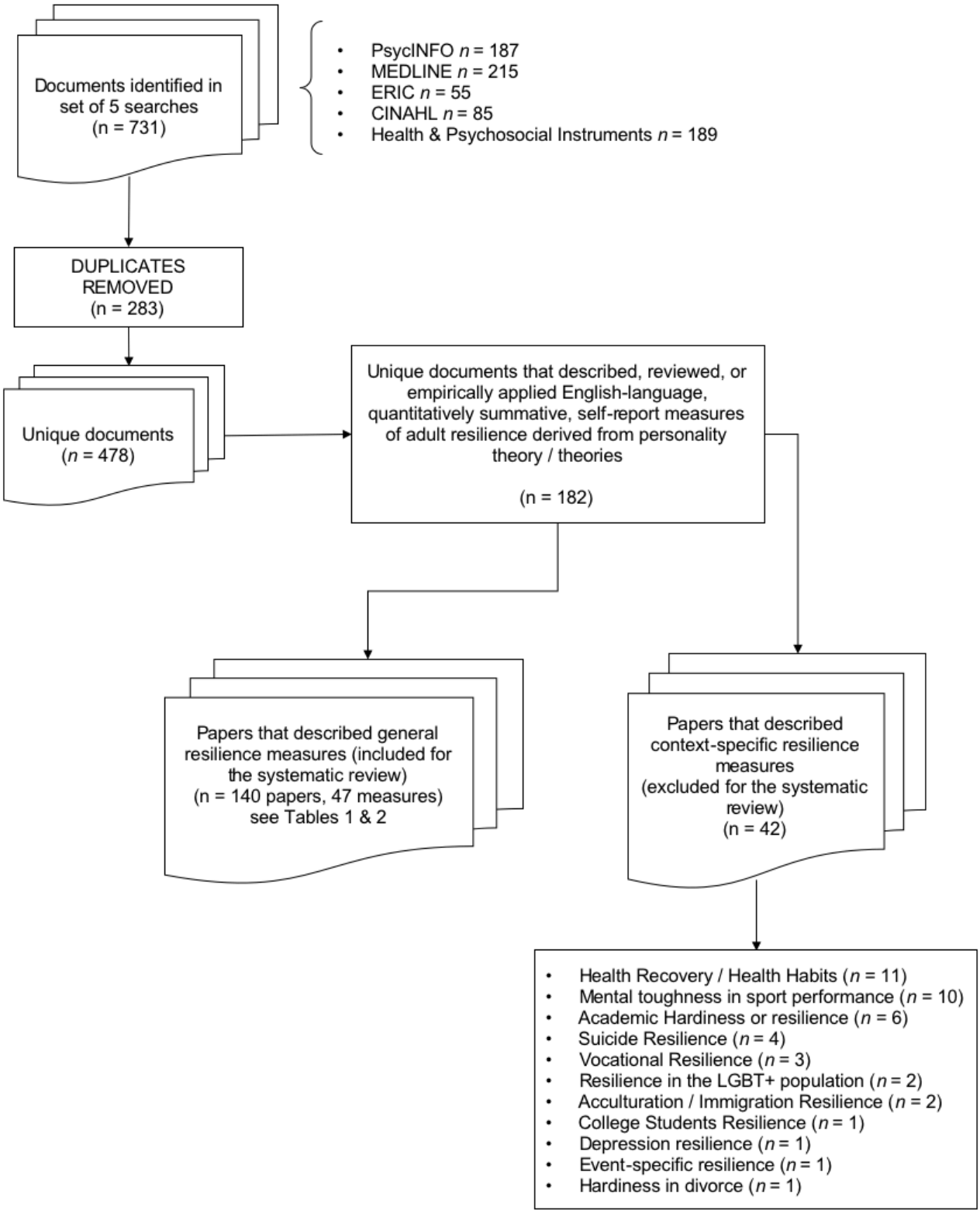


Figure 2. Flowchart of Inclusion/Exclusion Criteria for Study 2

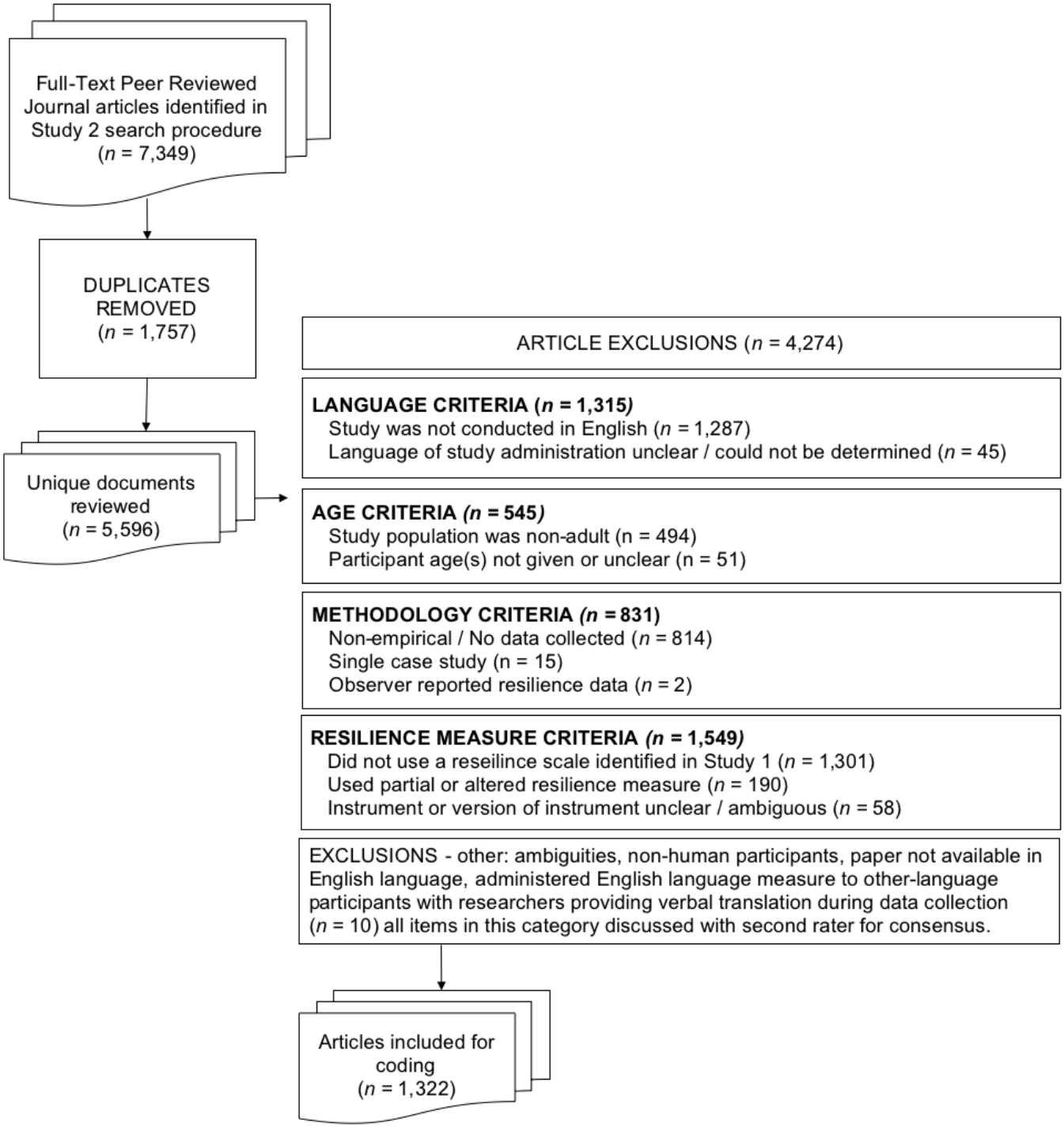


Figure 3. Resilience Construct Distribution Across Studies

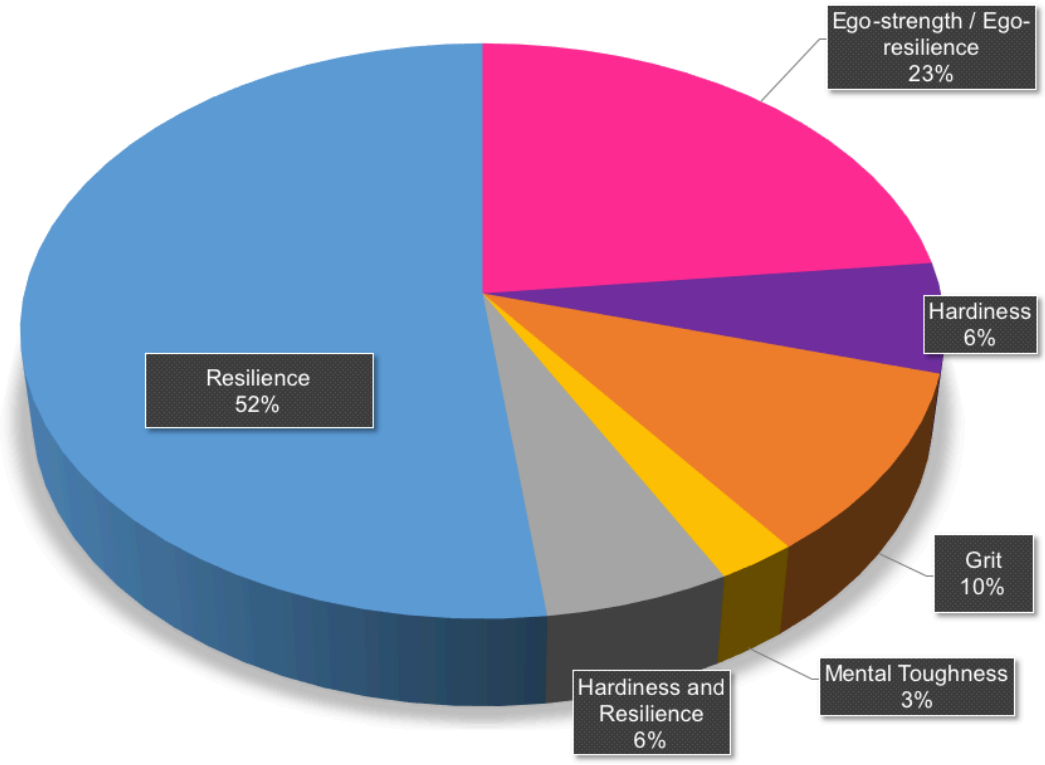


Figure 3. Percentage of resilience measure use frequency by resiliency construct (*n* = 1,193 studies).

Figure 4. Resilience Constructs Measured by Decade

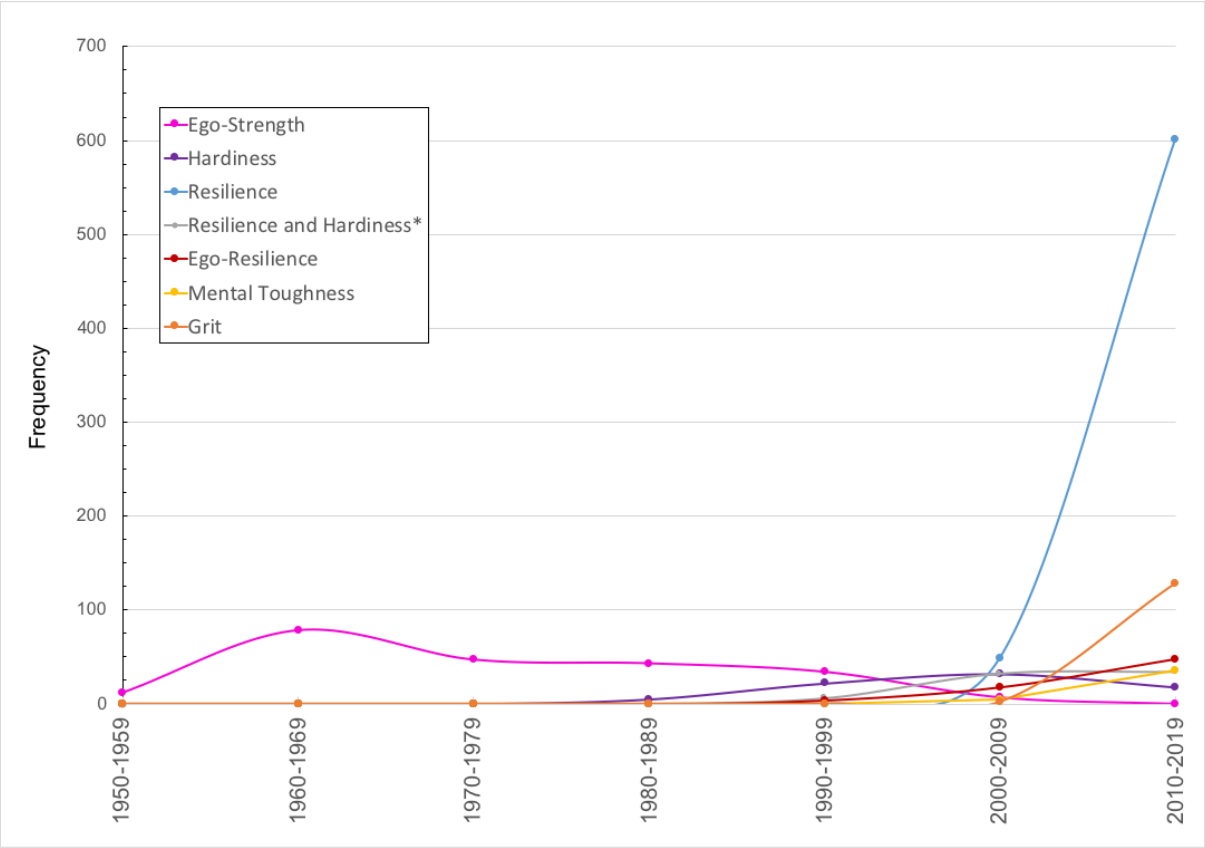


Figure 4. Timeline of resilience measure use for the measures identified in Study 1 by construct from the earliest ego-strength measure in 1957 to June 2019 (*n* = 1,239 measure applications in 1,193 studies).

Figure 5. Resilience Measure Use by Location

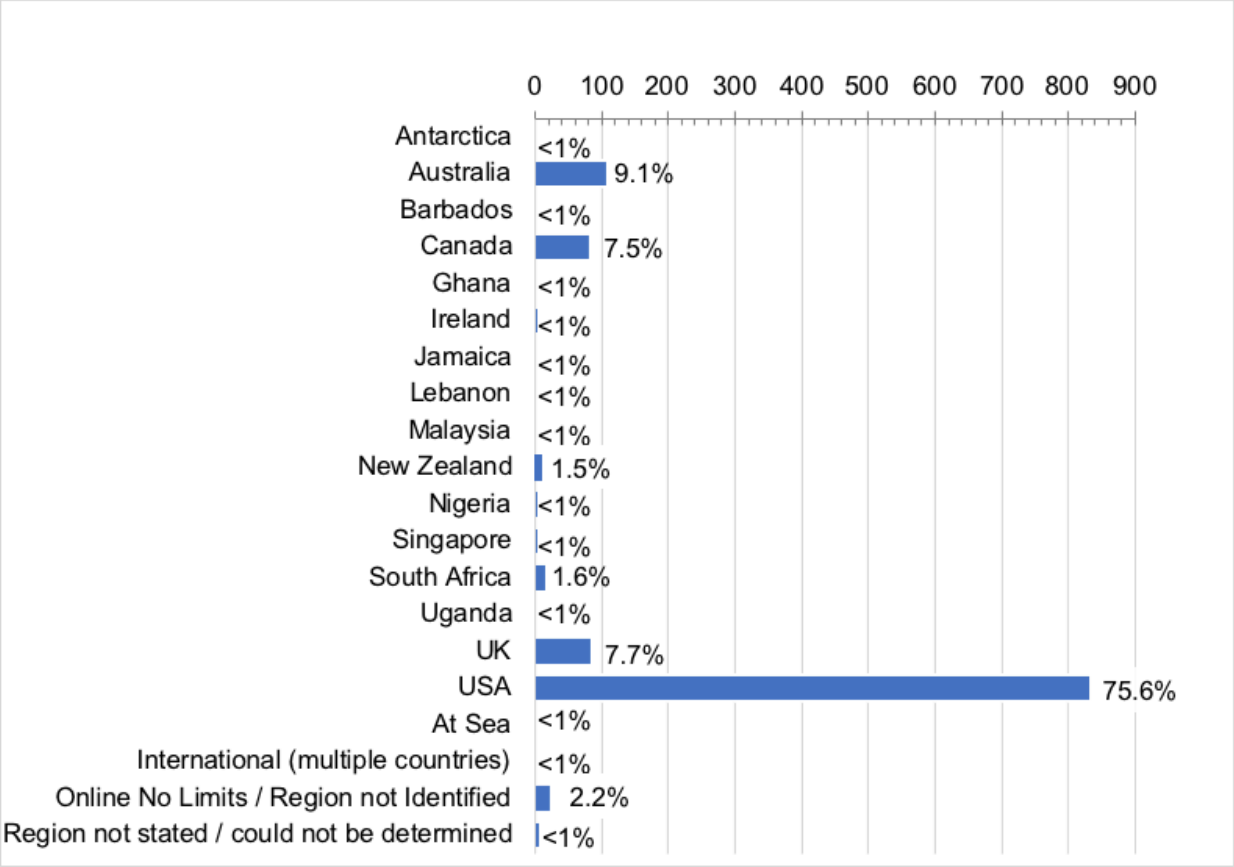


Figure 5. Frequency of resilience measure for the measures identified in Study 1 in 1,193 studies from 1957 to June 2019 by country/region.

Figure 6. Resilience Measure Use by Instrument.

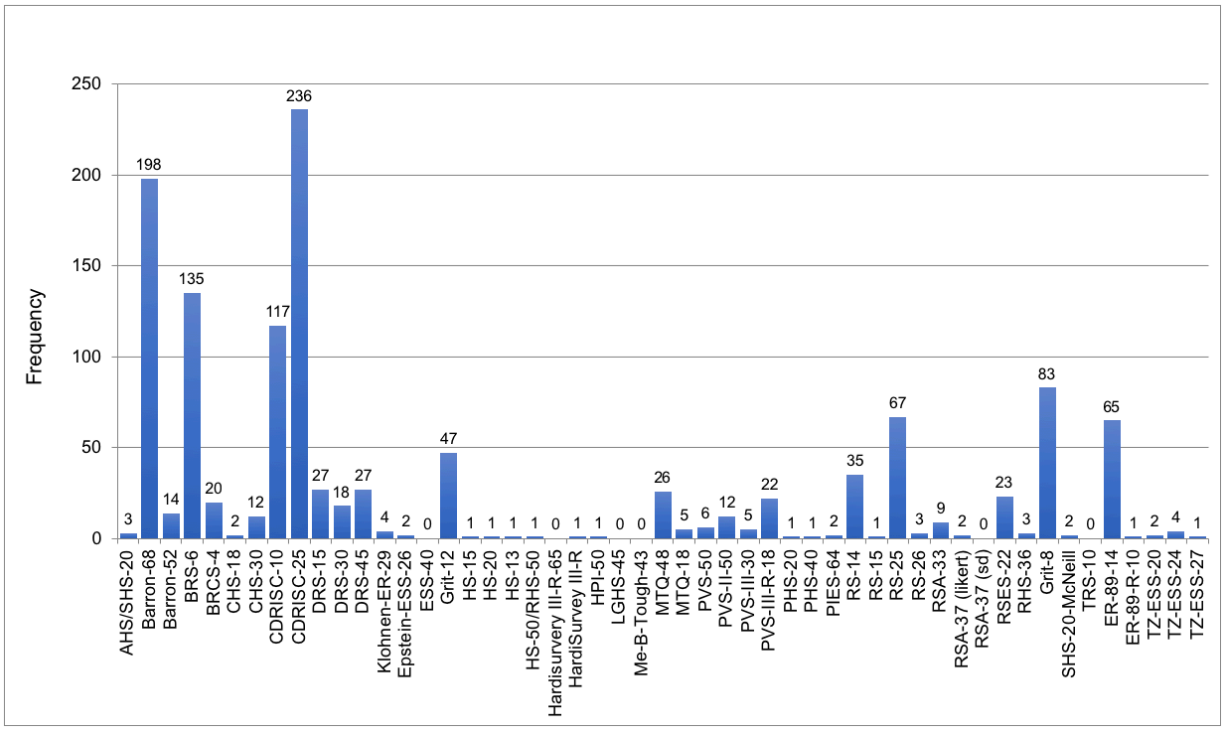


Figure 6. Frequency of resilience measure use by instrument for the measures identified in Study 1 in 1,193 studies from 1957 to June 2019.

Figure 7. Resilience Measure Use by Decade

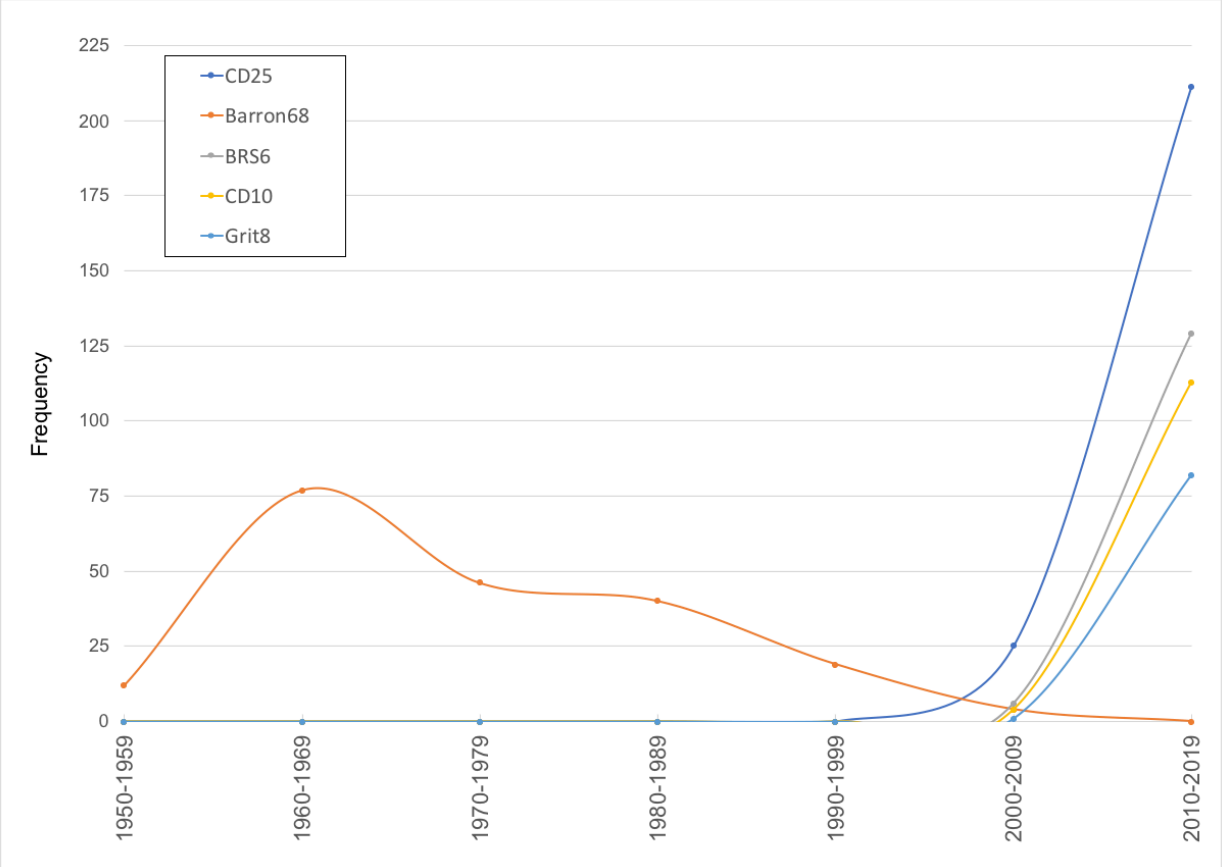


Figure 7. Frequency of resilience measure use by instrument for the 5 most frequently used resilience measures in 1,193 studies from 1957 to June 2019.

Project 2

The Brief Resilience Scale: Meta-Analyses of Reliability, Convergent Validity, and Concurrent Validity

Corresponding author:

Teresa A. Allan

Co-Authors:

John Hunsley

William Lawrence

University of Ottawa

May 17, 2021

Abstract

A series of meta-analyses were conducted to investigate the psychometric properties of scores on the Brief Resilience Scale (BRS) reported in 149 studies of adult participants. Reliability generalization methods were used to summarize 86 observations of internal consistency (mean Cronbach's α = .86) and validity generalization methods were used to summarize eight observations of convergent validity (M_r = .61). Concurrent validity was also investigated via validity generalization methods for the most frequently observed mental health correlates of BRS scores. For each set of mental health measures, an estimated summary effect was calculated (anxiety, M_r = -.47; depression, M_r = -.49; optimism, M_r = .44; social support, M_r = .24; and stress, M_r = -.53). Of the significant moderators conducted for concurrent validity estimates, the largest was sample category (accounting for 100% of the variance in the observed correlations between the BRS and measures of social support). Geographic location and participant age, accounted for 55%, and 24%, of the variation in the observed correlations between the BRS and significantly moderated the relation between the BRS and perceived stress. We discuss how the results we obtained relate to the BRS in these specific contexts and to resilience theory, and the potential broader implications of these findings.

Keywords:

Brief Resilience Scale, BRS, resilience, meta-analysis, adults, reliability, validity

Introduction

Psychological resiliency has been described in many ways, including ego-strength, ego-resilience, hardiness, mental toughness, resilience, and grit (Barron, 1953; Connor & Davidson, 2003; Duckworth, Peterson, Matthews, & Kelly, 2007; Freud, 1927; Kobasa, 1979; Strycharczyk & Clough, 2015). A systematic review of 1,322 empirical articles using characterologically focused general resilience measures for adults found the Brief Resilience Scale (BRS; Smith, Dalen, Wiggins, Tooley, Christopher, & Bernard, 2008) to be the third most frequently used resilience measure (Allan, Hunsley, & Lawrence, 2020), after the Connor-Davidson Resilience Scale, 25-item version (CD-RISC 25: Connor & Davidson, 2003) and the Barron Ego-Strength Scale, 68-item version (Barron, 1953). The BRS measures resilience conceptualized as a unidimensional characteristic of "bouncing back" from stressors. The instrument developers suggested that a single-factor parsimonious scale was both less laborious for individuals to complete and easier to interpret than existing multidimensional scales that included several sub-factors or multiple constructs (Smith et al., 2008). The BRS has six items, three of which are positively worded, and three of which are negatively worded and reverse-scored (e.g., *I tend to bounce back quickly after hard times, I have a hard time making it through stressful events*). Participants completing this measure rate each item on a 5-point Likert scale. Scores are summed and the range of possible scores is 5 to 30, with higher scores interpreted as indicating greater individual resilience (Smith et al., 2008). Given the extensive use of the measure, the two main purposes of the present research were to provide meta-analytic summaries of the reliability and validity evidence supporting the BRS.

Initial Psychometric Evidence

Smith et al.'s (2008) initial investigation of the psychometric properties of the BRS was

conducted in a series of four studies with two samples of undergraduate students (n = 128 and n = 64), a sample of 112 cardiac rehabilitation patients, and a sample of 50 women (20 with, and 30 without, symptoms of fibromyalgia). Internal consistency estimates calculated for each sample were Cronbach's α = .84, .87, .80, and .91, respectively. Cronbach's α can be interpreted such that values ranging from .80 to .89 represent *good* internal consistency, and values greater than or equal to .90 represent *excellent* internal consistency (Hunsley & Mash, 2018).

With respect to validity evidence, Smith et al. provided both a description of theory and the rationale underlying the development of items. Factor analyses of each of the four test development samples were conducted and items were found to load onto a single resilience factor and these loadings ranged from .67 to .91 for each item with an average loading of .76 (*SD* = .08) (Smith et al., 2008). Construct validity in the test development samples was assessed in terms of convergent validity, concurrent validity, and discriminant validity evidence. Convergent validity was examined with the CD-RISC-25 (Connor & Davidson, 2003) in the sample of 128 undergraduate students and found to be $r = .59$, $p < .01$. Convergent validity was also assessed between BRS scores and scores obtained from the Ego Resilience Scale 14-item version (Block & Kremen, 1996) in the sample of 128 undergraduate students ($r = .51$, $p < .01$), and in the sample of 50 women (20 fibromyalgia patients and 30 individuals reporting no symptoms of fibromyalgia) ($r = .49$, $p < .01$) (Smith et al., 2008). Concurrent validity in the test development samples was assessed by examining the correlation between BRS scores and 15 measures in four psychosocial domains (personality/characteristics, coping styles, social relationships, and outcomes related to physical and mental health). BRS scores were found to be positively correlated with: resilience measures, optimism, purpose in life, social support, adaptive coping, and positive affect; and were found to be negatively correlated with pessimism, negative social

interactions, maladaptive coping, perceived stress, fatigue, pain, as well as symptoms of anxiety and depression. Finally, Smith et al.'s (2008) examination of known groups validity in the test development samples yielded statistically significant differences between the group of patients being treated for fibromyalgia and the control group of individuals with no fibromyalgia symptoms, such that BRS scores could be used to differentiate to which group (symptoms/no symptoms) participants belonged.

Summarizing Reliability and Validity Evidence

Reliability and validity evidence described by instrument developers is an integral part of the measure development process, and the initial estimates provided by Smith et al. (2008) are very encouraging. These estimates, however, are not invariant properties of any psychometric instrument as measures can perform differentially across populations, settings, administrations, and contexts (Vacha-Hasse, 1998). Assuming that evidence for reliability and validity found in the initial validation sample apply to subsequent administrations of the instrument "…would be akin to saying that a fine piano that is well-tuned will always be in tune, and will sound just as good regardless of the kind of music that is played on it, or of who plays it…the quality of the sound a piano makes is a function not only of the instrument itself, but also of variables related to the music, the piano player, and the setting (e.g., the acoustics of the room) where is it played" (Urbina, 2004, p. 120). How well an instrument measures what it is intended to measure, and with what consistency, can vary across contexts, populations, and administrations. General patterns of test performance across contexts is a continual and ongoing process and multiple administrations, across a wide range of populations and contexts, are needed to establish general patterns of reliability and validity evidence (Hunsley & Mash, 2018).

Reliability generalization. As multiple measure administrations are necessary to

establish that test scores are likely to be reliable in a given sample, a method to summarize numerous observed estimates is needed. Reliability generalization is a meta-analytic procedure that summarizes effect sizes and estimates that specifically pertain to the investigation of reliability evidence (Borenstein, Hedges, Higgins, & Rothstein, 2009). The general premise is to generate an estimated average effect size (a summary effect) of specific statistics or estimates (e.g., Cronbach's α, Intra-class Correlation Coefficients, Pearson's *r*) weighted by the number of observed scores and score variability (Hunter & Schmidt, 2004; Vacha-Hasse, 1998). These summary effect sizes and respective confidence intervals can be used to describe reliability evidence across studies to gain a sense of how instrument scores and psychometric properties vary across populations and contexts. Further, reliability generalization also includes meta-analytic procedures to assess how certain sample and study attributes may moderate these estimates (i.e., do variables such as age, sex/gender identity, or geographic location explain any of the variability in either effect sizes or estimates observed across populations and contexts?) (Schmidt & Hunter, 1977; Vacha-Hasse, 1998).

Validity generalization. Validity generalization is meta-analytic procedure similar to reliability generalization except that the summaries generated relate to evidence of differing types of validity evidence (e.g., convergent, discriminant, predictive, and concurrent). Convergent validity is the degree of similarity between the scores on two measures that were designed to assess the same construct (i.e., two instruments that measure resilience). Discriminant validity is a type of validity evidence that indicates if an instrument can be used to distinguish between two populations. For example, there is evidence of discriminant validity if scores on the measure can be used to accurately distinguish between military veterans who have and do not have symptoms of PTSD. Both concurrent validity and predictive validity are types of

validity evidence that describe consistent relations between a predictor (e.g., resilience score) and the occurrence of criterion (outcome) variables (e.g., scores on a measures of mental health symptoms). The feature differentiating these two types of validity evidence is that concurrent validity evidence is collected when two measures are administered at the same time point (e.g., the correlation between scores on a resilience measure and scores on an anxiety measure, both collected at baseline); and predictive validity evidence is collected when two measures are administered at two differing timepoints (e.g., resilience scores collected in university students in first year and completion/non-completion of their program of study – such that the score on the resilience measure is being used to attempt to predict an outcome that occurs at a later point in time.

Similarly, potential moderators of the observed variation in validity evidence across contexts, populations, and study characteristics can also be investigated. For example, if scores on a measure yield consistent correlations with scores on measures of mental health in populations of older adults, but yields inconsistent correlations with scores on the same measures in young adults, a researcher whose participants are university undergraduates may want to select a measure with more adequate validity evidence for use with young adults, or at least bear in mind the observed variability of past administrations of the measure when interpreting and discussing study results or making clinical treatment recommendations.

The Current Study

Although the BRS has been frequently used in the research literature, there is comparatively little summative information about the BRS in the literature. In contrast, the frequently used Connor-Davidson resilience scales have been the focus of numerous studies related to evidence for the factor structure, reliability, and validity for these instruments;

extensive normative sample information is also available (Davidson, 2018). The Barron Ego-Strength (Es) subscale for the original Minnesota Multiphasic Personality Inventory has also been used and researched extensively and large-scale normative datasets are available for this measure (MMPI: Hathaway & McKinley, 1942). At present, there are no reliability/validity generalization summaries available for the BRS. Thus, the selection of the BRS for reliability and validity generalization analyses was based upon an identified gap in the available information related to the psychometric properties of the measure and the frequency of recent applications in the literature. Accordingly, the goals of this research were to conduct, for the BRS, (a) reliability generalization analyses for estimates of internal consistency, (b) validity generalization analyses for estimates of convergent validity and concurrent validity, and (c) moderator analyses to examine sample characteristics that might account for possible heterogeneity of reliability and validity estimates.

Materials and Methods

Data Collection

Literature search. Citation searching has been found to be an effective and efficient method to obtain a collection of documents that cite a single root document or collection of root documents (Jascó, 2004). Therefore, consistent with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Guidelines (Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009; Moher, Shamseer, Clarke, Ghersi, Liberati, Petticrew, Shekelle, & Stewart, 2015), a Web of Science citation search limited to peer-reviewed journal articles was conducted for the period of "no beginning date" to June 2019 for articles that cited the development document for the BRS: "*The Brief Resilience Scale: Assessing the ability to bounce back*" (Smith et al., 2008). The ProQuest Theses and Dissertations database was also

searched using “no beginning date” limiter to June 2019. A citation search of the ProQuest Theses and Dissertations database returned only two documents, therefore an expanded “anywhere in abstract” search using the same date range and the Boolean “Brief Resilience Scale” OR BRS* was used. In the sections that follow, all journal articles, theses, and dissertations are referred to as “articles.”

Supplemental data from study authors. Where data relevant to the meta-analyses were not provided in the article (e.g., an internal consistency statistic such as Cronbach’s α for the BRS specific to the sample) or where additional details related to the number of participants, age, sex/gender identification, or BRS mean and standard deviation for the sample was not reported, these coefficients, statistics, and datasets were requested using the email addresses given for corresponding authors. Where no email address was given, publications were searched in Google and on ResearchGate.com to ensure every effort was made to obtain as much complete data as was possible.

Inclusion and exclusion criteria.

Adult participants. For inclusion, the articles must have included at least one study where the BRS was completed by adult participants. Only articles where the mean age of study participants was greater than or equal to 18 years were included. Articles with studies that did not report the mean age of participants were excluded if the participants were described as “children,” “adolescents,” or primary/middle/high school students. Articles with studies that did not report the mean age of participants, but an age of 18+ years could be reasonably inferred from clear semantic identifiers were included. For example, participants described as “American military combat veterans” (given age requirements for military enrollment in the United States) were included. Similarly, studies that described participants as “older adults,” “successfully

aging," or "retirees" were included as these descriptors suggest it is reasonable to assume participants were 18 years or older. Studies that did not report the mean age of participants and did not use clear semantic identifiers of adulthood were excluded.

Language criteria. Only articles where resilience instruments were administered to all participants in the English language were included, as this was the language in which the BRS was originally developed. If the language of administration was not clearly stated, the study was coded as administered in the official language of the region of the participants. For example, if a study that met all other inclusion criteria described the collection of data from university students in the midwestern United States, and did not state the language of administration, the study was assumed to have been administered in the official language of that region (English) and was included. Similarly, if a study that was conducted in Madrid, Spain did not state the language of administration, it was assumed to have been conducted in the official language of that country (Spanish) and excluded.

Unique datasets. Some articles had data and participant overlap wherein a study had been described in more than one article. Identifying overlap was critical to avoid redundancy and prevent inflated participant counts (which would yield inaccurately skewed summary estimates). Therefore, when a given study was described in several articles, only the article containing the largest sample was used for data extraction purposes. The only exception to this was where the article containing the larger sample did not describe observed BRS scores; in those instances, the largest reported sample that described these results were retained and coded for the meta-analyses. The method described in Wood (2007) was the basis of the procedure used in the present analyses to identify duplicate studies, overlapping participants, and data redundancy. This method, based on first identifying overlapping authors, however, was not sufficient to

identify all study duplication, as large-scale national studies or long-term longitudinal studies conducted by research teams may have published articles containing the same data but do not share any authors between publications. Thus, to further eliminate redundancy, all studies that provided a study or dataset name (e.g., Project SERVE) were also assessed for overlap. Sub-samples were excluded if results for research items were already coded for a larger number of participants.

Coding Procedure. Data collected for the meta-analyses were organized into four levels: (a) article level attributes (such as title, year of publication, etc.), (b) study level attributes, as each article may describe more than one study, and (c) sample level attributes, as each study may include data from more than one sample (e.g., military veterans, undergraduate students, individuals living in areas impacted by natural disaster), and (d) research item level (e.g., a Pearson's *r* value given for observed BRS scores and observed scores on another measure would be a single research item). Here again, each study (and each group within each study) may have several research items. To identify each, unique reference numbers were assigned to corresponding data at each level using the variables: `Article_ID, Study_ID, Sample_ID,` and `ResearchItem_ID`.

Two raters (the first and third authors) participated in the coding of each retained that met inclusion criteria. All coded variables were entered into three spreadsheets (studies, samples, research items) by the two raters in accordance with an *a priori* established coding manual (Appendix B: Coding Manual). Rater 1 (the first author) coded 100% of articles. Rater 2 (the third author) coded 25% of the articles. Articles coded by Rater 2 were randomly selected using the article ID number and a random number generator in R.

Data Analysis Procedures

Software. The primary meta-analyses (reliability generalization, validity generalization, and assessment of moderators of unexplained variance) were carried out in the R statistical analyses environment (R Core Team; 2020) and used R packages metafor (Viechtbauer, 2010), and robumeta (Fisher, Tipton, & Zhipeng, 2017), with the code given in Quintana (2015) and Viechtbauer (2010). Analyses following procedures outlined in Borenstein et al. (2009), Quintana (2015), Schmidt and Hunter (1977), Thompson and Vacha-Haase (1998), Vacha-Haase (1998), and Viechtbauer (2010). As both the distributions of Cronbach's α and Pearson's *r,* tend to be skewed, Fisher's r-to-z transformation in the R package metafor was used (Feldt & Charter, 2006; Thompson & Vacha-Hasse, 2000; Viechtbauer, 2010). Subsequent to the analyses, the transformation process was reversed using the same R package. Thus, means and confidence intervals are reported in the same format as the original test statistics (Cronbach's α and Pearson's *r*). Supplementary analyses were also carried out in Excel, Numbers, R (R Core Team; 2020), SPSS version 20.0 (IBM Corp., 2011), and FileMakerPro version 16 (Apple, Inc., 2016). Analyses for the assessment of the possibility of publication bias followed procedures: Egger's z regression test and funnel plot of distribution asymmetry (examines shape of the distribution of standard error values around the summary effect size), Orwin's Fail-Safe *N* (estimates the number of additional studies that would be needed with an observed effect size below a selected threshold to reduce the observed summary effect produced by the model to non-significance), Rosenthal's Fail-Safe *N* (yields the number of null effect studies that would be needed to reduce the summary effect estimate produced by the model to non-significance) (Egger, Smith, and Schneider, 1997; Orwin, 1983; Rosenthal, 1977). Howell and Shield's Fail-Safe *N* was used in place of Orwin's and Rosenthal's Fail-Safe *N* calculations to assess the File Drawer Effect for estimates of internal consistency, as the aforementioned methods are not recommended to assess

potential publication bias on reported values of Cronbach's α (Howell and Shields; 2008). Three statistics were used to investigate variance heterogeneity for the reliability and validity generalization data: τ^2 (the standard deviation of the true effect size) estimated using T^2, the Q-statistic (used to assess the statistical significance of T^2), and I^2 (to evaluate the magnitude of heterogeneity across observed effects) (Dersimonian & Laird, 1986; Higgins, Thompson, Deeks, & Altman, 2003; Huedo-Medina et al., 2006). Inter-rater reliability was assessed using Cohen's κ_{POOLED} following the procedure outlined in DeVries, Elliott, Kanouse, and Teleki (2008).

Model selection. In studies where it is theorized that true effects vary, a random effects model is the appropriate analytic framework, as the model accounts for the variation in the observed effect size variance across studies (Quintana, 2015; Viechtbauer, 2010). A random effects model assumes that the true effect sizes across a collection of studies are sampled from an unknown distribution of effect sizes with a mean denoted by θ and a standard deviation denoted by τ^2 and that variation (not assumed to be error) must mathematically be included in the analyses of effects (Pigott, 2012). Therefore, the random-effects model uses an estimated weighted population mean, θ, and an estimation of the true population variance, τ^2 to describe the observed effect variation across studies (Borenstein et al., 2009; Quintana, 2015; Pigott, 2012). Thus, random effects models (REML and REML-Clustered analyses) were used for the summary effect size and moderator analyses in the current study.

Missing data. No data imputation was used for any of the analyses. First, many of the primary studies have already imputed missing values using a range of imputation techniques (e.g., simple procedures such as replacement with measures of central tendency, and fairly complex multivariate imputation processes inclusive of recursive iterative procedures). Data integrity/quality is reduced with each imputation procedure, thus performing additional

imputations on data that has already undergone imputation procedures would make the results of these analyses less replicable and interpretable. Second, given several definitions of resilience, from a theoretical standpoint it seems possible that individuals with higher levels of resilience may be more likely to be perseverant in the completion of studies, resilience training programs, or treatments than individuals with lower levels of resilience who may stop participation if they experience setbacks or feel discouraged (Duckworth & Quinn, 2009; Smith et al., 2008). Thus, it seems unlikely missing data in resilience-focused studies would be missing completely at random.

Results

Data Collection

Literature search. The citation search procedure yielded 861 full-text peer-reviewed journal articles. Duplicates (n = 290) were removed and (n = 571) unique articles were screened for inclusion (Figure 1). Abstract screening was not sufficient to determine if inclusion criteria were met as abstracts are brief and summative typically do not contain this level of detail. Screening for inclusion was conducted primarily by the first author as the inclusion/exclusion criteria were objective and defined *a priori* (e.g., there is little ambiguity in assessing if data were collected, if the BRS was used in the study, and if the mean age of participants was greater than or equal to 18 years where a mean age was given). However, in instances where the first author concluded ambiguity related to the article meeting/failing to meet inclusion criteria, the article was reviewed by the second rater and a consensus was reached. The final number of retained articles was 163. These articles described 149 unique studies.

Author-provided supplemental data. Of the 149 studies that met inclusion criteria, 17 contained all items of interest for the set of meta-analyses, and 123 authors were contacted for

additional supplemental information. Contact information was not available for nine authors. Forty-four of the contacted authors responded (35.8% response rate) and 32 of these (26.0%) provided some (n = 13) or all (n = 19) of the information requested. Sixteen emails were returned as undeliverable to the corresponding authors given address, and 63 emails appeared to have been sent without error, but no response was received (Figure 1).

Inter-rater Reliability

Rater 1 (the first author) coded 100% of articles (n = 123). Rater 2 (the third author) coded 25% of the articles (n = 31). Inter-rater reliability was assessed for variables related to study attributes (range κ = .96 to 1.0 for each variable, κ_{POOLED} = .98) for variables related to sample attributes (range from κ = .98 to 1.0, κ_{POOLED} = .99), and for variables related to research items (range from κ = .97 to 1.0, κ_{POOLED} = .99). Disagreements in coding were resolved by discussions between the two raters.

Reliability Generalization

Internal consistency. Of the 149 studies that used the BRS that met our inclusion criteria, internal consistency estimates specific to their study participants were available for 85 studies (87 samples of participants; 82 were given in the published articles and 5 were provided by study authors in response to the request for additional information). Cronbach's coefficient α was reported for 86 samples, and Zumbo's ordinal coefficient α ($\alpha_{ordinal}$ = .94) was reported for one study. As the $\alpha_{ordinal}$ coefficient is not statistically equivalent to Cronbach's α, it was not included in the meta-analyses for internal consistency. Thus 86 observations of Cronbach's α for the BRS (across 84 studies, 86 samples, and 26,372 participants) were included. The reported Cronbach's α estimates ranged from α = .363 to .94 (Figure 2). Eighty-four observations exceeded the α = .7 threshold that is recommended to be indicative of *adequate* internal

consistency, and of these, 62 estimates exceeded the α = .8 threshold indicative of *good* internal consistency, and of those 22 exceeded the .9 threshold indicative of *excellent* internal consistency (Hunsley & Mash, 2018).

Significant heterogeneity among Cronbach's α estimates was found (T^2 = .039, CI_{95} [0.03, 0.06], Q = 944.15, df(85), p<.0001, I^2 = 92.16, n = 86 samples). Given this, a random-effects model was used to calculate the summary effect model for these data. The resultant estimated mean internal consistency was Cronbach's α = .863, CI_{95} [0.85, 0.87], n = 86 (Figure 2). No evidence of publication bias for the sample of 80 included articles was found (Egger's z = 1.02, p = .307). The inconsequential threshold for the coefficient α_{TARGET} values were explored for α_{TARGET} = .80 (a value indicative of *good* internal consistency) and α_{TARGET} = .70 (interpreted as *adequate* internal consistency), given the assumption that studies where BRS internal consistencies *below adequate* would be those more likely to be unpublished (Hunsley & Mash, 2018) (Figure 3). The resulting Howell and Shields' Fail-safe values were N = 51 file drawer studies to reduce the mean internal consistency to below the threshold of *good* (α_{TARGET} = .80) and N = 212 file drawer studies to reduce the mean inconsistency to below the threshold of *adequate* (α_{TARGET} = .70).

Potential moderators of variability in internal consistency. The following potential moderators were assessed: Self-reported sex/gender identity (percentage female, male, transgender, and other), mean age of participants, country where study participants were located, sample category (five categories were derived from the authors descriptions of participants: university students, individuals selected for specific careers/vocations/or employment status, individuals selected for specific physical or mental health characteristics, individuals selected for other identified characteristics, and samples of individuals that were not selected for any specific

characteristics). The five preceding categories were also aggregated into two superordinate categories based on academic enrollment status at the time the study was conducted (Student / Non-student).

Of the variables related to self-reported sex or gender-identity, the percentage of male participants (n = 60, df(1), p = .016, R^2 = .014) and percentage of transgender participants (n = 60, df(1), p < .001, R^2 = .016) were found to be a significant moderators of the variability in internal consistency estimates and accounted for 1.4% and 1.6% of the observed variation, respectively. The direction of these small moderator effects were that, when the percentage of self-identified transgender participants increased, the estimated Cronbach's α increased; when the percentage of self-identified male participants increased, the estimated Cronbach's α decreased. Geographic location (country) was also found to be a significant moderator of the variability in internal consistency estimates (n = 86 observations across all categories, df(1), and p = .036, R^2 = .019). As location only accounted for 1.9% of variability, it is of little utility to provide a detailed breakdown of the correlational means and standard deviations by country. It is relevant to note, however, that the lowest observed internal consistency estimate (α = .36, substantially lower than the estimated summary effect size of α = .86), was based on scores from the only sample (n = 52) of participants located in Rwanda.

Sample category significantly moderated variation in estimates of internal consistency (n = 86 observations across all categories, df(1), p = .009, R^2 = .084). The means and standard deviations for each sample category were: Samples selected for physical or mental health characteristics (n = 12 reported estimates, M_α = .83, SD_α = .08), samples of students (n = 25 reported estimates, M_α = .85, SD_α = .04), samples selected for a specific vocation or profession (n = 17 reported estimates, M_α = .87, SD_α = .05), samples selected for specific other

characteristics (e.g., athletes, identification as LGBTQIA+); n = 13 reported estimates, M_{α} = .82, SD_{α} = .12), and samples that were not selected for any specific characteristics (n = 13 reported estimates, M_{α} = .89, SD_{α} = .03). Although these moderators are statistically significant, the difference across groups is relatively small (range of M_{α} = .82 to .89) and the entire moderation effect accounts for 8.4% of variability in the observed estimates of internal consistency. Finally, mean age of sample participants and student status of participants were not significant moderators of variation among internal consistency estimates.

Validity Generalization

Convergent validity was examined in terms of reported correlations between BRS scores and other measures of characterological resilience (operationally defined as measures that explicitly state in their development documents that these were designed to measure resilience, ego strength, hardiness, grit, or mental toughness, and that were not situationally or contextually specific). Concurrent validity was examined in terms of the most frequently reported correlations with related psychosocial constructs, grouped for analysis by construct. No studies beyond the initial validation sample described discriminant and predictive validity related to the BRS.

Convergent validity. Across four studies, eight observations of convergent validity were reported for five resilience instruments. These data were all either baseline measurements in longitudinal studies or data reported in cross-sectional studies. In each of these, participants completed the BRS at the same time point as the comparison instrument. The reported Pearson's r values ranged from r = .49 to .72. Significant variance heterogeneity (T^2 = .015, CI_{95} [<0.01, 0.08], Q =30.95, df(7), p<.0001, and I^2 = 77.52) was found among these observations of Pearson's r, thus a clustered hierarchical random effects model was used for the summary effect size analysis. The summary effect (estimated weighted mean Pearson's r) for these data was M_r

= .61, CI_{95} [0.55, 0.67], n = 8 (Figure 4). Possible publication bias for the sample of included articles was found (Egger's z = -2.09, p = .037). However, no evidence of publication bias was found using Rosenthal's Fail-safe N = 1,861, p_{TARGET} = .05; Orwin's Fail-safe N = 4, r_{TARGET} = .50). The value r = .50 was selected for the r_{TARGET} as the lowest value among recommended acceptable threshold for reported correlations of convergent validity between psychometric instruments assuming that values that are below this level are less likely to be reported in the published literature (Carlson & Herdman, 2012).

Potential moderators of variation in convergent validity. The potential moderators assessed in the reliability generalization analyses (age, and sample type – both using the five categories previously described, and also aggregated into two superordinate categories: student and non-students) were investigated for these convergent validity data. Of these, no significant moderators were found to explain variation in the reported correlations for convergent validity (n = 8, df(7), p <.001, R^2 = .614). Insufficient data were available to conduct moderation analyses of sex/gender identity. Country was also not analyzed as a potential moderator due to insufficient heterogeneity (all included samples were collected from participants located in the United States).

Post-hoc meta-analysis. Further constraint of the data to include only measures that the test developers defined as measuring the construct as *resilience* (i.e., omitting ego-resilience measures), yielded a summary effect size of M_r = .66 (n = 5, df(4), p = .035,CI_{95} [0.61, 0.71]) for reported Pearson's r values ranged from r = .51 to .72. Significant variance heterogeneity was evident (T^2 = .007, CI_{95} [<0.01, 0.15], Q =10.30, df(4), p=.036, and I^2 = 63.56), however, insufficient data were available to conduct analyses of moderators for this subset of the data. Convergent values above the mean were found between the BRS and the 25-item CD-RISC

(Conner & Davidson, 2003, r = .72, n = 396) and between the BRS and the 25-item Resilience Scale (Wagnild & Young, 1993, r = .70, n = 396). Convergent validity values below the mean were found between the BRS and the 25-item CD-RISC (Conner & Davisdon, 2003; r = .59, n = 128), the 4-item Brief Resilient Coping Scale (BRCS: Sinclair & Wallston, 2004; r = .51, n = 50), and 37-item Resilience Scale for Adults (RSA: Friborg, Barlaug, Martinussen, Rosenvinge, & Hjemdal, 2005, r = .65, n = 396). Thus, convergent validity both above and below the mean was found for the CD-RISC. Specifically for the CD-RISC, the sample with the greatest observed convergent validity value (r = .72) between the CD-RISC and the BRS in an internet-based sample not selected for specific characteristics, where the mean age of the participants was 37 years, and the lowest convergent validity value (r = .59) was a sample of university students with a mean age of 20.4 years.

Concurrent validity / BRS score correlated constructs. Meta-analyses were conducted to describe the summary effect sizes and potential moderators of variation in the summary effect size for each construct where there were more than 10 reported correlations between the BRS and measures of a given construct. The constructs investigated were: anxiety (Figure 5) , depression (Figure 6), optimism (Figure 7), social support (Figures 8 and 9), and stress (Figures 10 and 11). Each meta-analysis described an estimated summary effect size, analyses of potential publication bias, and analyses of the potential moderators using the same moderator variables as in the previous sections. Table 1 contains the summary effect data.

Anxiety. Sixteen studies examined correlations between the BRS and measures of anxiety. Significant heterogeneity was found among these observations of Pearson's r; thus, a random effects model was used for the summary effect size analysis (T^2 = .005, CI_{95} [0.00, 0.02], Q = 31.23, df(15), p<.008, I^2 = 56.97, n = 16 samples). The overall summary effect (estimated

weighted mean Pearson's *r*) for these data was *r* = -.47, CI_{95} [-0.51, -0.43], *n* = 4,410 participants (Figure 5). No evidence of publication bias for the sample of included articles was found (Egger's z = 1.02, *p* = .307; Rosenthal's Fail-safe *N* = 4,489, p_{TARGET} = .05; Orwin's Fail-safe *N* = 13, r_{TARGET} = -.30). The value *r* = -.30 was selected for the α_{TARGET} as a generally accepted threshold at which correlations between psychometric instruments are considered to be meaningful.

Mean age of participants, geographic location of participants (country), student/non-student status, and sample category (using the five categories described in previous analyses) were not found to be significant moderators of the correlation between participants BRS scores and their scores on measures of anxiety. Only three (18.8%) of the samples included sex/gender data specific to the participants described in the correlational analyses. Therefore, the variable self-reported sex/gender identity could not be used for moderator analyses as insufficient data were available.

Depression. Eighteen studies examined correlations between the BRS and measures of depression. Significant heterogeneity was found among these observations of Pearson's *r*, thus a random effects model was used for the summary effect size analysis (T^2 = .014, CI_{95} [0.01, 0.04], *Q* = 69.15, *df*(17), *p*<.001, I^2 = 76.81, *n* = 18 samples).The overall summary effect (estimated weighted mean Pearson's *r*) for these data was *r* = -.49, CI_{95} [-0.54, -0.43], *n* = 4,672 participants (Figure 6). No evidence of publication bias for the sample of included articles was found (Egger's z = 1.02, *p* = .307; Rosenthal's Fail-safe *N* = 5,894, p_{TARGET} = .05; Orwin's Fail-safe *N* = 14, r_{TARGET} = -.30). As in the previous analysis, the value *r* = -.30 was selected for the α_{TARGET} as a generally accepted threshold at which correlations between psychometric instruments are considered to be meaningful.

The mean age of participants, geographic location of participants (country), sample category (using the five categories described in previous analyses) , and student/non-student status were not found to be a significant moderators of the correlation between participants BRS scores and their scores on measures of depression. Similar to the previous set of meta-analyses, the variable self-reported sex/gender identity could not be used for moderator analyses as insufficient data, as four (22.2%) of the samples included sex/gender data specific to the participants described in the correlational analyses.

Optimism. Twelve studies examined correlations between the BRS and measures of optimism. Significant heterogeneity was found among these observations of Pearson's *r*; thus, a random effects model was used for the summary effect size analysis (T^2 = .032, CI_{95} [0.01, 0.11], Q = 65.33, df(11), p<.001, I^2 = 86.46, n = 12 samples). The overall summary effect (estimated weighted mean Pearson's *r*) for these data was r = .44, CI_{95} [0.35, 0.53], n = 2,635 participants (Figure 7). No evidence of publication bias for the sample of included articles was found (Egger's z = 1.33, p = .183; Rosenthal's Fail-safe N = 1,616, p_{TARGET} = .05; Orwin's Fail-safe N = 8, r_{TARGET} = .30). The value r = .30 was selected for the r_{TARGET} as a generally accepted threshold at which correlations between psychometric instruments are considered to be meaningful.

Mean age of participants, geographic location of participants (country), student/non-student status, and sample category (using the five categories described in previous analyses) were not found to be significant moderators of the correlation between participants BRS scores and their scores on measures of optimism. Only four (33.3%) of the samples included sex/gender data for the participants described in the correlational analyses; thus, the variable self-reported sex/gender identity could not be used for moderator analyses as insufficient data were available.

Social Support. Seventeen studies examined correlations between the BRS and measures of social support. Two studies used more than one measure of social support. Significant heterogeneity was found among these observations of Pearson's *r*; thus, a random effects model was used for the summary effect size analysis (T^2 = .004, CI_{95} [<0.001, 0.11], Q = 33.63, df(19), p<.001, I^2 = 46.50, n = 17 samples, 20 reported correlates). Thus, these data were analyzed using a clustered hierarchical random-effects model with a random-effects maximum likelihood estimator for Tau (i.e.,17 clusters and 20 correlational datapoints). The overall summary effect (estimated weighted mean Pearson's *r*) for these data was r = .24, CI_{95} [.20, .29], n = 3,334 participants (Figure 8).

Egger's regression test for asymmetry suggested possible publication bias, but visual inspection of the funnel plot suggests that this calculation was impacted by a single outlier value (Figure 9). No evidence of publication bias for the sample of included articles was found using Rosenthal's Fail-safe N = 1,508, p_{TARGET} = .05 and Orwin's Fail-safe N = 25, $\text{EffectSize}_{\text{TARGET}}$ = .12. The summary effect size (.24) divided by two was selected for the $\text{EffectSize}_{\text{TARGET}}$ as the threshold for Orwin's Fail-safe N for clustered analyses, it is recommended to use (Viechtbauer, 2010). (T^2 = .014, CI_{95} [0.00, 0.01], Q = 33.36, df(19), p<.001, I^2 = 46.50, n = 20 samples, 17 clusters).

The moderator of sample category (using the five categories described in previous analyses) accounted for 100% of the variance in the association between observed scores on the BRS and measures of social support (n = 20, df(4), p = .0001, R^2 = 1.0). For each sample type, the means and standard deviations were: Samples selected for characteristics related to mental and/or physical health (n = 8 reported correlates, M_r = .27, SD_r = .17), university students (n = 5 reported correlates, M_r = .25, SD_r = .05), samples selected for vocation or profession (n = 1

reported correlate, $r = .35$), samples selected for specific other characteristics ($n = 5$ reported correlates, $M_r = .17$, $SD_r = .02$), and samples not selected for any specific characteristic $n = 1$ reported correlate, $r = .35$).

Mean age of participants, geographic location, and student/non-student status were not found to be significant moderators of the correlation between participants BRS scores and their scores on measures of social support. The variable self-reported sex/gender identity could not be used for moderator analyses as insufficient data were available. Only six (35%) sample clusters included sex/gender data for the participants described in the correlational analyses).

Stress. Seventeen studies examined correlations between the BRS and measures of stress. Significant heterogeneity was found among these observations of Pearson's *r*; thus, a random effects model was used for the summary effect size analysis ($T^2 = .028$, CI_{95} [0.01, 0.07], $Q = 149.79$, $df(16)$, $p<.001$, $I^2 = 89.72$, $n = 17$ samples). The overall summary effect (estimated weighted mean Pearson's *r*) for these data was $r = -.53$, CI_{95} [-0.59, -0.46], $n = 5{,}654$ participants (Figure 10).

Egger's regression test for asymmetry suggests possible publication bias (Egger's z = -3.68, $p = <.001$). The observed pattern of asymmetry in these data is that the standard error of the summary effect size estimate is greater for correlations of greater magnitude, and lower for correlations of a lesser magnitude, rather than evenly distributed around the summary effect size (Figure 11). No evidence of publication bias for the sample of included articles was found using Rosenthal's Fail-safe $N = 7{,}007$, $p_{TARGET} = .05$ nor Orwin's Fail-safe $N = 18$, $\alpha_{\text{TARGET}} = -.30$. The value $r = -.30$ was selected for the r_{TARGET} as a generally accepted threshold at which correlations between psychometric instruments are considered to be meaningful.

Mean age of the participants, available for 15 (88%) of the studies, was found to be a

statistically significant moderator of the variability in the reported correlations between the BRS and measures of perceived stress and accounted for 23.92% of the observed variability in this relation (n = 15, df(1), p = .042, R^2 = .239, Figure 12). Geographic location of participants (country) was also found to be a significant moderator between the BRS and measures of perceived stress and accounted for 55.49% of the observed variability in this relation (n =17, df(4), p = .001, R^2 = .555). For each sample type, the means and standard deviations were: Samples from Australia (n = 4 reported correlates, M_r = -.42, SD_r = .12), samples from New Zealand (n = 2, M_r = -.60, SD_r = .06), samples from the United Kingdom (n = 1 reported correlate, r = -.24), samples from the United States (n = 9 reported correlates, M_r = -.59, SD_r = .09), and one sample was an online study where authors indicated participants were recruited through both Qualtrics and social media (e.g., Facebook, Reddit) and no regional limiters were used, (r = -.61, n = 1).

Sample category (using the five categories described in previous analyses) and student status were not found to be significant moderators of the correlation between participants BRS scores and their scores on measures of perceived stress. The variable self-reported sex/gender identity could not be used for moderator analyses as insufficient data were available. Only five (29.4%) of the samples included sex/gender data for the participants described in the correlational analyses.

Discussion

The goals of this research were to conduct, for the BRS, (a) reliability generalization analyses for estimates of internal consistency, (b) validity generalization analyses for estimates of convergent validity and concurrent validity, and (c) moderator analyses to examine sample characteristics that might account for possible heterogeneity of reliability and validity estimates.

Reliability Generalizations

Internal Consistency. Of the 149 studies that used the BRS that met our inclusion criteria, internal consistency estimates (Cronbach's α) were available for $n = 86$. Hunsley and Mash (2018) suggest it is appropriate to interpret Cronbach's α values of .7 as *adequate*, α = .8 as *good* and α = .9 as *excellent* internal consistency and suggest that sets of data with α values below .6 are not considered to be reliable enough to interpret meaningfully. The 86 reported Cronbach's α estimates in this set of meta-analyses ranged from α = .363 to .94. The distribution, using the aforementioned interpretation guidelines was: *below adequate,* 2.3% (2 samples); *adequate*, 14.0% (12 samples); *good*, 58.1% (50 samples); and *excellent*, 25.6% (22 samples). Using a random effects maximum likelihood (REML) model, as was appropriate for the observed significant heterogeneity of these data, a summary effect size (estimated mean internal consistency) for the BRS was calculated. The resulting value of Cronbach's α = .863 (CI_{95} [0.85, 0.87]) falls fully within the *good* range. This evidence is adequate to support selecting the measure for use with samples similar to those in the included studies (i.e., English-speaking adults, as most of the evidence in this study coming from participants located in Australia, Canada, New Zealand, the United Kingdom, and the United States). However, as with all unidimensional psychometric instruments, strong prior evidence of internal consistency does not negate the need to compute and report this reliability value for each administration of any psychometric instrument.

Significant heterogeneity (variation) was found for these data, therefore it was possible to assess potential moderators of the observed variation. The percentage of male participants and percentage of transgender participants were found to be significant moderators of internal consistency estimates, but only accounted for 1.37% and 1.60% of the variation, and the

percentage of female participants and other-identified participants accounted for 0% of the variability. Thus, in these data the moderating effects of self-identified sex/gender identity were almost inconsequential. Similarly, location (sorted by country) also explained an inconsequential amount of the variation 1.9%. The lowest observed internal consistency estimate, $\alpha = .36$, was found in the only sample where the BRS was used Rwandan participants. As this is a single observation, no conclusion can be drawn but it is possible that the *below adequate* internal consistency value was due to cultural or language considerations, as the instrument was developed based on North American conceptualizations of resilience and developed in the English language. Rwandan culture differs substantially from the United States and Rwanda has three official languages (Kinyarwanda, English, and French) therefore participants may not have been completing the measure in their first language, or cultural considerations may have impacted the applicability of the instrument items to the residents of this region. Sample category was also a statistically significant moderator of the variation in internal consistency scores. However, the mean α range across groups was minimal (.82 to .89) and accounted for only 8.4% of the variance. Thus, although statistically significant likely due to the number ($n = 86$) of internal consistency estimates included the analysis, all of these fall into the range of *good* and it is unlikely that small observed differences such as these are clinically or practically meaningful. Neither the mean age nor student status of participants significantly moderated observations of internal consistency.

Test-Retest Reliability. There were insufficient data to conduct analyses to assess test-retest reliability, as such data were only available for three samples. Two of these samples were Intra-class Correlation Coefficients (ICCs) for the validation sample for the development of the BRS: $ICC_{1\ month} = .62$ in a sample of university students, and $ICC_{3\ months} = .62$ in a sample of

female medical patients. Another study reported Pearson's $r_{3\ months}$ = .80 in a sample of employed adults. Given the variability in timeframe (one sample with a one-month time-delay between administrations and two samples with a time-delay of three months) these data, if combined for meta-analyses could not be interpreted meaningfully. Given the lack of data available there is insufficient information to form conclusions related consistency of measurement for periods longer than three months.

Further, although it is possible that both the underlying resilience construct and the measure could yield stable and consistent measurements across time (i.e., resilience does not change with time and the measure yields consistent results), it is also possible that resilience varies over time, or that scores on the instrument may vary across subsequent administrations. Given that the majority of the studies were conducted at a single time point and correlational in nature, additional test-retest reliability studies would need to be conducted to establish a base of evidence for conclusions related to test-retest reliability. Thus, for both pre- and post-test and repeated-measures designs, it would be useful for researchers to consider examining and reporting ICCs for scores on the BRS.

Validity Generalizations

Convergent Validity. Four studies used measures described by the authors as measures of resilience (n = 8 observations). Using a random-effects model, a summary effect size (for Pearson's r correlations ranging from r = .49 to .72) a summary effect size of M_r = .61, CI_{95} [0.55, 0.67] was obtained. A generally accepted benchmark for Pearson's r values to be meaningful in psychological studies is an r value that is greater than or equal to .3, thus the summary effect size for these data meet that benchmark of potential relevance. Pearson (1900) did not offer interpretative benchmarks for this strength of association statistic, however Cohen

(1992) set the most commonly used thresholds to interpret the product-moment correlation as: *r* = .10 to .29 indicates a small effect, *r* = .30 to .49 suggests a medium/moderate effect, and values greater than or equal to *r* = .50 to be a large effect. The summary effect size for these convergent data is in the large effect size range. However, as is often the case in the psychological literature, there is little agreement upon a precise value for an acceptable benchmark value to determine that defines adequate evidence of convergent validity. For example, Campbell and Fiske (1959) suggest that any correlational magnitude that is non-zero where the strength of association between a measure and measures of the same construct exceeds the strength of association between the measure and measures of unrelated constructs constitutes evidence for convergent validity (and in that article, no specific magnitudes for interpretation are given). Conversely, Carlson and Herdman (2012) indicate that, where the correlation is between measures intended to assess similar constructs, values above *r* = .70 are recommended, and correlations between measures intended to measure similar constructs that are "below *r* = .50 should be avoided" (Carlson & Herdman, 2012, p. 19). Abma, Rovers, and van der Wees, (2016) describe the acceptable threshold for convergent validity to be any value greater than or equal to *r* = .50, and Furr and Bacharach (2013) describe that in the behavioural sciences even relatively low convergent validity values (i.e., values less than *r* = .40) can still be acceptable. Finally, Campbell and Fiske (1959) also suggest that low strengths of association for convergent validity may also be due to a) one measure does not adequately capture the construct, b) neither measure adequately captures the construct, or c) the construct is not being measured independent of non-trait attributes captured by the measure (Campbell & Fiske, 1959). It is further possible, that low convergent validity values are found amongst resilience measures due to the differing definitions of resilience that underlie the measures (i.e., each instrument may be measuring a definitionally

unique construct of resilience).

Therefore, based upon the aforementioned recommendations, we suggest that convergent validity greater than $r = .50$ are *acceptable,* and values greater than or equal to $r = .70$ are *ideal.* Thus, the summary effect size produced between the BRS and resilience constructs (inclusive of overlapping constructs such as Ego-resilience, $M_r = .61$) the convergent validity finding is below our newly defined quantitative threshold for *ideal* internal consistency, but falls within the range of *acceptable*. As noted in the results section, if the data are constrained to only examine convergent validity between measures that are defined as instruments to assess resilience (i.e., omitting ego-resilience measures) the convergent validity summary effect size increased to $M_r = .66$). Of the five samples included in this post-hoc exploratory analysis, two were analyses of convergent validity between the BRS and the 25-item Conner-Davidson Resilience Scale (CD-RISC: Connor & Davidson, 2003). The sample of participants with a greater magnitude of observed convergent validity between the BRS and the CD-RISC were older ($r = .72$, mean age = 37 years, $n = 396$) and were not selected for particular characteristics. The sample of participants with the lesser magnitude of observed convergence between the BRS and the CD-RISC were younger ($r = .59$, mean age = 20.4, $n = 128$) and were university students. It is possible that age or sample type (student vs. non-student samples) may moderate this relation, however, two studies do not comprise a pattern, and further studies would be needed to investigate potential moderators of this relation.

Concurrent Validity.

Summary effect sizes. Meta-analyses were conducted to investigate the most frequently reported mental health measures where participants scores on these measures are correlated with their observed scores on the BRS. Summary effect sizes for each investigated mental health

construct were calculated (anxiety, n= 16 studies, r = -.47, CI_{95} [-0.51, -0.43]; depression, n= 18 studies, r = -.49, CI_{95} [-0.54, -0.43]; optimism, n= 18 studies, r = .44, CI_{95} [0.35, 0.53]; social support, n= 17 studies, r = .24, CI_{95} [.20, .29]; and stress, n= 17 studies, r = -.53, CI_{95} [-0.59, -0.46]). Of these, the strength of association between the BRS and social support was the smallest. This may be due to the fact that, unlike the other constructs, social support is not exclusively a feature/inherent characteristic of the individual but is affected by both circumstance and the behaviour of others.

Potential moderators. Among the data-driven moderators that were selected, that also had adequate data to be included in the meta-analysis, sample category accounted for 100% of the variance in the relation between the BRS and measures of social support. Geographic location of participants (country) was found to be a significant moderator between the BRS and measures of perceived stress and accounted for 55.49% of the observed variability in this relation. Mean age of the participants, available for 15 (88%) of the studies, was found to be a statistically significant moderator of the variability in the reported correlations between the BRS and measures of perceived stress and accounted for 23.92% of the observed variability in this relation. As mean age increased, the magnitude of the relation decreased (range 19.22 to 69.13 years, grand mean = 35.08 years, SD = 16.03 years, n = 15, df(1), p = .042, R^2 = .239, Figure 12). This finding suggests that the BRS may be more sensitive to detecting resilience in younger adults but may also suggest that younger adults may perceive (or tend to self-report) higher base-rates of resilience than do older adults. For example, it would be reasonable for populations of university students in the United States who have mental and physical health that is adequate enough to be enrolled in higher education to report feeling more able to “bounce back quickly after hard times” as these individuals may have had less life experience related to adversity than

populations of older individuals who may have been selected for their experience with physical or mental health problems or specific careers with a high exposure to stressors (e.g., military personnel, first responders, and nurses).

No statistically significant moderators were found for to account for the observed variation among the strengths of association between the BRS and measures of anxiety, measures of depression, and measures of optimism among the data-driven moderators that were selected for this study. As there was significant heterogeneity amongst these observed relations (I^2 = 57.0%, 76.8%, 86.5%) and these were in the moderate level of heterogeneity range (50.0%-74.9%) and high level of heterogeneity range (≥ 75.0%) (Higgins & Thompson, 2002). However, it would be erroneous to assume that a lack of statistically significant moderators among apparent and identified potential moderators equates to a lack of a moderation effect (Hak, Van Rhee, & Suurmond; 2016).

It is relevant to note, most of the studies included in this sample lacked sufficient self-identified sex/gender identity information applicable to the reported correlations between the BRS and the mental health measures that were the foci of these analyses. Given that the lifetime prevalence of several clusters of anxiety symptoms and clusters of low mood symptoms is higher in females than males in the United States (CPES and NCS; conducted from 1990 to 1992; McLean, Asnaani, Litz, & Hoffmann, 2011; DSM-5: American Psychiatric Association, 2013), biological sex at birth or self-reported gender identity may explain some of the observed variation. Although this is a possibility, it would also not be appropriate to conclude that a yet-to-be-identified moderator variable would explain enough of the observed variation to be practically relevant as the data may contain many artefacts (i.e., noise) such that each artefact introduces some variation, but the amount of the impact from each artefact is too small to be relevant. These

missing data and potential impacts are further discussed in the limitations section to follow.

Convergent vs. concurrent validity evidence. As described above, the summary effect size for the BRS with other resilience scales was $r = .61$. The summary effect sizes for the strength of association between the BRS and constructs indicative of positive mental health range related to internal experiences were of similar magnitude: optimism, $r = .44$; anxiety $r = -.47$; and depression, $r = -.49$, and perceived stress $r = -.53$. Only the mean correlation with social support ($r = .24$) was notably lower in magnitude than the convergent validity estimate. Although the each of the mental health measures is lower than the observed strength of association between the BRS and other measures of resilience, these results suggest there is significant overlap between the construct of resilience as measured by the BRS with indicators of mental health. When convergent and concurrent validity scores are found to be similar in magnitude, questions that may be useful to consider are: Why is the convergent validity between resilience measures typically lower than an ideal level (i.e., $r = .70$)?; and/or: If the concurrent and convergent validity values are statistically similar, is the BRS measuring something more than overall psychological distress at the time of assessment?

Low convergent validity between resilience measures. Lesser magnitudes of convergent validity may also indicate that a) either the BRS or the comparator measure does not adequately capture the construct, b) neither measure adequately captures the construct, or c) the construct is not being measured independent of the influence of contextual variables/moderators, or d) low convergent validity values may be due to the inconsistency in definitions of resilience across measures.

The developers of the BRS indicate that the focus of the BRS differs from other resilience measures in that it conceptualizes resilience as a unidimensional characteristic of

focused specifically on the perseverant ability to bounce back (i.e., recover from stress, and "make it through difficult times with little trouble"), and that other resilience measures such as the Connor Davidson Resilience Scale (CD-RISC: Connor & Davidson, 2003) and the Resilience Scale (RS: Wagnild & Young, 1993) capture a broader range of personality characteristics and define resilience as a multi-dimensional cluster of characteristics distributed across two or more factors (Smith et al., 2008, p.196). These differentials – unidimensional vs. multidimensional structure and depth vs. breadth in the theoretical structure and definitions of resilience that inform resilience measurement – may be why convergent validity values are less than what would be typically expected for two measures assessing the same construct. Previous systematic reviews of resilience measures consistently describe that a notable challenge to the study and measurement of resilience is lack of definitional and theoretical consistence amongst measures of this construct (Ahern, Kiehl, Sole, & Byers, 2006, O'Neal 1999; Pangallo, Zibarras, Lewis, & Flaxman, 2015; Salisu & Hashim, 2017; Smith-Osborne & Bolton, 2013; Wagnild & Young, 1991; Windle, Bennet, & Noyes, 2011). In short, a lack of consistency in the definition of resilience as a construct likely explains the lack of convergent validity between resilience measures.

High concurrent validity with indicators of mental health. One conceptualization of resilience is that it is an enduring characterological trait (either a unidimensional aspect of personality or multi-dimensional/multi-factorial cluster of pervasive personality traits, as described in the prior section) that have developed in response to, or are being observed because of, noted experience with adversity, and that the outcome of being resilient is an ability to bounce back from adversity and moderate the negative effects of stress (Campbell-Sills & Stein, 2007; Connor & Davidson, 2003; Sinclair & Wallston, 2004; Smith et al., 2008; Wagnild &

Young, 1993). Block and Kremen (1996) define characterological ego-resilience as the individual's "effectiveness of adaptation, in their ability to equilibrate and re-equilibrate in response to their ever-changing being and the ever-changing world" (p. 349). Within such a conceptual framework, it would be expected that individuals with higher trait resilience would, across time, endorse indicators of mental health and such relations have been noted throughout the resilience literature (Connor & Davidson, 2003; Markstrom & Marshall, 2007; Nishi, Uehara, Yoshikawa, Sato, Ito, & Matsuoka, 2013; Smith et al., 2008). Thus, in accordance with the allostatic load theory of psychological distress and mental health symptoms, it would be expected that individuals who are experiencing less distress are more likely to report feeling more resilient/more able to face and overcome challenges in moments because their personal resources are not depleted (e.g., they are not feeling drained or overwhelmed) (Guidi, Lecente, Sonino, & Fava, 2021).

The observed strengths of association with indicators of mental health are similar to reported for other resilience measures. Test developers reported that perceived stress was significantly correlated with the Resilience Scale (RS: Wagnild & Young, 1993; r = -.32 and r = -.67) and with the 25-item Connor-Davidson Resilience scale (CD-RISC: Connor & Davidson, 2003; r = -.76). Optimism was reported to be correlated with Brief Resilient Coping Scale (BRCS: Sinclair & Wallston, 2004; r = .50, r = .40). Social support was reported to be positively correlated to the 25-item Connor-Davidson Resilience scale (CD-RISC: Connor & Davidson, 2003; r = .36). Symptoms of depression were found to be negatively correlated with Brief Resilient Coping Scale (BRCS: Sinclair & Wallston, 2004; r = -30). Hu, Zhang, & Wang (2015) conducted a meta-analysis that included 60 studies that administered characterological measures of resilience, and the authors combined data for all age groups (children, adolescents, and

adults). They identified 76 observations of strengths of association between resilience and life satisfaction or positive affect (which they aggregated and categorized as negative indicators of mental health), and 35 observations of strengths of association between resilience and depression, anxiety, and negative affect (which they aggregated and categorized as positive indicators of mental health). The summary effect sizes for positive and negative mental health indicators were found to be r = -.36, r = .50, respectively (Hu et al., 2015). In terms of concurrent validity with mental health indicators, consistent with the resilience literature, the data from the present study support that: individuals who are resilient experience (or report) fewer mental health symptoms, and/or individuals with fewer mental health symptoms experience/report feeling more resilient (Barron; 1953; Connor & Davidson, 2003; Duckworth, 2007; Freud, 1927; Kobasa, 1979; Smith et al., 2008; Strycharczyk & Clough, 2015).

A question raised in the course of the present study was whether, given the strength of association between resilience and these positive indicators of mental health, how do we know that resilience is something more than self-reported adequate to good mental health? A common criticism of resilience and resilience theory is that it is a 'circular problem' is that resilience is thought to contribute to good mental health and good mental health is an indicator of resilience. To differentiate resilience as a character trait from indicators of mental health: we can consider that mental health symptoms may be episodic in nature, may persist chronically, or may change in response to intervention, or may increase and decrease due to the presence of certain stimuli; and that resilience as a personality trait and overcoming adversity should remain fairly consistent across the life span and should be pervasive across contexts (American Psychiatric Association, 2013; World Health Organization, 2020). The 'circular problem' critique also does not consider the role of adversity in the development (and observation) of resilience nor that characterological

traits are theorized to be consistent and enduring. Thus, adequate data are not available to support that that resilience is invalid/circular, but there is also not much data available at this time to counter that point, because most of the studies that use resilience measures are cross-sectional and data observations represent a single time-point.

A given sample where the relation between resilience and mental health is observed at a single point in time is likely to contain a) true positives – resilient individuals facing adversity that report indicators of mental health, b) false positives – mentally healthy individuals who report resilience in the absence of having experienced adversity and mentally healthy individuals who over-report/over-estimate their resilience or under-report psychological distress, c) true negatives – individuals low in resilience who report current psychological distress related to adversity, and d) false negatives – individuals low in resilience that report psychological distress in the absence of specific adversity, individuals who under-estimate their resilience or over-report psychological distress, and also individuals that are resilient but at the time of data collection were experiencing a low point that they have not yet "bounced back from". As all of the analyses (and nearly all of the data available) were focused on a single-timepoint / cross-sectional designs, causal relations, and neither stability nor change that may occur in these relations across time points can be inferred. Thus, to assess resilience using the definition that resilience is bouncing back or standing strong in response to adversity, longitudinal or repeated measures designs would be needed to draw conclusions related to the transience or stability of resilience, and the consistency, or inconsistency, of the strengths of association between resilience measures observed at multiple points in time.

The observed amount of variation unexplained by the strengths of association found in this sample of studies suggests that although what is being measured by the BRS overlaps/tends

to correlate with a positive outlook, positive situational appraisal, decreased perceived stress, and fewer self-reported mental health symptoms; there is more to resilience than a positive future-focused outlook and a lack of mental health symptoms. As the literature related to characterological resilience would suggest, there may be other personality constructs that relate to and explain resilience, or resilience may be a unique construct of personality independent of these (Block & Kremen, 1996; Seligman, 2019; Smith et al., 2008). Thus, relations between resilience and other personality constructs may be a direction for future resilience research.

Further, the BRS, used in the present meta-analyses, is premised upon the definition that resilience develops in response to experience with adversity, and very few studies included a measure of adversity/adverse experience. For example, 24.2% of the research participants in the concurrent validity meta-analyses were undergraduate university students, with age means for these studies ranging from 19.22 years to 21.09 years. Although is it possible that some of these research participants have overcome significant life events and adversity, these individuals represent a younger subset of the population, and are likely to be individuals with adequate enough physical and mental health, and adequate enough external resources to be enrolled in post-secondary education programs, and as such this segment of the population that is younger, healthier, and socio-economically privileged may be over-represented in the resilience research literature (Black, Stone, Hutchinson, & Suarez, 2007; McClellan, 2014). Without measurement of adversities that have been overcome, a number of false positives in the detection of resilience may comprise some of the noise/unexplained variation in these data (item b, above). Therefore, to ensure that studies of resilience are capturing resilience specifically, and not just indicators of mental health, resilience research could also include some measure of adversity. For example, certain types of adversity could be assessed using a life events checklist such as the LEC-5 series

of instruments (Weathers, Blake, Schnurr, Kaloupek, Marx, & Keane, 2013). Alternately, adversity could be assessed by examining participants current allostatic load, current stressors, and quality of life, as was done in the early hardiness studies (Guidi et al., 2021; Funk, 1992; Kobasa, 1979). Adversity might also be examined through a socio-cultural lens by assessing privilege, intersectionality effects, and/or many other potential indicators (Black et al., 2014). One meta-analytic study did include a binary indicator ("yes"/"no") for adversity (though the study included children and adults and no operational definition was provided to clarify what was, and was not, coded as adversity); and reported that the associations between characterological measures of resilience and mental health indicators were significantly moderated by adversity. Thus, temporal data, (i.e., longitudinal designs to study resilience), inclusive of clearly defined assessment for adversity are needed in order to have adequate evidence to conclude that resilience differ from in-the-moment indicators of adequate mental health. That noted, because the strengths of association between characterological measures of resilience and mental health indicators are not higher (i.e., much variance remains unaccounted for), it is possible to hypothesize that there is more to resilience than adequate mental health.

Limitations

Although there were sufficient studies to conduct all of the planned analysis, there was a specific limitation with respect to examining moderation effects. Perhaps the most salient limitation of this research is that for the majority of the analyses self-reported sex/gender identity information was not available. Most study authors provided self-reported sex/gender identity information on their overall participant sample (e.g., n = 300, 150 males, 120 females, 10 transgender, and 5 other and 15 missing data), but they often did not provide the self-reported sex/gender identity information relevant to correlations based on data from less than the full

sample (e.g., n = 200 participants – who could have been any combination of the 300 participants). When sex/gender data is not given for the participants whose scores are described by the reported correlation it is not possible to investigate self-reported sex/gender identity information as a moderator. That impact may be particularly pronounced in instances when resilience was measured only following an intervention where there may have been a sex/gender effect that impacted participant attrition.

Another limitation is that for the moderators that were found on these data collected from previous studies, there is still no guarantee that a moderation will occur in any subsequent research. When there is an absence of significant moderators it is possible that a) other variables moderate this relation that were not investigated by researchers, b) the relation is direct and unmoderated, c) both scores on the BRS and another instrument are moderated by a third variable, or d) the relation is spurious. Which of these applies to specific populations in a given context can only be clarified when additional research and additional data analyses and meta-analyses are conducted to continue to build broader bases of psychometric evidence to describe the BRS and other resilience measurement tools.

Other limitations of the current research is that results sets were limited to publications in peer-reviewed journal articles, theses, and dissertations. Thus, information in the grey literature were not included. Another limitation of the current research is that, for practical and scope-management purposes, we limited our sample to resilience instruments administered in the English language and excluded translated versions of English-language measures. Thus, the summary offered here does not include information related to translations of the BRS and non-English-language administration contexts.

Directions for Future Research

The BRS is one of the most commonly used resilience measures, with cross-sectional studies reporting reliability and validity data. However, there are limited studies on test-retest reliability information for the BRS. The instrument developer reported Intra-class Correlation Coefficients (ICCs) for two samples (Smith et al., 2008). $ICC_{1\ month} = .62$, in a sample of $n = 48$ university students, and $ICC_{3\ months} = .62$ in a sample of $n = 50$ women some of which were chronic pain patients. ICCs can be interpreted to be *adequate* for coefficient values of .70 to .79 for intervals of several days to several months between measure administrations (Hunsley & Mash, 2018).

Only two samples described relations between administrations of the BRS at more than one timepoint (Pearson's $r_{3\ months} = .80$ $n = 208$ employed adults, and the other reported there was no statistically significant difference between BRS scores at T1 and T2 (with a three-month interval and no intervention between), in a sample of 115 university students. Given the lack of test-retest data, and the differential time-delays between instrument administrations, these data could not be combined meaningfully in a meta-analysis. Another challenge to investigating test-retest reliability for any measure is that even when time-delays are consistent in duration, unless there is no intervention, studies may not be suited to combine meta-analytically (as differing interventions may impact pre- and post-test measurements of resilience differentially). Separate, independent studies with no intervention between administration would be better suited for this purpose. However, even in those, a researcher would not be able to hold constant the participants actual amount of resilience as it may also change over time. Although it is a challenge to collect test-retest information related to resilience, there are two reasons why such information would be of utility. First, if test-retest correlations in the absence of an intervention are calculated and the magnitudes of these correlations are high, this finding would support that the measure is

adequately capturing resilience as an enduring characterological trait. Second, information supporting test-retest reliability of the BRS is needed if the instrument were to be used to determine if any change in BRS scores following a stressor or intervention are due to the stressor/intervention (i.e., are due to true changes in the construct being measured) rather than measurement fluctuations over time.

Another direction for future study of the BRS scale would be to conduct studies under-studied or under-serviced populations. For example, no studies were found where this instrument was used with samples of homeless individuals, immigrants, forensic populations/incarcerated individuals, or mental health professionals, and very few collected BRS resilience data for LGBTQIA+ populations (n = 7 studies). All of these populations face unique stressors, and may be uniquely resilient, thus inclusion in research would be useful to fill gaps in the existing literature.

Finally, as outlined in previous sections there is a need for longitudinal studies inclusive of clearly defined adversity assessment in order to assess the hypothesis that resilience, and resilience as measured by the BRS, is measuring an enduring characterological trait, and to determine if that trait is correlated with a measurable "bouncing back" following adversity. To clarify these queries, such studies would also need to include a wide enough range of adversity/non-adversity, and a wide enough range of "bounced back/did not bounce back" if adversity (and or specific types of adversity) can be considered as a potential moderator of the development of resilience. Overall, a thorough exploration of this construct remains to be conducted.

Conclusions

Differential test performance, when it is predictable (i.e., moderated by an observable or

measurable variable), is neither inherently "good" nor "bad." Many instruments perform differentially across populations and this is not inherently problematic, so as long as clinicians and researchers are aware of prior evidence of reliability and validity or norming data that may provide a general idea of how the instrument is likely to perform with specific clients or research participants, and recognize that past evidence for reliability does not apply to subsequent studies. Thus, the assessment of how each instrument performs in each group of study participants is essential to the meaningful interpretation of the scores on the instrument.

To summarize the results of the present research, scores on the BRS generally yield "good" internal consistency values, with little variation due to sample characteristics or study characteristics. There is also consistent convergent validity evidence that the BRS is measuring what other characterological measures of resilience are measuring, but that the associations between BRS scores and scores on conceptually related are consistent with the expectation outlined in the resilience literature. These associations, as noted in our discussion, were observed at a single point in time and because sufficient longitudinal data are not available, conclusions about BRS scores and conclusions about characterological resilience in general are limited to these specific cross-sectional snapshots. As with all characterological measures of resilience, in order to determine if the hypothesis of characterological resilience is supported by the evidence studies that examine the temporal sequence of resilience development are needed. For example, it is necessary to know if adversity was, or is, present and preceded one or more resilient outcomes, or if resilience was, or is, present and was found to be predictive one or more resilient outcomes. It is also necessary to clearly define and measure "bounced back"/"did not bounce back" as an outcome. To do so, we would recommend using at least two resilience instruments: a depth measure such as the BRS that is a unidimensional instrument specifically focused on only

the “bouncing back” aspect of resilience and an established multi-dimensional breadth measure that captures a wider range of criteria hypothesized to comprise characterological resilience as the next steps to further explore resilience, related constructs, and resilient outcomes.

Acknowledgement(s)

Gratitude to Dr. Bruce Smith (BRS developer) for his correspondence.

References

Abma, I., Rovers, M., & van der Wees, P. (2016). Appraising convergent validity of patient-reported outcome measures in systematic reviews: constructing hypotheses and interpreting outcomes. *BMC Research Notes, 9*, 226–226. https://doi.org/10.1186/s13104-016-2034-2

Ahern, N. R., Kiehl, E. M., Sole, M. L., & Byers, J. (2006). A review of instruments measuring resilience. *Issues in Comprehensive Pediatric Nursing, 29*, 103–125. https://doi.org/10.1080/01460860600677643

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). *Standards for educational and psychological testing.* Washington, DC: AERA.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. https://doi.org/10.1176/appi.books.9780890425596

Apple, Inc. (2016). FileMaker Pro 16. [Computer Software] Cupertino, California: Apple, Inc.

Barron, F. (1953). An ego-strength scale which predicts response to psychotherapy. *Journal of Consulting Psychology, 17*, 327-333.

Black, L. L., Stone, D. A., Hutchinson, S. R., & Suarez, E. C. (2007). The Development and Validation of the Social Privilege Measure. *Measurement and Evaluation in Counseling and Development*, *40*, 16–32. https://doi.org/10.1080/07481756.2007.11909803

Block, J., & Kremen, A. M. (1996). IQ and ego-resiliency: Conceptual and empirical connections and Separateness. *Journal of Personality and Social Psychology, 70*, 349-361.

Borenstein, M., L.V. Hedges, J.P.T. Higgins, and H.R. Rothstein. (2009). *Introduction to meta-analysis*. Chicester/West Sussex, UK: John Wiley & Sons, Inc.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin, 56*, 81-105. http://dx.doi.org/10.1037/h0046016

Campbell-Sills, L., & Stein, M. B. (2007). Psychometric analysis and refinement of the Connor-Davidson Resilience Scale (CD-RISC): Validation of a 10-item measure of resilience. *Journal of Traumatic Stress, 6*, 1019-1028. https://doi: 10.1002/jts.20271

Carlson, K., & Herdman, A. (2012). Understanding the impact of convergent validity on research results. *Organizational Research Methods*, *15*, 17–32. https://doi.org/10.1177/1094428110392383

Cohen, J. (1992). A power primer. Psychological Bulletin, 112, 155- 159.

Connor, K. M., & Davidson, J. R. T. (2003). Development of a new resilience scale: The Connor-Davidson resilience scale (CD-RISC). *Depression and Anxiety, 18*, 76–82.

Davidson, J. R. T. (2018). *Connor-Davidson Resilience Scale (CD- RISC) Manual*. Retrieved November 18, 2019, from http://www.cd-risc.com.

DerSimonian, R. & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7*, 177-188.

De Vries, H., Elliott, M., Kanouse, D., & Teleki, S. (2008). Using Pooled Kappa to summarize interrater agreement across many items. *Field Methods*, *20*, 272–282. https://doi.org/ 10.1177/1525822X08317166

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology, 92*, 1087-1101. https://doi.org/10.1037/0022-3514.92.6.1087

Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale

(Grit-S). *Journal of Personality Assessment, 91,* 166-174.

Egger., M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629-647.

Feldt, L., & Charter, R. (2006). Averaging Internal Consistency Reliability Coefficients. *Educational and Psychological Measurement*, *66*, 215–227. https://doi.org/10.1177/0013164404273947

Fisher, Z., Tipton, E., & Zhipeng, H. (2017). Robumeta: Robust variance meta-regression. R package version 2.0. https://CRAN.R-project.org/package=robumeta

Friborg, O., Barlaug, D., Martinussen, M., Rosenvinge, J. H., & Hjemdal, O. (2005). Resilience in relation to personality and intelligence. *International Journal of Methods in Psychiatric Research, 14*, 29–42. https://doi.org/10.1002/mpr.15

Freud, S. (1927). *The ego and the id.* Honolulu, HI: Hogarth Press.

Funk, S. C. (1992). Hardiness: A review of theory and research. *Health Psychology, 11*, 335-345.

Furr, R. M., & Bacharach, V. R. (2013). *Psychometrics: An introduction (Second edition).* SAGE.

Guidi, J., Lecente, M., Sonino, N, & Fava, G. A. (2021). Allostatic load and its impact on health: A systematic review. *Psychotherapy and Psychosomatics, 90*, 11-27. https://doi: 10.1159/000510696

Hak, T., Van Rhee, H. J., & Suurmond, R. (2016). *How to interpret results of meta-analysis*. *(Version 1.3)*. Rotterdam, The Netherlands: Erasmus Rotterdam Institute of Management. www.erim.eur.nl/research-support/meta-essentials/downloads

Hathaway, S. R., & McKinley J. C. (1942). *Manual for the Minnesota Multiphasic Personality Inventory.* Minneapolis, MN: University of Minnesota Press.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., Altman, D. G. (2009). Measuring inconsistency in meta-analyses. *British Medical Journal, 237*, 557-560

Howell, R., & Shields, A. (2008). The file drawer problem in reliability generalization: A strategy to compute a Fail-Safe N with reliability coefficients. *Educational and Psychological Measurement*, *68*, 120–128. https://doi.org/10.1177/0013164407301528

Hu, T., Zhang, D., & Wang, J. (2015). A meta-analysis of the trait resilience and mental health. *Personality and Individual Differences*, *76*, 18–27. https://doi.org/10.1016/j.paid.2014.11.039

Huedo-Medina, T., Sanchez-Meca, J., Marin-Martinez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: *Q* statistic or I^2 index? *Psychological Methods, 11,* 193-206.

Hunsley, J. & Mash E. J. (2018). *A guide to assessment that work*. (2nd ed.). New York, NY: Oxford University Press.

Hunter, J. E. & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error bias in research findings*. (2nd ed.). Thousand Oaks, CA: Sage Publications.

IBM Corp. (2011). IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp.

Jascó, P. (2004). Savvy searching: Citation searching. *Online Information Review, 28*, 454-460. Retrieved from http://search.proquest.com/docview/194484842/

Kobasa, S. C. O. (1979). *Stress, personality, and health: An overlooked possibility.* (Unpublished doctoral dissertation). University of Chicago, Chicago, Illinois, USA.

Markstrom, C. A., & Marshall, S. K. (2007). The Psychosocial Inventory of Ego Strengths: Examination of theory and psychometric properties. *Journal of Adolescence, 30*, 63-79. https://doi: 10.1016/j.adolescence.2005.11.003

McClellan, M. J. (2014). *Awareness of Privilege and Oppression Scale-2: Construction and initial validation*. Theses and Dissertations - Educational, School, and Counseling Psychology. https://uknowledge.uky.edu/edp_etds/22

McLean, C. P., Asnaani, A., Litz, B. T., & Hofmann, S. G. (2011). Gender differences in anxiety disorders: Prevalence, course of illness, comorbidity and burden of illness. *Journal of Psychiatric Research, 45*, 1027–1035. https://doi.org/10.1016/j.jpsychires.2011.03.006

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-Analyses: The PRISMA Statement. *PLoS Med, 6*. 1-9. https://doi.org/10.1371/journal.pmed1000097

Moher, D., Shamseer, L., Clarke, M, Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., & Stewart, L. A. (2015). Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement. *Systematic Review, 4*. 1-9. https://doi.org/: 10.1186/2046-4053-4-1

Nishi, D., Uehara, R., Yoshikawa, E., Sato, G., Ito, M., & Matsuoka. (2013). Culturally sensitive and universal measure of resilience for Japanese populations: Tachikawa Resilience Scale in comparison with Resilience Scale 14-item version. *Psychiatry and Clinical Neurosciences, 2013, 67*, 174-181. https://doi:10.1111/pcn.12028

O'Neal, M. R. (1999). *Measuring resilience*. (Unpublished doctoral dissertation). Retrieved from https://search-proquest-com.proxy.bib.uottawa.ca/docview/62411455?accountid=14701

Orwin, R. (1983). A Fail-Safe N for effect size in meta-analysis. *Journal of Educational Statistics*, *8*, 157–159. https://doi.org/10.2307/1164923

Pangallo, A., Zibarras, L., Lewis, R., & Flaxman, P. (2015). Resilience through the lens of interactionism: A systematic review. *Psychological Assessment, 27,* 1-20.

https://doi.org/10.1037/pas0000024

Pearson, K. (1900). X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series 5*, 157-175. https://doi.org/10.1080/14786440009463897

Pigott, T. D. (2012). *Advances in meta-analysis.* New York, NY: Springer.

Quintana D. S. (2015). From pre-registration to publication: A non-technical primer for conducting a meta-analysis to synthesize correlational data. *Frontiers in Psychology, 6*, 1-9. https://doi.org/10.3389/fpsyg.2015.01549

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Rosenthal, R. (1979). The "File Drawer Problem" and tolerance for null results, *Psychological Bulletin, 3, 638-641*. http://dx.doi.org/10.1037/0033-2909.86.3.638

Salisu, I., & Hashim, N. (2017). A critical review of the scales used in resilience research. *IOSR Journal of Business and Management, 19*, 23-33.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology, 62*, 529-540.

Seligman, M. E. (2019). Positive Psychology: A Personal History. *Annual Review of Clinical Psychology*, *15*, 1–23. https://doi.org/10.1146/annurev-clinpsy-050718-095653

Sinclair, V. G., & Wallston, K. A. (2004). The development and psychometric evaluation of the Brief Resilient Coping Scale. *Assessment, 11,* 94-101

Smith, B. W., Dalen, J., Wiggins, K., Tooley, E., Christopher, P., & Bernard, J. (2008). The Brief Resilience Scale: Assessing the ability to bounce back. *International Journal of*

Behavioral Medicine, 15, 194-200. https://doi.org/10.1080/10705500802222972

Smith-Osborne, A., & Bolton, K. W. (2013). Assessing resilience: A review of measures across the life course. *Journal of Evidence-Based Social Work, 10*, 111-126.

Thompson, B. & Vacha-Haase, T. (2000). Psychometrics is datametrics: The test is not reliable. *Educational and Psychological Measurement, 60*, 174-195. https://doi.org/10.1177/0013164400602002

Strycharczyk, D., & Clough, A. (2015). *Developing mental toughness: Coaching strategies to improve performance, resilience, and wellbeing.* London, UK: KoganPage.

Urbina, S. (2004). *Essentials of psychological testing*. Hoboken, NJ: John Wiley & Sons, Inc.

Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement*, *58*, 6–20.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*, 1–48. https://www.jstatsoft.org/v036/i03.

Wagnild, G., & Young, H. M. (1991). Another look at hardiness. *Journal of Nursing Scholarship, 23*, 257-260. https://doi.org/10.1111/j.1547-5069.1991.tb00682.x

Wagnild, G. M., & Young, H. M. (1993). Development and psychometric evaluation of the Resilience Scale. *Journal of Nursing Measurement, 1*, 165-178.

Weathers, F. W., Blake, D. D., Schnurr, P. P., Kaloupek, D. G., Marx, B. P., & Keane, T. M. (2013). *The Life Events Checklist for DSM-5 (LEC-5)*. Instrument available from the National Center for PTSD at www.ptsd.va.gov

Windle, G., Benett, K. M., & Noyes, J. (2011). A methodological review of resilience measurement scales. *Health and Quality of Life Outcomes, 9*, 1-18.

https://doi.org/10.1186/1477-7525-9-8

Wood, J. (2007). Methodology for dealing with duplicate study effects in a meta-analysis. *Organizational Research Methods, 11*, 79–95. https://doi.org/10.1177/1094428106296638

World Health Organization (2020). *International statistical classification of diseases and related health problems (11th ed.)*. https://icd.who.int/

Project 2 Tables and Figures

Table 1. Estimated Summary Effects Sizes by Correlated Construct

			CI95 estimate			Heterogeneity of effects			
Construct	k	Mean *r*	LB	UB	*p* model	*Tau*-squared	*Q*	*p(Q)*	I-squared
Anxiety	16	-0.47	-0.51	-0.43	<.001	0.005	31.23	0.008	57.0%
Depression	18	-0.49	-0.54	-0.43	<.001	0.014	69.15	<.001	76.8%
Optimism	12	0.44	0.34	0.53	<.001	0.005	65.33	0.02	86.5%
Social Support	17	0.24	0.20	0.28	<.001	0.004	33.63	<.001	46.5%
Stress	17	-0.53	-0.59	-0.46	<.001	0.028	149.8	<.001	89.7%

Figure 1. Inclusion/Exclusion Criteria Flowchart

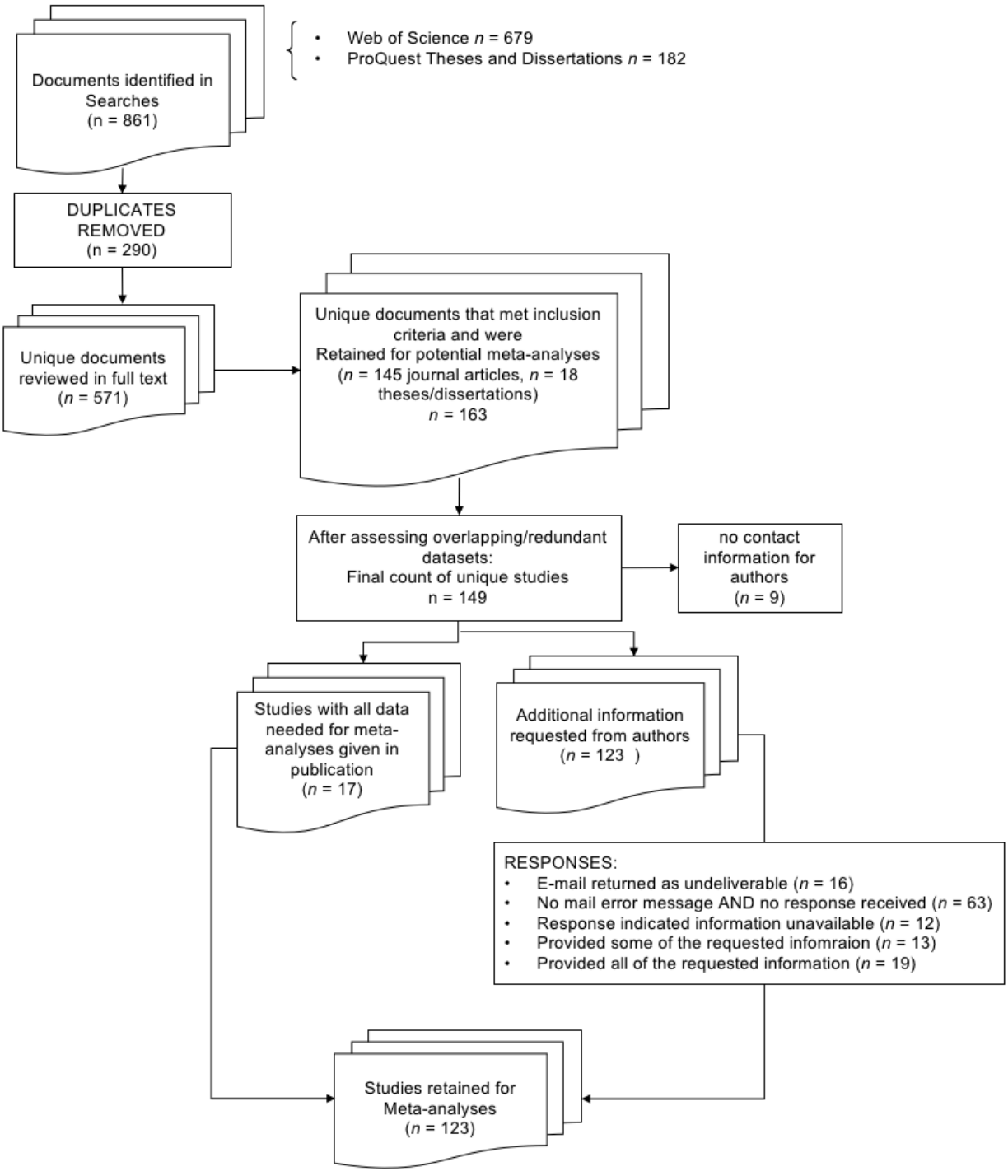


Figure 2. Internal Consistency Forest Plot

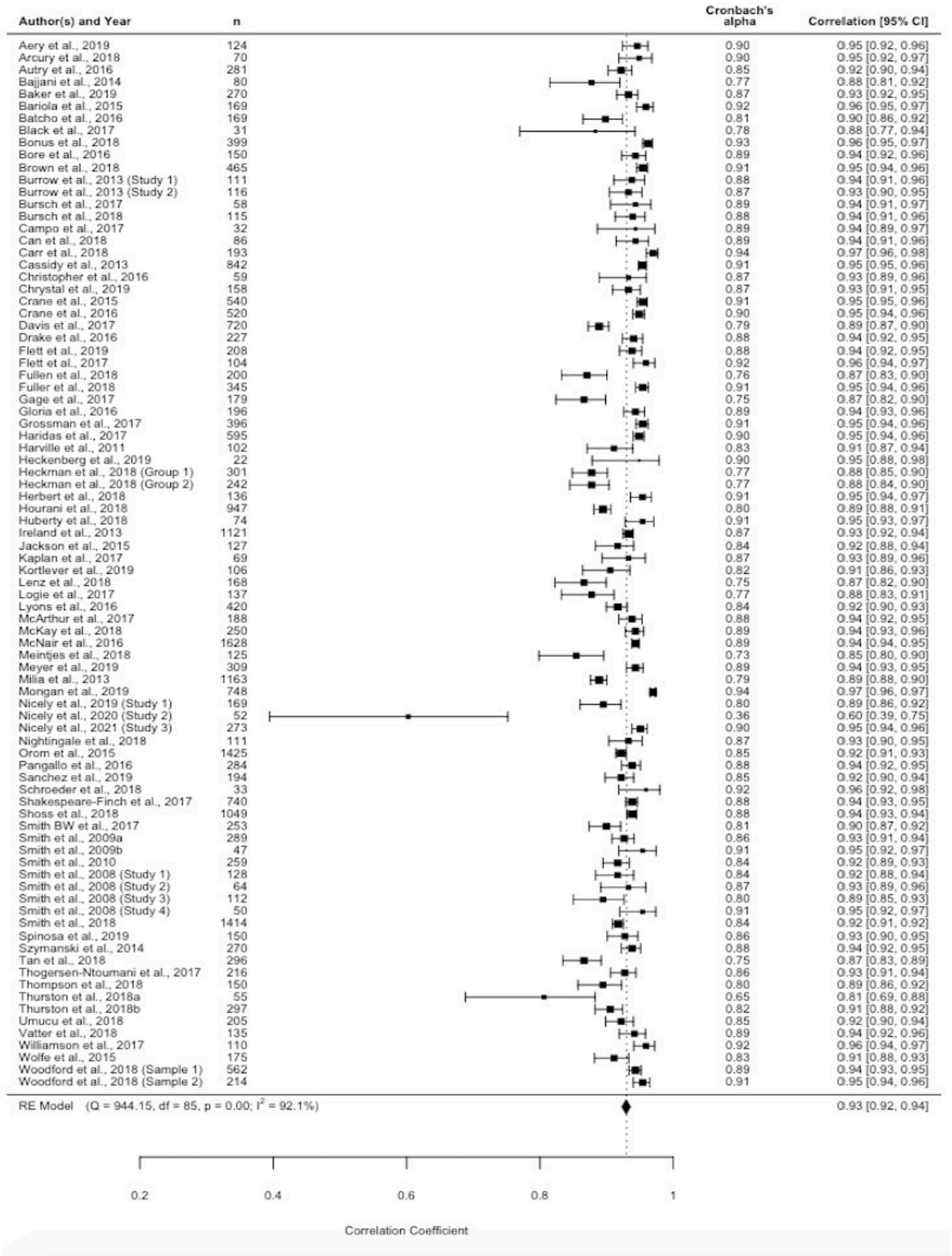


Figure 3. Publication Bias Analysis for Included Internal Consistency Values

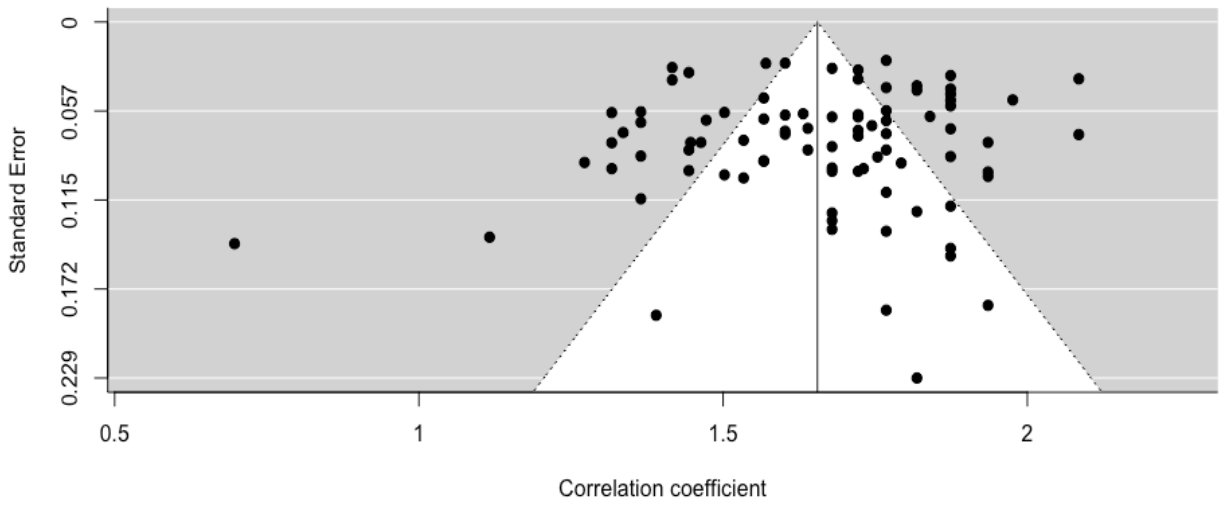


Figure 3. Funnel plot visualization of publication bias analysis, Non-significant asymmetry (Egger's z = 1.02, *p* = .307).

Figure 4. Convergent Validity Forest Plot

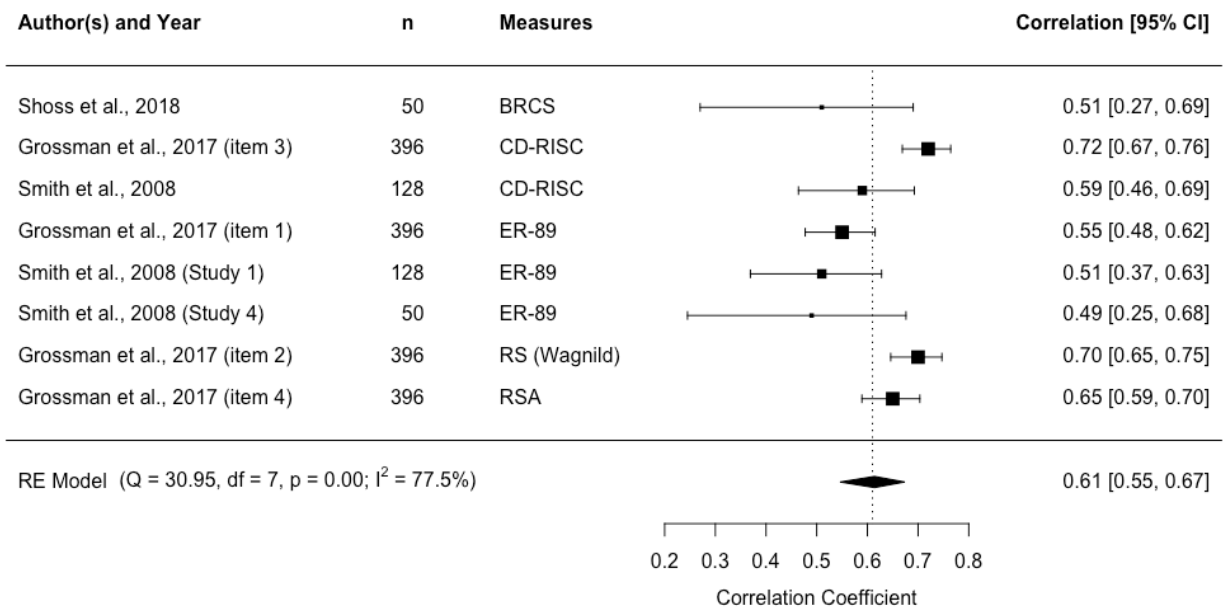


Figure 4. Measure abbreviations: BRCS (Brief-Resilient Coping Scale), CD-RISC (Connor-Davidson Resilience Scale, 25-item version), ER-89 (Ego Resilience scale, 14 item version), RS (Resilience Scale, 25-item version), RSA (Resilience Scale for Adults, 37-item version).

Figure 5. Correlations Between BRS Scores and Measures of Anxiety

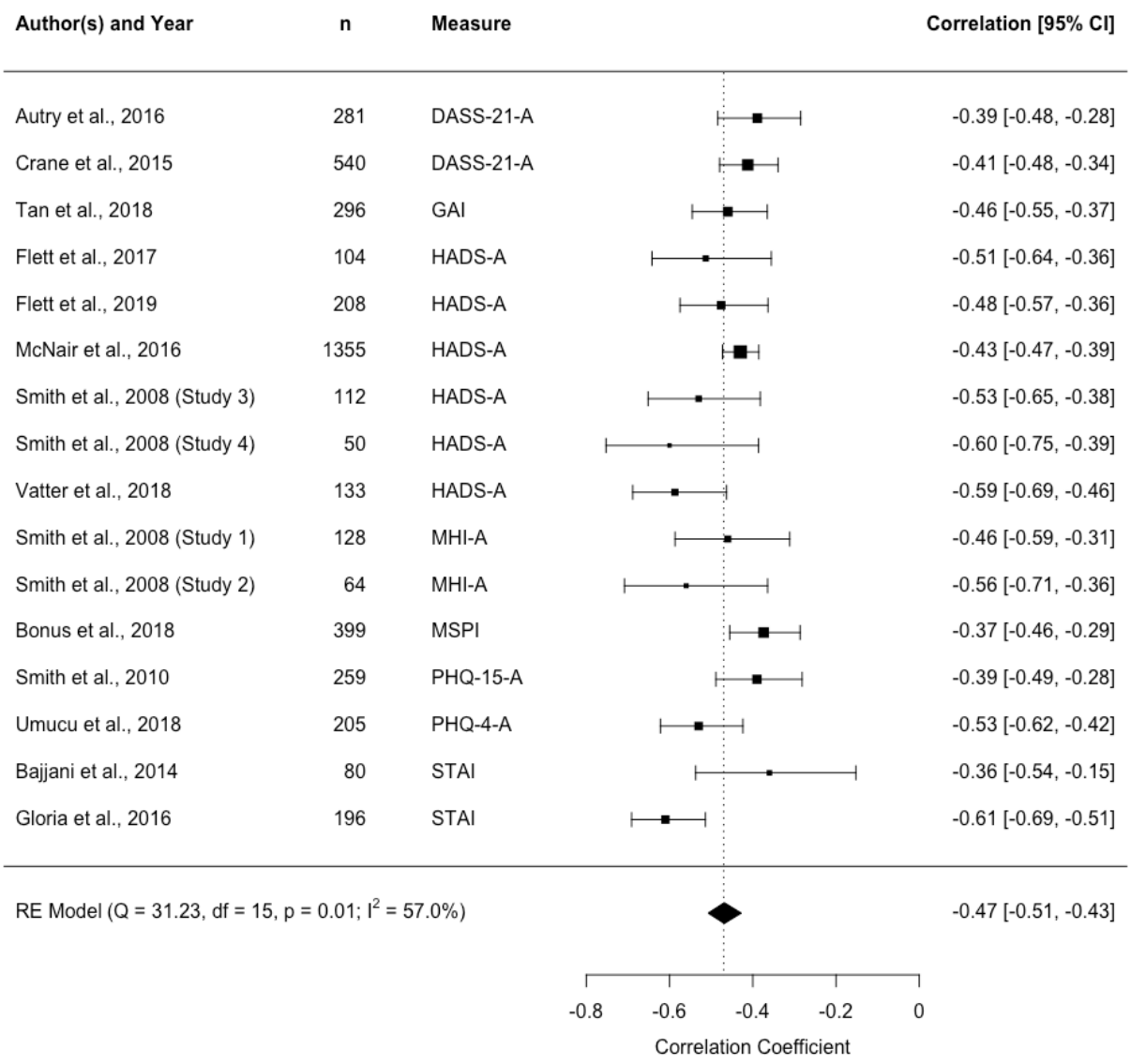


Figure 5. Measure abbreviations: DASS-21-A (Depression Anxiety Stress Scale, Anxiety Subscale from the 21-item version), GAI (Geriatric Anxiety Inventory), HADS-A (Hospital Anxiety and Depression Scale, Anxiety Subscale), MHI-A (Mental Health Inventory, Anxiety Subscale), STAI (State-Trait Anxiety Inventory, 20-item version).

Figure 6. Correlations Between BRS Scores and Measures of Depression

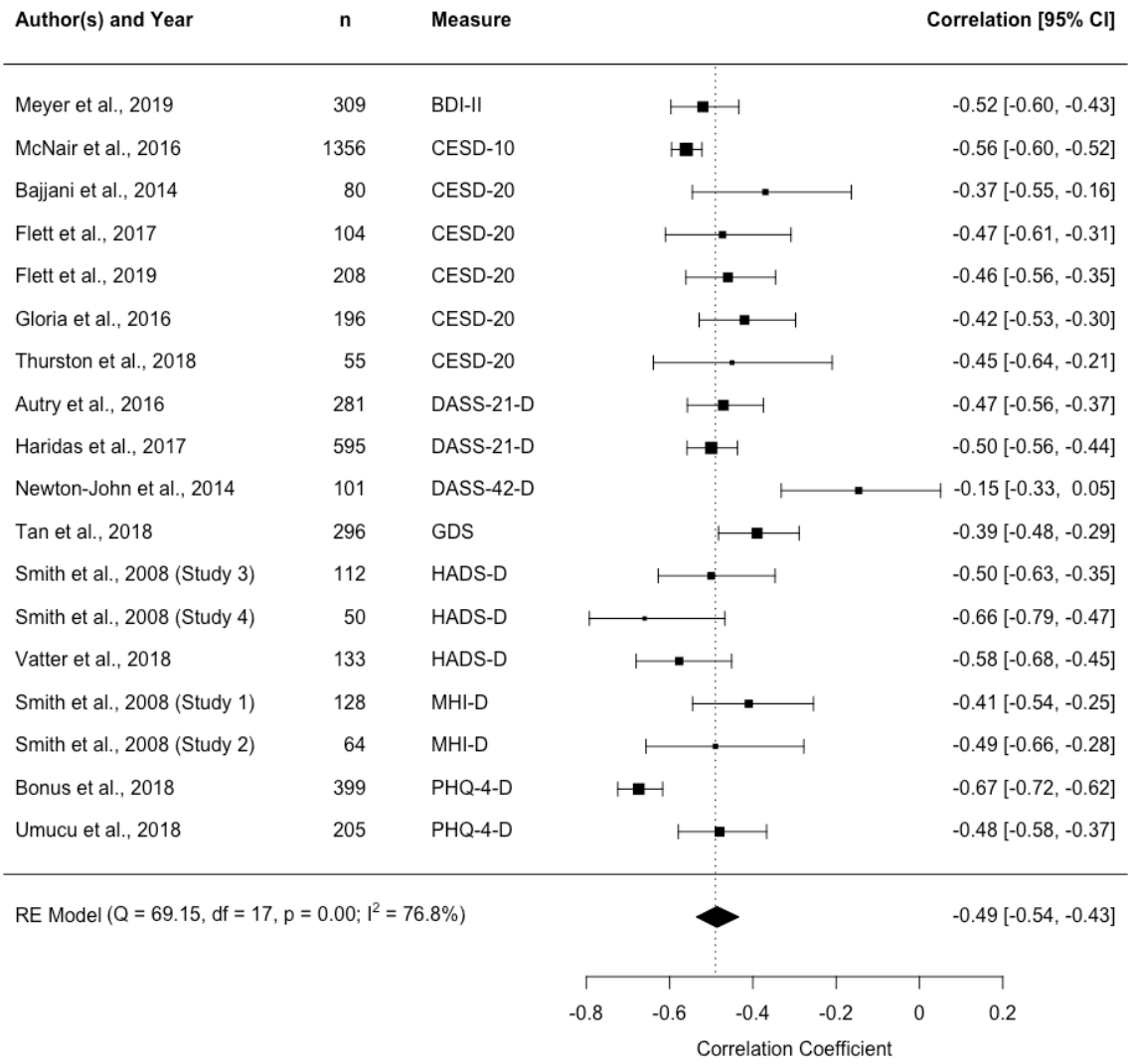


Figure 6. Measure abbreviations: BDI-II (Beck Depression Inventory, version 2), CESD-10 (Center for Epidemiological Studies Depression Scale, 10-item version), CESD-20 (Center for Epidemiological Studies Depression Scale, 20-item version), DASS-21-D (Depression Anxiety Stress Scale, Depression Subscale from the 21-item version), DASS-42-D (Depression Anxiety Stress Scale, Depression Subscale from the 42-item version), GDS (Geriatric Depression Scale), HADS-D (Hospital Anxiety and Depression Scale, Depression Subscale), MHI-D (Mental Health Inventory, Depression Subscale), PHQ-4-D (Patient Health Questionnaires, Depression Subscale from the 4-item version).

Figure 7. Correlations Between BRS Scores and Measures of Optimism

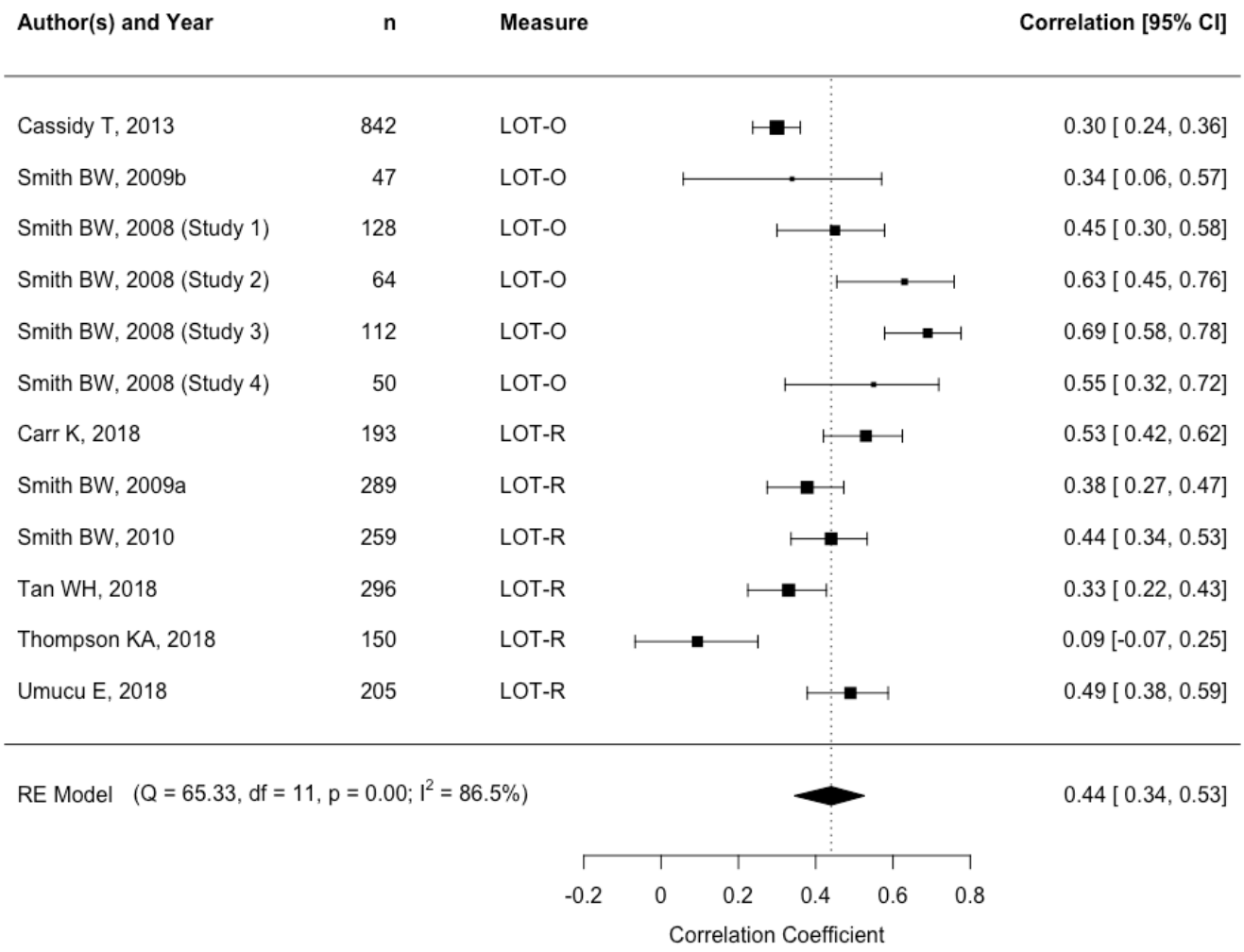


Figure 7. Measure abbreviations: LOT-O (LOT-O Life Orientation Test, 10-item Optimism Subscale), LOT-R (Life Orientation Test - Revised, 10-item version)

Figure 8. Correlations Between BRS Scores and Measures of Social Support

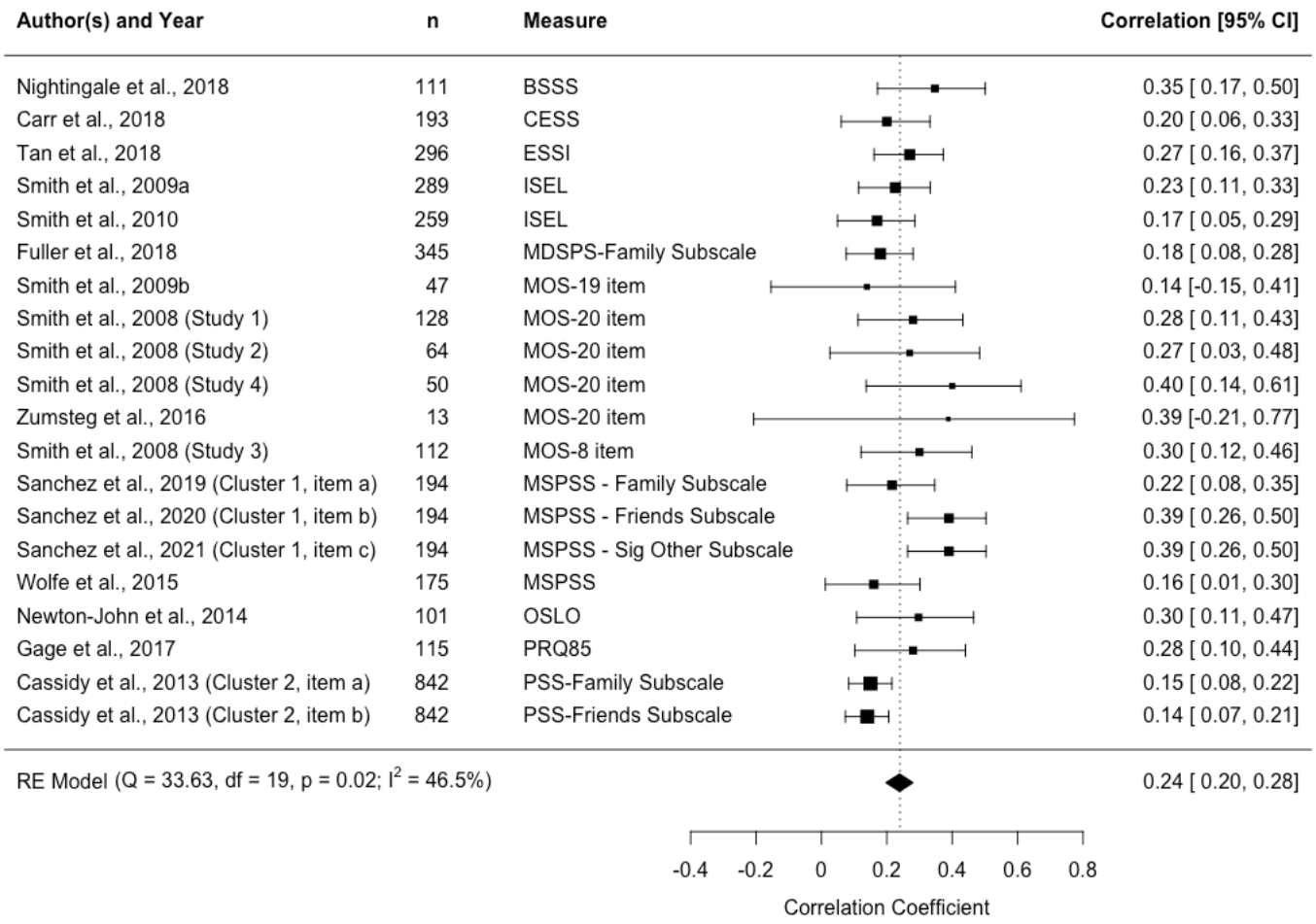


Figure 8. Measure abbreviations: BSSS (Berlin Social Support Scale), CESS (Communication-Based Emotional Support Scale 13-item version), ISEL (Interpersonal Social Support Evaluation List, 12-item version), MDSPS (Multi-Dimensional Scale of Perceived Social Support, Family Support subscale, 1991 version), MOS-19 (Medical Outcome Study Social Support Survey 19-item version), MOS-20 (Medical Outcome Study Social Support Survey, 20-item version), MOS-8 (Medical Outcome Study Social Support Survey, 8-item version), MSPSS-Family (Multidimensional Scale of Perceived Social Support Family Support Subscale, 12-item, 1998 version), MSPSS-Friends (Multidimensional Scale of Perceived Social Support Friends Support Subscale, 12-item, 1998 version), MSPSS-Sig Other (Multidimensional Scale of Perceived Social Support Significant Other Support Subscale, 12-item, 1998 version), PRQ85 (Personal Resource Questionnaire, Social Support Subscale), PSS-Family (Perceived Social Support, Family Subscale), PSS-Friends (Perceived Social Support, Friends Subscale).

Figure 9. Funnel Plot to Examine Potential Publication Bias, BRS ~ Social Support

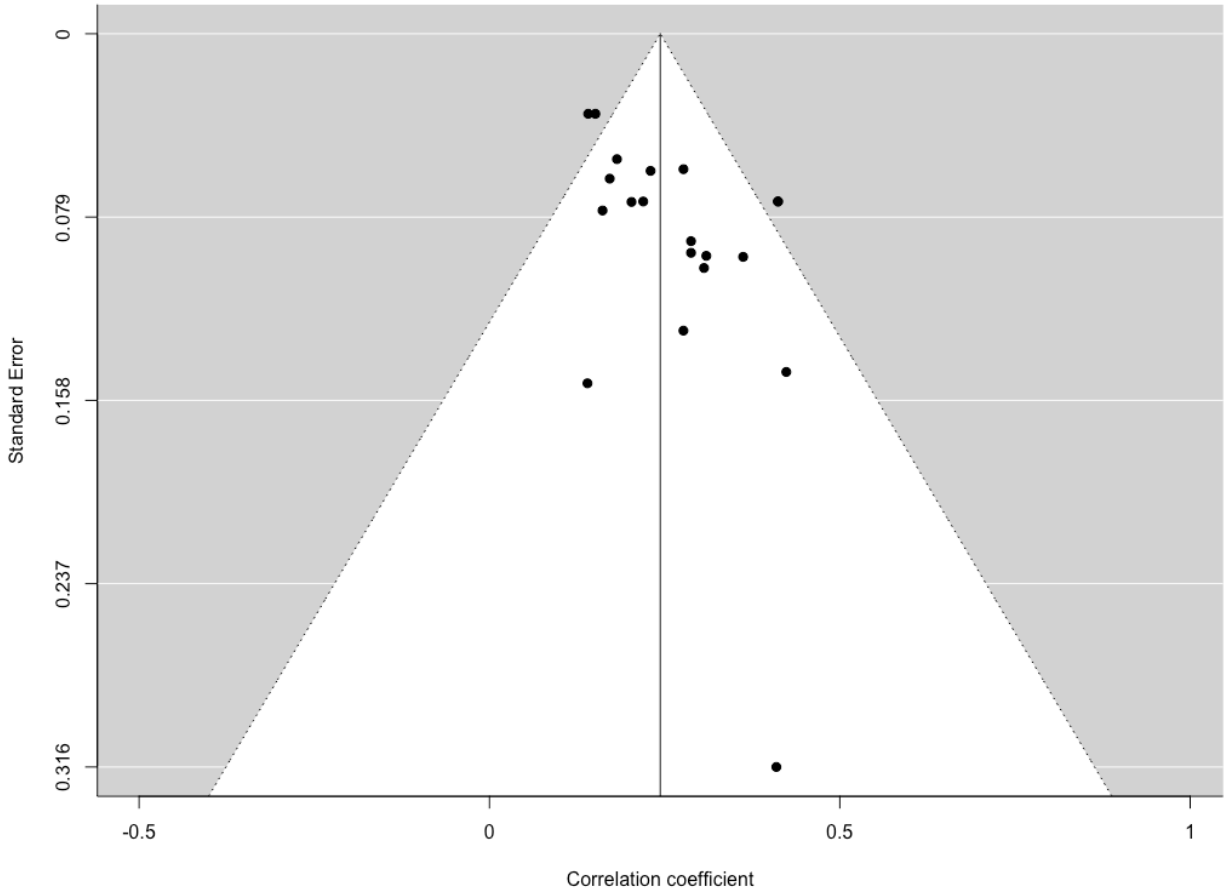


Figure 9. Funnel plot visualization of publication bias analysis, Significant asymmetry (Egger's z = 3.17, *p* = .002).

Figure 10. Correlations Between BRS Scores and Measures of Stress

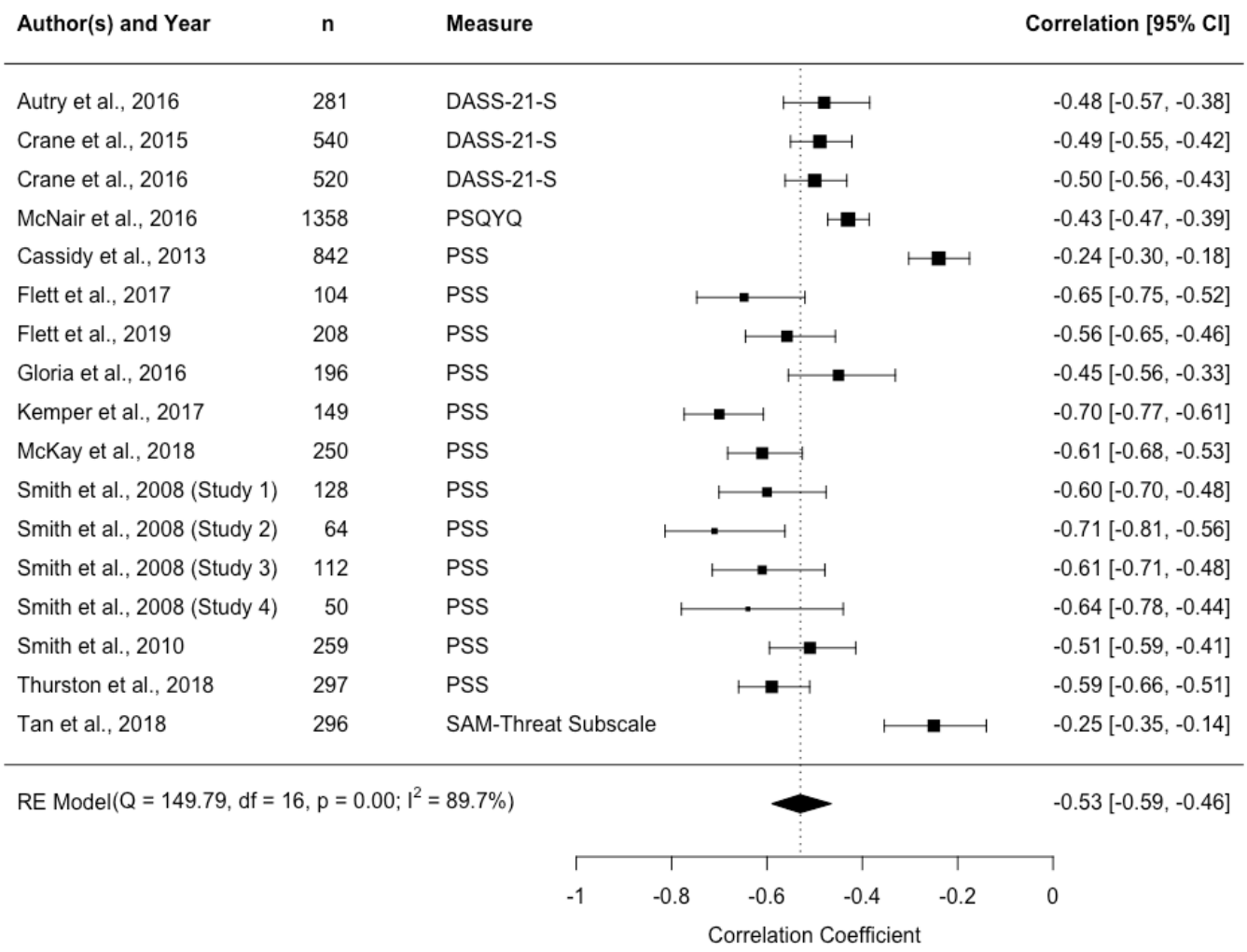


Figure 10. DASS-21-S (Depression Anxiety Stress Scale, Stress Subscale from the 21-item version), PSQYQ (Perceived Stress Questionnaire for Younger Women, 11-item version), PSS (Perceived Stress Scale, 10-item version), SAM-Threat (Stress Appraisal Measure, Threat Subscale)

Figure 11. Funnel Plot to Examine Potential Publication Bias, BRS ~ Stress

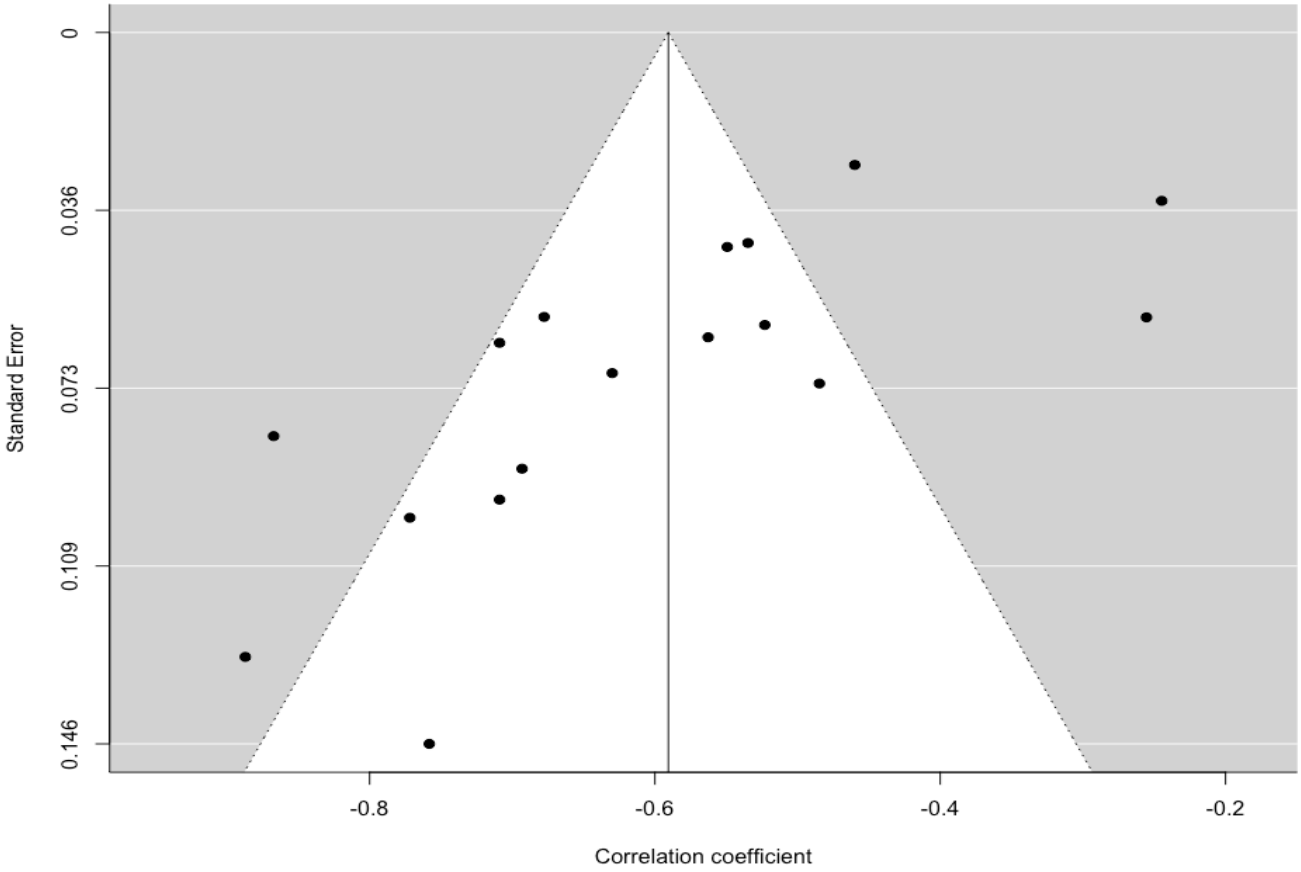


Figure 11. Funnel plot visualization of publication bias analysis, Significant asymmetry (Egger's z = -3.68, *p* = .001).

Figure 12. Mean Age as a Moderator of BRS and Perceived Stress Scores

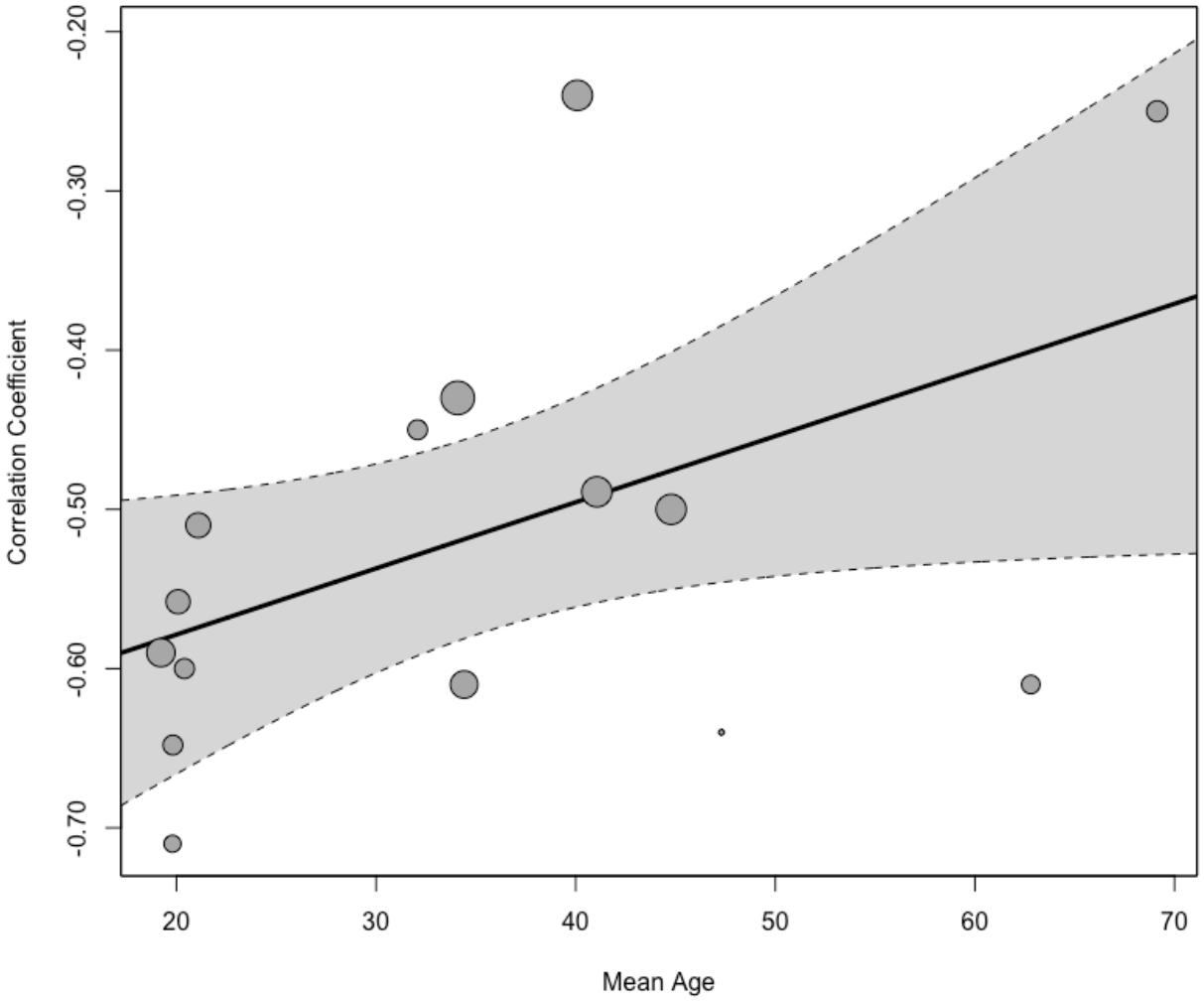


Figure 12. Mean Age (range 19.22 to 69.13 years, grand mean = 35.08 years, SD= 16.03 years) as a significant moderator of the relation between observed scores on the BRS and measures of perceived stress (*n* = 15, *df*(1), *p* = .042, R^2 = .239.

General Discussion

More than a century has passed since William James noticed and sought to define and measure a new construct to explain exceptional outcomes of survival, perseverance, return to homeostasis, and thriving in the following extreme circumstances and/or adversity (James, 1907). Resilience entered the lexicon of psychological scientists in the 1920's and has emerged and re-emerged several times in the form of overlapping, but very similar, constructs: ego-strength, grit, hardiness, mental toughness, and resilience. (Barron, 1953; Freud, 1927; Clough et al., 2007; Funk, 1992; Garmezy, 1974; King et al., 2003; Klohnen, 1996; Kobasa, 1979; Lyons et al., 2009; Maddi, 2002; Padesky & Mooney, 2012). Resilience also has been identified to be a collection of internal and external resources independent of the individual, as a dynamic process independent of character, a static characterological trait (i.e., within a framework that assumes personality is generally consistent in adulthood), and a dynamic characterological trait (i.e., within a framework that assumes personality in adults is malleable/amenable to change.) (Barron, 1953; Bartone, Ursano, Wright, & Ingraham, 1989; Connor & Davidson, 2003; Padesky & Mooney, 2012; Rutter, 2012; Smith et al, 2008; Wagnild & Young, 1993; Windle et al., 2011). Resilience has been treated as both a predictor and a criterion variable (Seligman, 2019; Smith et al., 2008; Kobasa, 1979; Maddi, 2002; Padesky & Mooney, 2012) . Resilience measures have been developed for specific contextual adversities such military combat, minority stress/discrimination, academic achievement and high-stress workplaces (Benishek, Feldman, Shipon, & Mecham, 2005; King et al., 2003; Pollock & Duffy, 1990; Sheard, Golby, & van Wersch, 2009; Smith & Gray, 2009). Instruments have also been developed to measure general resilience as a characterological trait that is defined as resistance to psychological distress and disruption (i.e., equilibration and re-equilibration after experiencing adversity; and

characterological general resilience has been defined both as a unidimensional trait specific only to "bouncing back", and as a multi-dimensional trait comprised of between two and eight resilience personality factors (Barron, 1953; Bartone et al., 1989; Connor & Davidson, 2003; Klohnen, 1996; Smith et al., 2008; Wagnild & Young, 1993).

Critiques and Challenges of Measuring Characterological Resilience

Definition Inconsistency. A common critique in the resilience literature is the substantial inconsistency among definitions of resilience (Funk, 1992; van Breda, 2018). Previous systematic reviews have identified up to 19 measures of resilience and hardiness spanning the lifespan and spanning conceptual definitions (Ahern, Kiehl, Sole, & Byers, 2006; O'Neal, 1999; Pangallo, Zibarras, Lewis, & Flaxman, 2015; Smith-Osborne & Bolton, 2013; Windle, Bennett, & Noyes, 2011). The largest review described 12 measures of adult characterological resilience that had some validation evidence and one that had not yet been psychometrically evaluated (Windle et al., 2011). In Project 1, 51 adult characterological general (i.e., non-context-specific) resilience instruments were identified. By resiliency construct, there were 20 hardiness instruments, 13 resilience instruments, seven ego strength instruments, three ego resilience measures, two grit measures, and three instruments that measured both hardiness and resilience in a single instrument. These ranged in length from four to 106 items, and the number of resilience factors that comprised an overall resilience score ranged from one (for unidimensional measures) to eight (for multi-dimensional measures), although the Three C's (commitment, challenge, and control) was the most frequently used model. These results support the critiques of resilience theory that state that one of the most prominent issues with measuring resilience is the lack of agreement upon a definition for resiliency.

Circular Problem Critique. A common criticism of resilience and resilience theory is

that it is a ‘circular problem;’ in other words, resilience is thought to contribute to good mental health and good mental health is an indicator of resilience. Related to defining characterological resilience, a question raised in the course of Project 2 was whether, given the strength of association between resilience and indicators of mental health, how do we know that resilience is something more than self-reported quality indicative of adequate to good mental health? The ‘circular problem critique’ does not consider three key issues: a) mental health symptoms may be episodic in nature, may persist chronically, or may change in response to intervention, or may increase and decrease due to the presence of certain stimuli; b) resilience as a personality trait should remain fairly consistent across the life span and should be pervasive across contexts, and; c) the role of adversity in the development (and observation) of resilience. Thus, adequate data are not available to support the position that resilience is invalid/circular but, as was found in the series of meta-analyses in Project 2, there is also not much data available at this time to counter that point, because most of the studies that use resilience measures are cross-sectional and data observations represent a single time-point.

As elaborated in the discussion for Project 2, in samples where the relation between resilience and mental health is observed at a single point in time is likely to contain a) true positives – resilient individuals facing adversity that report indicators of mental health, b) false positives – mentally healthy individuals who report resilience in the absence of having experienced adversity and mentally healthy individuals who over-report/over-estimate their resilience or under-report psychological distress, c) true negatives – individuals low in resilience who report current psychological distress related to adversity, and d) false negatives – individuals low in resilience that report psychological distress in the absence of specific adversity, individuals who under-estimate their resilience or over-report psychological distress,

and also individuals that are resilient but at the time of data collection were experiencing a low point that they have not yet "bounced back from". As all of the analyses (and nearly all of the data available) for inclusion in Project 2 were focused on a single-timepoint / cross-sectional designs, causal relations and the stability or change that may occur in these relations across time points cannot be inferred. To assess resilience using the definition that resilience is bouncing back or standing strong in response to adversity, longitudinal or repeated measures designs would be needed to draw conclusions related to a) the transience or stability of resilience and b) the consistency, or inconsistency, of the strength of associations between resilience measures observed at multiple points in time; such data (at least for the BRS that was examined in Project 2) are simply not available.

Proliferation of new measures. In addition to numerous definitions having been developed for resilience, as was found in Study 1 in the last two decades there has been a sharp increase in the number of instruments developed to measure characterological resilience constructs. Some of these instruments (e.g., CD-RISC 10-item and 25-item instruments, Barron Es Scale for the original MMPI) have extensive psychometric information available (Barron, 1953; Block & Kremen, 1996; Campbell-Sills & Stein, 2007; Connor & Davidson, 2003, 2015; Gottesman, 1959; Klohnen, 1996). However, most of the instruments (n = 37, 72.5%) identified in Project 1, as of 2019, have been used in 20 or fewer studies. Of the instruments that were identified as the five most frequently used in Project 1, the Connor-Davidson Resilience Scale, 25-item version, the Barron Ego-Strength Scale, 68-item version, the Brief Resilience Scale, 6-item version, the Connor-Davidson Resilience Scale, 10-item version, and the Short Grit Scale, 8-item version (Grit-8: Duckworth & Quinn, 2009) (Project 1, Figure 6), no meta-analytic summaries of psychometric data were found for the BRS and the Short Grit Scale, (administered

in *n* = 135 studies, and *n* = 83 studies, respectively). Thus, the BRS was selected to be the focus of the meta-analyses in Project 2. However, before I move the discussion to the meta-analyses of the BRS, the results of Project 1 brought to my attention other areas that are problematic in the resilience research literature: researchers altering measures and the lack of adversity and diversity in the participants that are the most common focus in the resilience literature.

Measure Alteration. As described in the results of Project 1, Study 2, 190 articles were located where researchers had used either incomplete resilience instruments (*n* = 140) altered some portion of the wording of items or response options or developed a novel scoring procedure (*n* = 39), or measured resilience using select items from an instrument that they altered (*n* = 11). Most (73.7%, *n* = 140) of these measure administrations occurred in the most recent decade (2010-2019). Three studies were excluded from the meta-analyses in Project 2 due to altered item wording, altered measure scoring, or both. Although the rationale for researchers making changes to measures is often related to ethical considerations (such as shorter measures being a more efficient use of the participants' time or reducing participant fatigue) or is attributed to item relevance (i.e., authors removed items, or altered items they believed were not an appropriate fit for their participants), measure alteration can be extremely problematic.

It is a common misconception that altered, shortened, or revised instruments retain evidence of reliability and validity that have been found in studies that used an original or previous version of the instrument. Data providing evidence of validity do not apply to subsequently revised versions of the instrument (Haynes, Smith, & Hunsley, 2019). When items or responses are altered in any way—be it wording, context, or the addition or removal of responses or items—the instrument effectively becomes a novel scale with no psychometric evidence that scores on the new measure capture the construct of interest reliably, accurately, and

completely (Smith, McCarthy, & Anderson, 2000; Urbina, 2014). Unfortunately, it is not uncommon for researchers to describe an altered version or partial instrument as "reliable and valid" on the basis that scores on the original instrument had evidence of reliability and validity.

Lack of Diversity in Resilience Research. As was described in the results of Project 1 the proportion of studies of resilience and resiliency constructs where participants were university students was 29.1%. For the concurrent validity meta-analyses in Project 2, 24.2% of the research participants were undergraduate university students, and most of these samples of these participants in both projects were located in the United States. Although is it possible that some of these research participants have overcome significant life events and adversity, these individuals represent a younger subset of the population and are likely to be individuals with adequate enough physical health, mental health, and external resources to be enrolled in post-secondary education programs, and as such this segment of the population that is younger, healthier, and socio-economically privileged may be over-represented in the resilience research literature (Black et al., 2007; McClellan, 2014). Socio-economic privilege may be particularly notable in this subset of participants as tuition paid by students for American college and university programs averages $39,045 USD per year (HSBC, 2021).

In addition to noting that some types of samples were highly prevalent, in Project 1 it was also found that some participant categories were under-represented. Very few studies included participants who were: in treatment for substance use or reported substance-use related difficulties (2.8%); forensic/justice-involved participants (.01%), participants dealing with homelessness (.01%), socioeconomically disadvantaged populations/individuals living in poverty (.01%), or individuals who reported having experienced assault or abuse (.01%). Each of these sample categories is a context where research could be beneficial to the planning of programs,

services, and supports for underserved populations. Resilience research related to both visible and non-visible minority status, intersectionality, and the adversities and daily stressors faced by these populations (e.g., ethnic minorities, LGBTQIA+ persons, differently-abled persons) is also fairly rare in the current base of resilience literature.

Lack of Adversity in Resilience Research. As described previously, many of the samples included in present base of resilience research literature are university students. In addition to the possibility that university students may not be representative of populations that have faced and overcome adversity, another salient consideration is that very few resilience studies assess adversity although it would be possible to do so using a measures such as a life events checklist (e.g., LEC-5 series of instruments; Weathers, Blake, Schnurr, Kaloupek, Marx, & Keane, 2013) or by examining participants' current allostatic load, current stressors, and quality of life privilege, intersectionality effects, and/or many other potential indicators (Black et al., 2014; Guidi et al., 2021; Funk, 1992; Kobasa, 1979).

Psychometric Properties of the BRS

The Brief Resilience Scale (BRS: Smith et al., 2008) was selected to be the focus of the meta-analyses in Project 2 because it was the most frequently applied resilience instrument identified in Project 1 that did not have substantial psychometric summary information available. Of the 149 studies that were retained from the literature search, 86 provided (either in the initial publication, or in response to my request) information related to internal consistency. All studies that reported internal consistency statistics reported Cronbach's α. Per Hunsley and Mash (2018) guidelines for the interpretation of Cronbach's α, the summary effect size yielded by the reliability generalization meta-analysis (α = .863 (CI_{95} [0.85, 0.87]) fell within *good* range values. Data were insufficient to assess test-retest reliability, as 96.6% of the included studies

measured resilience at a single point in time and the few studies that did provide test-retest information used differing time intervals between administrations, thus these data could not be combined meta-analytically (e.g., test-retest data ranged from one to three months). Although initial data are promising (ICC = .62, r = .80), too few samples are available to establish a base of evidence for conclusions related to test-retest reliability for the BRS. A summary effect size for convergent validity with other characterologically based measures of resilience was found to be M_r = .61, CI_{95} [0.55, 0.67]; and concurrent validity with mental health indicators were found to be: anxiety, r = -.47; depression, r = -.49; optimism, r = .44; social support, r = .24; and stress, r = -.53. As described in the discussion for Project 2, given that the observed correlations among resilience measure was close to the observed correlations between resilience measures and indicators of mental health, how then do we differentiate resilience from mental health? One explanation I propose is that convergent validity is low between resilience measures related to the inconsistent definitions for resilience (some measures are depth measures, e.g., the BRS, that focus only on measuring the unidimensional trait of bouncing back from adversity, and other measures are breadth measures (e.g., the 25-item CD-RISC that captures five different factors theorized to comprise resilience). Thus these measures are targeting different aspects of resilience and do not share an operational definition for the construct. The concurrent validity summary estimates (as described in the discussion for Project 2) are consistent with similar findings in the resilience literature (Connor & Davidson, 2003; Markstrom & Marshall, 2007; Nishi, Uehara, Yoshikawa, Sato, Ito, & Matsuoka, 2013; Smith et al., 2008).

The Role of Moderators

A fundamental assumption and rationale for conducting meta-analyses of all types of effect sizes (treatment effects sizes, reliability effect sizes, validity effect sizes) is that these will

differ from study to study. All observed variation is assumed to have some degree of heterogeneity that is an aggregate of systematic and unsystematic variation (Field, 2005; Howell, 2014). Statistical artefacts such as sampling error and measurement are the predictable, random, within-group unsystematic sources of error variance (Hunter & Schmidt, 2004). Other artefacts are in some way attributable to 'the ecosystem' of all factors and variables that may impact, influence, or confound the outcome of a study (hence, these are the systematic artefacts). Ideally, error variance reflects only true variation – i.e., the variation is not due to artefacts (Haynes et al., 2011; Hunter & Schmidt, 2004).

In Project 2, the following potential moderators were assessed: Self-reported sex/gender identity (percentage female, male, transgender, and other), mean age of participants, country where study participants were located, sample category (five categories were derived from the authors descriptions of participants: university students, individuals selected for specific careers/vocations/or employment status, individuals selected for specific physical or mental health characteristics, individuals selected for other identified characteristics, and samples of individuals that were not selected for any specific characteristics). The five preceding categories were also aggregated into two superordinate categories based on academic enrollment status at the time the study was conducted (Student / Non-student).

Significant heterogeneity was observed in the reliability generalization analyses of internal consistency. Small, and practically inconsequential amounts of the explained variance were accounted for by sex and geographic location (1.4-1.9% of the explained variation); sample category explained 8.4% of the observed variation in estimates of internal consistency. No significant moderators were found to explain variation in the reported correlations for convergent validity. Of the significant moderator analyses conducted for concurrent validity estimates, the

largest impact was found for sample category (accounting for 100% of the variance in the observed correlations between the BRS and measures of social support). Geographic location and participant age, accounted for 55%, and 24%, of the variation in the observed correlations between the BRS perceived stress.

For the variables where there was an absence of significant moderators it is possible that a) other variables moderate this relation that were not investigated by researchers, b) the relation is direct and unmoderated, c) both scores on the BRS and another instrument are moderated by a third variable. Which of these applies to specific populations in a given context can only be clarified when additional research and additional data analyses and meta-analyses are conducted to continue to build broader bases of psychometric evidence to describe the BRS and other resilience instruments.

Limitations of the Studies

Project 1. A limitation of the Project 1 is that results sets were limited to publications in peer-reviewed journal articles. Thus, information in the grey literature and theses and dissertations were not included. Another limitation of the systematic reviews in Project 1 is that, for practical and scope-management purposes, we limited our sample to resilience instruments administered in the English language and excluded translated versions of English-language measures. Thus, the summaries offered here do not include information related to the total number of resilience instruments available and the absolute frequency of use for all characterological resilience measures.

Project 2. Although there were sufficient studies to conduct all of the planned analysis in Project 2, there was a specific limitation with respect to examining some potential moderation effects. Perhaps the most salient limitation of this research is that for the majority of the analyses

self-reported sex/gender identity information was not available. Most study authors provided self-reported sex/gender identity information on their overall participant sample but often did not provide the self-reported sex/gender identity information relevant to correlations based on data from less than the full sample which could have been any combination of the participants. When sex/gender data is not given for the participants whose scores are described by the reported correlation it is not possible to investigate self-reported sex/gender identity information as a moderator.

Similar to Project 1, a limitation of the Project 2 is that results sets were limited to publications in peer-reviewed journal articles, theses, and dissertations. Thus, information in the grey literature were not included; however, as described in each section, analyses of publication bias were conducted and the results of these analyses do not suggest that the results reported herein were significantly impacted by publication biases or file drawer effects. Further, limiting inclusion to peer-reviewed publications and committee-reviewed theses and dissertations ensures that some amount of quality assurance/review has occurred for the included publications; whereas grey literature may or may not have had any type of review. Another limitation of the current research is that, for practical and scope-management purposes, we limited our sample to resilience instruments administered in the English language and excluded translated versions of English-language measures. Thus, the summary offered here does not include information related to translations of the BRS and non-English-language administration contexts.

Recommendations for Researchers

Based upon the results of Projects 1 and 2, the following recommendations for resilience researchers became more salient.

Statistical and Reporting Considerations. As noted throughout this document, a

common error found in the psychological literature is the assumption is that because an instrument has been found to have adequate psychometric evidence with a similar population in past research (e.g., reasonable internal consistency, test-retest reliability) that these past results apply to current administrations of an instrument. Although it is true that having a sense of how an instrument had performed previously is helpful in determining if a selected measure is likely to be interpretable in a new research study, it is not a guarantee. The absence of 42.2% (even after contacting authors and requesting the missing data) of internal consistency values for the BRS examined in Project 2 indicate that, although internal consistency is being reported more frequently for samples than it has been in the past, this is an area that is still in need of improvement.

Similarly, in Project 2 for most of the meta-analyses it was not possible to assess self-reported sex or gender identity as a potential moderator of unexplained variation in nearly all of the analyses because overall participant count did not match the count of participants described in the reported results. Further, many studies reported only the number of males or only the number of females, with no indication of other sexes/gender identities nor details related to the count of participants where sex/gender data was missing. As described in the limitations the impacts of missing gender data may be particularly pronounced in instances when resilience was measured only following an intervention where there may have been a sex/gender effect that impacted participant attrition.

Instrument integrity is another area of importance. When partial instruments are administered or when instruments are altered (wording of items, response options, novel scoring procedures), prior validation data no longer apply to the “new” instrument. Although some alterations may have little impact, the problem with altering instruments is that these become

uninterpretable in terms of the test developers' recommendations, and in terms of discussing the scores obtained from altered instruments as consistent/inconsistent with the results of previous studies where researchers did not make changes to the measure. This is particularly problematic for multi-dimensional measures, as omitting specific items may significantly alter the breadth or factorial structure of the measure. Essentially, the recommendation for this is that unless there is a very strong reason to do so, the benefits of altering an instrument (i.e., clarify an item, shorten administration time, changing the response scale to match other measures in a given study) typically do not outweigh the interpretation problems caused by this practice.

Longitudinal Research is Needed. The majority of studies in Project 1 (87.7%) and Project 2 (96.6%) measured resilience at a single point in time. Although correlational and cross-sectional study designs can be useful, informative, and save resources (i.e., participant and researcher time, costs to conduct the research), the temporal nature of the construct cannot be examined with such designs. Without longitudinal and repeated-measures designs, particularly in resilience research, many assumptions underlying the research go untested (e.g., that resilience is stable and differs from indicators of mental health at a single point in time, that scores on the measures have adequate temporal stability/test-retest reliability). Multiple time-point studies are especially relevant to the study of the characterological aspects of resilience, given that the term resilience has come to be used describe not only resilience as a static personality trait but also as dynamic variable, a process, and to describe several different types of outcomes.

Adversity. Operational definitions related to the theoretical role of adversity needs to be clearly stated in resilience research. If the research is premised upon the resilience theory that a) adversity must have occurred for resilience to form, or b) resilience indicates that an adversity has/or is being overcome or "bounced back" from, then adversity needs to be measured and

described in a way that is contextually relevant to the research. All too frequently in the resilience research the study involves a sample of university students completing a measure at a single time point, with scores on the measure used to differentiate 'resilient' participants from 'non-resilient' participants, without a discussion of whether adversity led to higher resilience scores, or alternately, whether higher resilience scores led to the overcoming of an adversity. This recommendation is also connected to the prior one in that, although useful to examine a single point in time, single-time point cross-sectional research is not helpful to elucidate temporal patterns or to develop and assess hypotheses related to the predictive validity of resilience instruments.

Under-represented Populations. A final recommendation is to expand the scope of resilience research toward greater inclusivity of typically neglected populations. As described previously, in addition to noting that some types of samples were highly prevalent, in Project 1 it was also found that some participant categories were under-represented. Very few studies included participants who were: in treatment for substance use or reported substance-use related difficulties, forensic/justice-involved participants, participants dealing with homelessness, socioeconomically disadvantaged populations/individuals living in poverty, or individuals who reported having experienced assault or abuse. Each of these sample categories is a context where research could be beneficial to the planning of programs, services, and supports for underserved populations. Resilience research related to both visible and non-visible minority status, intersectionality, and the adversities and daily stressors faced by these populations (e.g., ethnic minorities, LGBTQIA+ persons, differently-abled persons) is also fairly rare in the resilience literature.

Conclusion

Resilience as a character trait has been of interest to researchers in psychology and psychiatry for over a century. Research on the topic began with no clearly defined constructs, no measures, and only questions about how some individuals managed to withstand or "bounce back" from war, earthquakes, and a massive fire that consumed much of San Francisco in 1906. The human condition is rife with adversity, and it always will be – be it daily stressors from a high intensity workplace, systemic problems such as discrimination and poverty, or unexpected conditions such as natural disaster, incidents, and accidents.

At present, a substantial proportion of the resilience literature is focused upon university students in the United States, and very little resilience research is conducted with individuals in treatment for substance use/substance-use related difficulties, forensic/justice-involved populations, participants dealing with homelessness, socioeconomically disadvantaged populations/individuals living in poverty, and individuals who reported having experienced assault or abuse. Resilience research is also lacking in minority populations where individuals often experience inequality and differential treatment due to ethnicity, sexual orientation, and gender identity. For researchers, clinicians, social program developers, and law and policy makers, learning more about contextual resilience, resilience in terms of protective factors and risk factors, and general characterological resilience is useful in the development of programs and the provision of equal and inclusive services to underserviced populations.

The essence and foundation of the scientific method is working from what we know toward what we do not know using ethically acceptable, systematic, evidence-based research methods. A part of this process is the design of quality research, which includes understanding how the instruments to be used in a study have used previously, and the types of reliability and validity evidence that are available (or unavailable) for that instrument in the population to be

studied. My contribution to this process is a set of summaries of evidence related to structure and use of instruments to measure resilience, ego-strength, hardiness, grit, and mental toughness spanning the time period from 1953 to 2019. I hope that my research will serve as an informative and useful starting point for the researcher who is interested in the tools available to investigate resilience in adults using a characterologically based instrument.

Further, differential test performance, when it is predictable (i.e., moderated by an observable or measurable variable), is neither inherently "good" nor "bad." Many instruments perform differentially across populations and this is not inherently problematic, so as long as clinicians and researchers are aware of prior evidence of reliability and validity or norming data that may provide a general idea of how the instrument is likely to perform with specific clients or research participants, and recognize that past evidence for reliability does not necessarily apply to subsequent studies. Thus, the assessment of how each instrument performs in each group of study participants is essential to the meaningful interpretation of the scores on the instrument.

To summarize the results of the present research, scores on the BRS generally yield "good" internal consistency values, with little variation due to sample characteristics or study characteristics. There is also consistent convergent validity evidence that BRS scores measure what scores on other characterological measures of resilience are measuring, and that the associations between BRS scores and scores on conceptually related measures are consistent with the expectation outlined in the resilience literature. As with all characterological measures of resilience, in order to determine if the hypothesis of characterological resilience is supported by the evidence studies that examine the temporal sequence of resilience development are needed. For example, it is necessary to know if adversity was, or is, present and preceded one or more resilient outcomes, or if resilience was, or is, present and was found to be predictive one or

more resilient outcomes.

These projects have brought to light that numerous measures of characterological resilience are available with varying amounts of psychometric information available for each measure. Some characterological resilience measures are unidimensional and focus on the depth of resilience (e.g., the BRS, measures the "bouncing back" aspect of resilience), and others are multi-dimensional, and allow researchers to collect information on wider range of criteria hypothesized to comprise characterological resilience. Thus, using at least one depth measure and at least one breath measure is recommended. In closing, as we move forward in the study of resilience, it is important to clearly define what type of resilience in being measured in each study, which definition is being used, and which theoretical base underlies it. If characterological resilience is the type of resilience being studied, measurement ideally needs to occur at more than one time point, and the relation of that timepoint to the occurrence of adversity needs to be addressed (i.e., either measured or described). Despite that hundreds of studies have been conducted, many aspects of resilience remain unknown. Although there are many possible paths to discovery using the scientific method (and many possible permutations and combinations of these paths), following the above noted recommendations related to measure selection and study design would be a reasonable and evidence-based next step to further explore resilience, related constructs, and resilient outcomes.

References (General Introduction and General Discussion)

Ahern, N. R., Kiehl, E. M., Sole, M. L., & Byers, J. (2006). A review of instruments measuring resilience. *Issues in Comprehensive Pediatric Nursing, 29*, 103–125.

Alessandri, G., Vecchione, M., Caprara, G., & Letzring, T. D. (2012). The Ego Resiliency scale revised: A cross-cultural study in Italy, Spain, and the United States. *European Journal of Psychological Assessment, 28*, 139-146.

Allport, G. W. (1921) Personality and character. *Psychological Bulletin, 18,* 441-455.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational, & Psychological Testing (US). (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. Arlington, VA: American Psychiatric Publishing.

Apple, Inc. (2016). FileMaker Pro 16. [Computer Software] Cupertino, California: Apple, Inc.

Barron, F. (1953). An ego-strength scale which predicts response to psychotherapy. *Journal of Consulting Psychology, 17*, 327-333.

Bartholomew, T. T., A. Robbins, K., Joy, E. E., Kang, E., & Maldonado Aguiñiga, S. (2020). Clients' resilience and distress in psychotherapy: a preliminary meta-analysis. *Counselling Psychology Quarterly*, 1–19. https://doi.org/10.1080/09515070.2020.1777389

Bartone, P.T. (1991). Development and Validation of a Short Hardiness Measure. Paper presented at the annual convention of the American Psychological Society, Washington, DC.

Bartone PT: A short hardiness scale. (1995). Paper presented at the annual convention of the

American Psychological Society, New York.

Bartone, P. T., Ursano, R. J., Wright, K. M., & Ingraham, L. H. (1989). The impact of a military air disaster on the health of assistance workers. *The Journal of Nervous and Mental Disease, 177*, 317-328.

Baruth, K. E., & Carroll, J. J. (2002). A formal assessment of resilience: The Baruth Protective Factors Inventory. *The Journal of Individual Psychology, 58*, 235–244.

Benishek, L. A., Feldman, J. M., Shipon, R. W., Mecham, S. D. (2005). Development and evaluation of the Revised Academic Hardiness Scale. *Journal of Career Assessment, 13*, 59-76. https://doi.org/10.1177/1069072704270274

Block, J., & Kremen, A. M. (1996). IQ and ego-resiliency: Conceptual and empirical connections and separateness. *Journal of Personality and Social Psychology, 70*, 349-361.

Borenstein, M., L.V. Hedges, J.P.T. Higgins, and H.R. Rothstein. (2009). *Introduction to meta-analysis*. Chicester/West Sussex, UK: John Wiley & Sons, Inc.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2014). Comprehensive Meta-Analysis version 3.3.070. [Computer Software] Englewood, NJ: Biostat

Calhoun, L. G., & Tedeschi, R. G. (2013). *Posttraumatic growth in clinical practice*. Routledge.

Camara, W. J., Nathan, J. S., & Puente, A. E. (2000). Psychological test usage: Implications in professional psychology. *Professional Psychology: Research & Practice, 2*, 141-154.

Campbell-Sills, L., & Stein, M. B. (2007). Psychometric analysis and refinement of the Connor-Davidson Resilience Scale (CD-RISC): Validation of a 10-item measure of resilience. *Journal of Traumatic Stress, 6*, 1019-1028.

Canavan, J. (2008). Resilience: Cautiously welcoming a contested concept. *Child Care in Practice, 14*, 1-7.

Cicchetti, D. (2010). Resilience under conditions of extreme stress: a multilevel perspective. *World Psychiatry, 9*, 145-154.

Clough, P. J., Marchant, D. C., & Earle, K. (2007). *Mental Toughness Questionnaire 48: Technical manual*. Wrexham, UK: AQR.

Connor, K. M., & Davidson, J. R. T. (2003). Development of a new resilience scale: The Connor-Davidson resilience scale (CD-RISC). *Depression and Anxiety, 18*, 76–82.

Costa, P T. & McCrae, R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five Factor Model (NEO-FFI) Professional manual*. Lutz, Florida: Psychological Assessment Resources.

Credé, M., Harms, P., Niehorster, S., & Gaye-Valentine, A. (2012). An evaluation of the consequences of using short measures of the Big Five personality traits. *Journal of Personality and Social Psychology, 102*, 874-888.

Credé, M., Tynan, M. C., & Harms, P. D. (2017). Much ado about grit: A meta-synthesis of the grit literature. *Journal of Personality and Social Psychology, 113*, 492-511.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281-302.

Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement, 64*, 391–418.

DerSimonian, R. & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7*, 177-188.

De Vries, H., Elliott, M., Kanouse, D., & Teleki, S. (2008). Using Pooled Kappa to summarize

interrater agreement across many items. *Field Methods*, *20*, 272–282. https://doi.org/10.1177/1525822X08317166

Duckworth, A. L., Peterson, C., Matthews, M., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Personality Processes and Individual Differences, 92*, 1087-1101.

Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (Grit-S). *Journal of Personality Assessment, 91,* 166-174.

Dunn, T. J., Baguley, T., Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105*, 399-412.

Egger., M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629-647.

Eschleman, K. J., Bowling, N. A., & Alarcon, G. M. (2010). A meta-analytic examination of hardiness. *International Journal of Stress Management, 17*, 277-307. https://dx.doi.org/10.1037/a0020476

Field, A. (2005). *Discovering statistics using SPSS*. (2nd ed.). London: Sage.

Fisher, R. A. (1925). *Statistical methods for research workers.* Edinburgh, UK: Oliver and Boyd.

Fletcher, D., & Sarkar, M. (2013). Psychological resilience: A review and critique of definitions, concepts, and theory. *European Psychologist, 18*, 12-23.

Freud, S. (1905). Three essays on the theory of sexuality. Retrieved from: online database: The complete works of Sigmund Freud (May 21, 2018)
URL: https://www.valas.fr/IMG/pdf/Freud_Complete_Works.pdf

Freud, S. (1927). *The ego and the id.* Honolulu, HI: Hogarth Press.

Friborg, O., Hjemdal, O., Rosenvinge, J. H., & Martinussen, M. (2003). A new rating scale for adult resilience: What are the central protective resources behind healthy adjustment? *International Journal of Methods in Psychiatric Research, 12*, 65–76.

Friborg, O., Barlaug, D., Martinussen, M., Rosenvinge, J. H., & Hjemdal, O. (2005). Resilience in relation to personality and intelligence. *International Journal of Methods in Psychiatric Research, 14*, 29–42.

Friborg, O., Martinussen, M., & Rosenvinge, J. H. (2006). Likert-based vs. semantic differential-based scorings of positive psychological constructs: A psychometric comparison of two versions of a scale measuring resilience. *Personality and Individual Differences, 40*, 873 – 884.

Funk, S. C. (1992). Hardiness: A review of theory and research. *Health Psychology, 11*, 335-345.

Galton, F. (1884). Measurement of character. *Fortnightly Review, 36, 179–185.*

Garmezy, N. (1974). Children at risk: The search for the antecedents of schizophrenia part I. Conceptual models and research methods. *Schizophrenia Bulletin, 8,* 14-90.

Garmezy, M. (1993). Children in poverty: Resilience despite risk, *Psychiatry, 56*, 127-136.

Gillespie, B. M., Chaboyer, W., & Wallis, M. (2007). Development of a theoretically derived model of resilience through concept analysis. *Contemporary Nurse, 25*, 124-135.

Goodwin, J. C. (2008). *Research in psychology: Methods and design.* Hoboken, NJ: John Wiley & Sons, Inc.

Gottesman, I. I. (1959). More construct validation of the ego-strength scale. *Journal of Consulting Psychology, 4*, 342-346.

Haynes, S. N., Smith, G. T., & Hunsley, J. (2011). *Scientific foundations of clinical assessment.* New York, NY: Routledge.

Hathaway, S. R., & McKinley J. C. (1942). *Manual for the Minnesota Multiphasic Personality Inventory.* Minneapolis, MN: University of Minnesota Press.

Higgins, J. P. T. & Green, S. (2008). *Cochrane handbook for systematic reviews of interventions.* West Sussex, UK: John Wiley & Sons Ltd.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., Altman, D. G. (2009). Measuring inconsistency in meta-analyses. *British Medical Journal, 237*, 557-560

Howell, D. C. (2014). *Fundamental Statistics for the Behavioral Sciences*. (8th ed.). Belmont, CA: Wadsworth Cengage Learning

Howell, R., & Shields, A. (2008). The file drawer problem in reliability generalization: A strategy to compute a Fail-Safe N with reliability coefficients. *Educational and Psychological Measurement*, *68*, 120–128. https://doi.org/10.1177/0013164407301528

HSBC. (2021). The Value of Education: The price of success. Retrieved from: https://www.us.hsbc.com/content/dam/hsbc/us/en_us/value-of-education/HSBC_VOE5_USA_FactSheet_508r2.pdf

Hu, T., Zhang, D., & Wang, J. (2015). A meta-analysis of the trait resilience and mental health. *Personality and Individual Differences, 76*, 18–27. https://doi.org/10.1016/j.paid.2014.11.039

Huedo-Medina, T., Sanchez-Meca, J., Marin-Martinez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? *Psychological Methods, 11,* 193-206.

Hunsley, J. & Mash E. J. (2018). *A guide to assessment that work*. (2nd ed.). New York, NY: Oxford University Press.

Hunsley, J., & Meyer, G. J. (2003). The incremental validity of psychological testing and

assessment: Conceptual, methodological, and statistical issues. *Psychological Assessment, 15*, 446-455.

Hunter, J. E. & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error bias in research findings*. (2nd ed.). Thousand Oaks, CA: Sage Publications.

Inter-Rater Reliability. (2018). In *StatisticsHowTo.com*. Retrieved April 16, 2018, from http://www.statisticshowto.com/inter-rater-reliability/

James, W. (1907). The energies of men. *Science, New Series, 25*, 321-332.

Jascó, P. (2004). Savvy searching: Citation searching. *Online Information Review, 28*, 454-460.

Jew, C. L., Green, K. E., & Kroger, J. (1999). Development and validation of a measure of resiliency. *Measurement and Evaluation in Counseling and Development, 32*, 75-89.

Kelly, D. R., Matthews, M. D., & Bartone, P. T. (2014). Grit and hardiness as predictors of performance among west point cadets. *Military Psychology, 26*, 327-342.

King, L. A., King, D. W., Vogt, D. S., Knight, J., & Samper, R. E. (2003). Deployment Risk and Resilience Inventory (DRRI): A collection of measures for studying deployment-related experiences in military personnel and veterans. *Military Psychology, 18*, 89-120.

Klohnen, E. C. (1996). Conceptual analysis and measurement of the construct of ego-resiliency. *Journal of Personality and Social Psychology, 70*, 1067-1079.

Kobasa, S. C., Maddi, S. R., & Kahn, S. (1982). Hardiness and health: A prospective study. *Journal of Personality and Social Psychology, 42*, 168-177.

Kobasa, S. C. O. (1979). *Stress, personality, and health: An overlooked possibility.* (Unpublished doctoral dissertation). University of Chicago, Chicago, Illinois, USA.

Luthar, S. S., Cicchetti, D., & Becker, B. (2000), The construct of resilience: A critical evaluation and guidelines for future work. *Child Development, 71*, 543-562.

Lyons, D. M., Parker, K. J., Katz, M., & Schatzberg, A. F. (2009). Developmental cascades linking stress inoculation, arousal regulation, and resilience. *Frontiers in Behavioral Neuroscience, 3*, 1-6.

Maddi, S. (2002). The story of hardiness: Twenty years of theorizing, research, and practice. *Consulting Psychology Journal: Practice and Research, 54*, 175-185.

Maddi, S. R., Erwin, L. M., Carmody, C. L., Villarreal, B. J., White, M., & Gundersen, K. K. (2013). Relationship of hardiness, grit, and emotional intelligence to internet addiction, excessive consumer spending, and gambling. *The Journal of Positive Psychology, 8*, 128-134.

Maddi, S. R., Harvey, R. H., Khoshaba, D. M., Lu, J. L., Persico, M., & Brow, M. (2006). The personality construct of hardiness, III: Relationships with repression, innovativeness, authoritarianism, and performance. *Journal of Personality, 74,* 575–597.

Maddi, S. R., & Khoshaba, D. M. (1994). Hardiness and mental health. *Journal of Personality Assessment, 63*, 265-274.

Massachusetts Institute of Technology (2018). 638 Primary Personality Traits. Retrieved from: http://ideonomy.mit.edu/essays/traits.html

Masten, A. S. (2001). Ordinary magic: Resilience processes in Development. *American Psychologist, 56*, 227-238.

Matthews, M. D. (2008). Toward a positive military psychology. *Military Psychology, 20*, 289-298.

Messick, S. (1990). *Validity of test interpretation and use*. Princeton, NJ: Educational Testing Service.

Mettle (2018). In *Merriam-Webster Online Dictionary.* Retrieved April 27, 2018, from

https://www.merriam-webster.com/dictionary/mettle

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-Analyses: The PRISMA Statement. *PLoS Med, 6*. https://doi.org/10.1371/journal.pmed1000097

Moher, D., Shamseer, L., Clarke, M, Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., & Stewart, L. A. (2015). Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement. *Systematic Review, 4*. 1-9. https://doi.org/: 10.1186/2046-4053-4-1

Nicholls, A. R., Polman, R. C. J., Levy, A. R., & Backhouse, S. H. (2008). Mental toughness, optimism, pessimism, and coping among athletes. *Personality and Individual Differences, 44*, 1182-1192.

Nowack, K. M. (1989). Coping style, cognitive hardiness, and health status. *Journal of Behavioral Medicine, 12*, 163-178.

Nunberg, H. (1942). Ego strength and ego weakness. *Psychoanalytic Journal for the Arts and Sciences, 3*, 25-40.

O'Neal, M. R. (1999). *Measuring resilience*. Retrieved from https://search-proquest-com.proxy.bib.uottawa.ca/docview/62411455?accountid=14701

Orwin, R. (1983). A Fail-Safe N for Effect Size in Meta-Analysis. *Journal of Educational Statistics*, *8*, 157–159. https://doi.org/10.2307/1164923

Oshio, A., Taku, K., Hirano, M., & Saeed, G. (2018). Resilience and big five personality traits: A meta-analysis. *Personality and Individual Differences, 127*, 54–60. https://doi.org/10.1016/j.paid.2018.01.048

Padesky, C. A, & Mooney, K. A. (2012). Strengths-based cognitive-behavioural therapy: A four-

step model to build resilience. *Clinical Psychology and Psychotherapy, 19*, 283-290.

Pangallo, A., Zibarras, L., Lewis, R., & Flaxman, P. (2015). Resilience through the lens of interactionism: A systematic review. *Psychological Assessment, 27,* 1-20.

Pearson, K. (1900). X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series 5*, 157-175.

Piedmont R.L. (2014) Reliability Coefficient. In: Michalos A.C. (eds) *Encyclopedia of Quality of Life and Well-Being Research*. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-0753-5_2463

Pigott, T. D. (2012). *Advances in meta-analysis.* New York, NY: Springer.

Pollock, S. E., & Duffy, M. E. (1990). The Health-Related Hardiness Scale: development and psychometric analysis. *Nursing Research, 39*, 218-222.

Quintana D. S. (2015). From pre-registration to publication: A non-technical primer for conducting a meta-analysis to synthesize correlational data. *Frontiers in Psychology, 6*, 1-9. https://doi.org/10.3389/fpsyg.2015.01549

Rafilson, F. (1991). The case for validity generalization. *Practical Assessment, Research & Evaluation, 2*, 1-3.

Reeves, T. D., & Marbach-Ad, G. (2016). Contemporary test validity in theory and practice: A primer for discipline-based education researchers. *CBE Life Sciences Education, 15,* 1-9.

Resilio. (2018). In *Latin-Dictionary.com*. Retrieved April 26, 2018, from http://www.latin-dictionary.net

Resilience. (2018). In *Online Etymology Dictionary*. Retrieved April 26, 2018, from https://www.etymonline.com/word/resilience

Rimfeld, K., Kovas, Y., Dale, P. S., & Plomin, R. (2016). True grit and genetics: Predicting academic achievement from personality. *Journal of Personality and Social Psychology, 111*, 780-789.

Rosenthal, R. (1979). The "File Drawer Problem" and tolerance for null results, *Psychological Bulletin, 3, 638-641*. http://dx.doi.org/10.1037/0033-2909.86.3.638

Rouse, S. V. (2007). Using reliability generalization methods to explore measurement error: An illustration using the MMPI-2 PSY-5 scales. *Journal of Personality Assessment*, *88*, 264-275.

Rutter, M. (2012). Resilience as a dynamic concept. *Development and Psychopathology, 24,* 335-344.

Rutter, P.A., Freedenthal, S., & Osman, A. (2008). Assessing protection from suicidal risk: Psychometric properties of the suicide resilience inventory. *Death Studies, 33*, 142-153. doi: 10.1080.07481180701801295

Salisu, I., & Hashim, N. (2017). A critical review of the scales used in resilience research. *IOSR Journal of Business and Management, 19*, 23-33.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology, 62*, 529-540.

Schmidt, F. L., Hunter, J. E., Pearlman, K., & Hirsh, H. R. (1985). Forty questions about validity generalization and meta-analysis. *Personnel Psychology*, *38*, 697-798.

Schmidt, F. L., Law, K., Hunter, J. E., Rothstein, H. R., Pearlman, K., & McDaniel, M. (1993). Refinements in validity generalization methods: Implications for the situational specificity hypothesis. *Journal of Applied Psychology, 78*, 3-12.

Schmidt, F. L., Hunter, J. E., & Urry, V. W. (1976). Statistical power in criterion-related

validation studies. *Journal of Applied Psychology, 61*, 473-485.

Seligman, M. E. (2019). Positive Psychology: A Personal History. *Annual Review of Clinical Psychology*, *15*, 1–23. https://doi.org/10.1146/annurev-clinpsy-050718-095653

Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist, 55*, 5-14.

Sheard, M., Golby, J., & van Wersch, A. (2009). Progress toward construct validation of the Sports Mental Toughness Questionnaire (SMTQ). *European Journal of Psychological Assessment, 25*, 186-193. https://doi.org/10.1027/1015-5759.25.3.186

Sinclair, V. G., & Wallston, K. A. (2004). The development and psychometric evaluation of the Brief Resilient Coping Scale. *Assessment, 11,* 94-101.

Smith, B. W., Dalen, J., Wiggins, K., Tooley, E., Christopher, P., & Bernard, J. (2008). The Brief Resilience Scale: Assessing the ability to bounce back. *International Journal of Behavioral Medicine, 15,* 194-200.

Smith, M. S. & Gray, S. W. (2009). The Courage to Challenge: A new measure of hardiness in LGBT adults. *Journal of Gay & Lesbian Social Services, 21*, 73-89. https://doi.org/10.1080/10538720802494776

Smith-Osborne, A., & Bolton, K. W. (2013). Assessing resilience: A review of measures across the life course. *Journal of Evidence-Based Social Work, 10*, 111-126.

St. Clair-Thompson, H., Giles, R., McGeown, S. P., Putwain, D., Clough, P., & Perry, P. (2017). Mental toughness and transitions to high school and to undergraduate study. *Educational Psychology, 7*, 792-809.

Stamp, E., Crust, L. Swann, C., Perry, J., Clough, P., & Marchant, D. (2014). Relationships between mental toughness and psychological wellbeing in undergraduate students.

Personality and Individual Differences, 75, 170-174.

Streiner, D. L. (2003) Starting at the beginning: an introduction to coefficient alpha and internal consistency, *Journal of Personality Assessment, 80*, 99-103.

Strycharczyk, D., & Clough, A. (2015). *Developing mental toughness: Coaching strategies to improve performance, resilience, and wellbeing.* London, UK: KoganPage.

Tedeschi, R. G., & Calhoun, L. G. (1996). The Posttraumatic Growth Inventory: Measuring the positive legacy of trauma. *Journal of Traumatic Stress, 9*, 455–471. https://doi.org/10.1002/jts.2490090305

The Connor-Davidson Resilience Scale Online User Guide. (2015, Nov. 10). Retrieved from: http://www.connordavidson-resiliencescale.com/user-guide.php

Thompson, S. G. & Higgins, J. P. T. (2002) How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine, 21*, 1559-1573.

Tupes, E. C., & Christal, R. E. (1961). *Recurrent personality factors based on trait ratings*. USAF ASD Technical Report No. 61-97. Lackland Air Force Base, Texas: U.S. Air Force.

Ulman, L. S. (2011). Systematic reviews and meta-analyses. *Journal of the Canadian Academy of Child and Adolescent Psychiatry, 20*, 57-59.

Urbina, S. (2004). *Essentials of psychological testing.* Hoboken, NJ: John Wiley & Sons, Inc.

Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement*, *58*, 6–20.

Vaishnavi, S., Connor, K., & Davidson, J. R. (2007). An abbreviated version of the Connor-Davidson Resilience Scale (CD-RISC), the CD-RISC2: Psychometric properties and applications in psychopharmacological trials. *Psychiatry research, 152*, 293-297.

van Breda, A. (2018). A Critical review of resilience theory and its relevance for social work. Social Work, 54, 1–18. https://doi.org/10.15270/54-1-611

Vassar, M., & Bradley, G. (2010). A reliability generalization study of coefficient alpha for the Life Orientation Test. *Journal of Personality Assessment, 9*, 362-370.

Vecchione, M., Allesandri, G., Barbaranelli, C., & Gerbino, M. (2010). Stability and change of ego-resilience from late adolescence to young adulthood: a multiperspective study using the ER-89R scale. *Journal of Personality Assessment, 92*, 212-221.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36, 1–48. https://www.jstatsoft.org/v036/i03.

Wagnild, G., & Young, H. M. (1991). Another look at hardiness. *Journal of Nursing Scholarship, 23*, 257-260.

Wagnild, G. M., & Young, H. M. (1993). Development and psychometric evaluation of the Resilience Scale. *Journal of Nursing Measurement, 1*, 165-178.

Weinberg, R., Freysinger, V., & Mellano, K. (2018). How can coaches build mental toughness? Views from sport psychologists. *Journal of Sport Psychology in Action, 9*, 1-10.

Wermelinger-Ávila, M. P., Lucchetti, A. L. G., & Lucchetti, G. (2017). Association between depression and resilience in older adults: A systematic review and meta-analysis. *International Journal of Geriatric Psychiatry*, *32*, 237–246. https://doi.org/10.1002/gps.4619

Windle, G. (2011). What is resilience? A review and concept analysis. *Reviews in Clinical Gerontology, 21*, 152-169.

Windle, G., Benett, K. M., & Noyes, J. (2011). A methodological review of resilience measurement scales. *Health and Quality of Life Outcomes, 9*, 1-18.

Windle, G., Markland, D. A., & Woods, R. T. (2008). Examination of a theoretical model of psychological resilience in older age. *Aging & Mental Health, 12*, 285-292.

Winwood, P. C., Colon, R., & McEwen, K. (2013). A practical measure of workplace resilience: Developing the Resilience at Work Scale. *Journal of Occupational and Environmental Medicine, 55*, 1205-1212.

Young, W. C., & Budynas, R. G. (2002). *Roark's formulas for stress and strain*. New York, NY: McGraw-Hill.

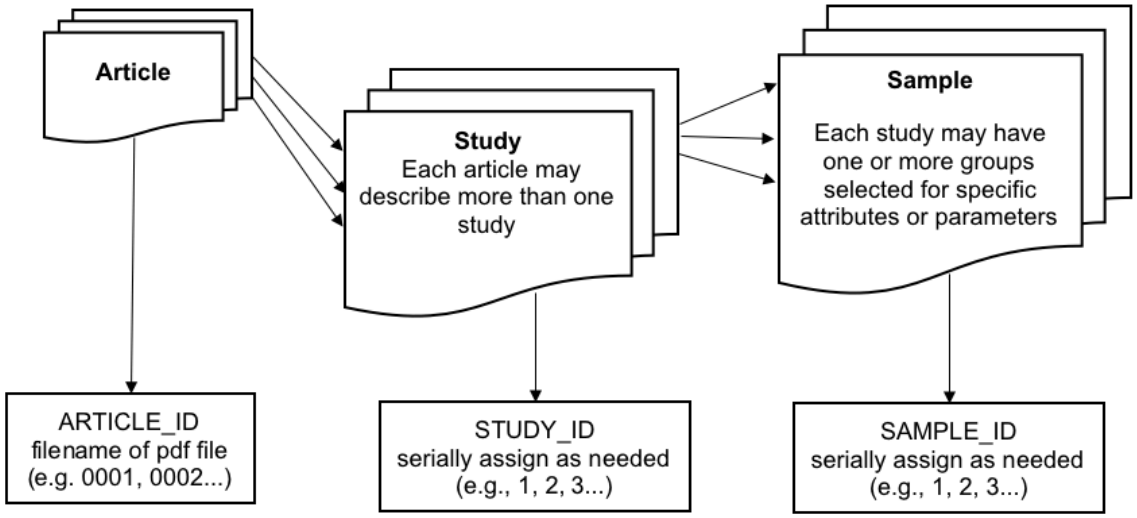
Youngstrom, E. A., Van Meter, A., Frazier, T. W., Hunsley, J., Prinstein, M. J., Ong, M., & Youngstrom, J. K. (2017). Evidence-based assessment as an integrative model for applying psychological science to guide the voyage of treatment. *Clinical Psychology (New York, N.Y.)*, *24*, 331–363. https://doi.org/10.1111/cpsp.12207

Younkin, S. L. (1992). *The development and validation of the Psychological Hardiness Scale.* (Unpublished doctoral dissertation). Ohio State University, Columbus, Ohio, USA.

Appendix A: Coding Manual Study 1

Coding Manual

For each journal article, use the supplied spreadsheets (titled: StudyData.xlsx, and SampleData.xlsx) to code Rater 2 variables as follows. The relation between identification numbers (`article_ID`, `study_ID`, and `sample_ID`) are many-to-one in that each article may contain one or more studies, and one or more samples, but each study and each sample may relate/cascade from only one article.



Article Spreadsheet (ArticleData.xlsx)

The article spreadsheet has been provided for your reference and contains the article identification number – which will match the .pdf file name for each article to be coded (e.g., 0001.pdf) title of the journal article, publication date, journal name). Please enter the coding status of "1" in the column "COMPLETE" for each article you have coded.

Study Spreadsheet (StudyData.xlsx)

The study spreadsheet will be used to code the following variables, using one row to represent each study. Fill in the `article_ID` variable as above, and serially assign unique values incrementing by 1 (primary key values) to each study as needed. Please refer to provided reference tables to code each variable. Variables will be coded using a categorical binary structure unless otherwise indicated (1 = applies to this article, 0 = does not apply to this article.)

- **Resilience Instruments Used.** There are 51 resilience instruments of interest for this study (listed below). One or more resilience instruments may have been used in each study. For each study (note that each article may contain more than one study) code a 1 in for each of the following instruments that were used. More than one instrument may have been used in each study please code all that apply.

— Abridged Hardiness Scale/Short Hardiness Scale (AHS/SHS-20)
— Barron's Ego Strength Scale, original for MMPI (Barron-68)
— Barron's Ego Strength Scale, revised for MMPI-2 (Barron-52)
— Brief Resilience Scale (BRS-6)
— Brief Resilient Coping Scale (BRCS-4)
— Cognitive Hardiness Scale (CHS-18)
— Cognitive Hardiness Scale (CHS-30)
— Connor-Davidson Resilience Scale (CDRISC-10)
— Connor-Davidson Resilience Scale (CDRISC-25)
— Dispositional Resilience Scale (DRS-15)
— Dispositional Resilience Scale (DRS-30)
— Dispositional Resilience Scale (DRS-45)
— Ego Resilience Scale for California Personality Inventory (CPI)(Klohnen-ER-29)
— Ego Strength Scale (Epstein-ESS-26)
— Ego Strength Scale (ESS-40)
— Grit Scale / Original Grit Scale (Grit-12)
— Hardiness Scale (HS-15)
— Hardiness Scale (HS-20)
— Hardiness Scale (HS-13)
— Hardiness Scale 50-item version/Revised Hardiness Scale (HS-50/RHS-50)
— HardiSurvery III-R (HardiSurvery III-R-65)
— HardiSurveyIII-R (HardiSurvey III-R)
— Hardy Personality Inventory (HPI-50)
— Lang and Goulet Hardiness Scale (LGHS-45)
— MeBTough Assessment (Me-B-Tough-43)
— Mental Toughness Questionnaire (MTQ-48)
— Mental Toughness Questionnaire (MTQ-18)
— Personal Views Survey (PVS) (PVS-50)
— Personal Views Survey II (PVS-II-50)
— Personal Views Survey III (PVS-III-30)
— Personal Views Survey III-R (PVS-III-R-18)
— Psychological Hardiness Scale (PHS-20)
— Psychological Hardiness Scale (PHS-40)
— Psychosocial Inventory of Ego Strength (PIES-64)
— Resilience Scale (RS-14)
— Resilience Scale (RS-15)
— Resilience Scale (RS-25)
— Resilience Scale (RS-26)
— Resilience Scale for Adults, 33-item Likert response English version (RSA-33)

— Resilience Scale for Adults, 37-item Likert response English version (RSA-37)
— Resilience Scale for Adults, 37-item semantic differential version, English) (RSA-37sd)
— Response to Stressful Experiences Scale (RSES-22)
— Revised Hardiness Scale (RHS-36)
— Short Grit Scale (Grit-8)
— Short Hardiness Scale (SHS-20-McNeill)
— Tachikawa Resilience Scale, English version) (TRS-10)
— The Ego Resilience Scale / ER89 (ER-89-14)
— The Ego Resilience Scale-Revised (ER-89-R-10)
— Thomas-Zander Ego Strength Scale (TZ-ESS-20)
— Thomas-Zander Ego Strength Scale (TZ-ESS-24)
— Thomas-Zander Ego Strength Scale (TZ-ESS-27)

- **Resilience Measure Count.** If more than one resilience instrument identified in the previous section was used in the study, using binary (1/0) to select the appropriate number of measures from the following variable list to indicate how many of the resilience measures on the aforementioned list were used in each study. Code a 1 for <u>only one</u> of the following variables.

  — Two Instruments
  — Three Instruments
  — Four Instrument
  — Five Instruments
  — FLAG – NUMBER OF INSTRUMENTS NOT ON LIST
  (if selected please describe using notes column)

- **Resilience Measure Chronology.** Resilience measures may have been administered to study participants at one point in time or many. Using the following structure, <u>code all timepoints with a 1 that are applicable to each study</u>.

  Timepoints: T1 is used to describe "Time 1" (the first time/baseline time data was collected from participants using the resilience measure(s). Respectively T2, T3, T4, and so on, to describe subsequent timepoints.

  EVENT: If an event occurred between measurement timepoints code (circumstance that occurred not due to the experimenter – e.g., graduation from military boot camp) code event with a 1.

  INTERVENTION: If an intervention occurred (e.g., treatment, training, or specific experience provided by researchers as a part of the research) code intervention with a 1. If resilience measures were administered to the participants during the intervention (e.g., at T1/baseline, and at the mid-point of an intervention (3 weeks into a 6-week yoga class), that timepoint during the intervention would be coded, in this instance as T2_INT

= 1, indicating that researchers made the second measurement T2, while the intervention was ongoing.

POST: If resilience measure data was collected directly following an event or intervention this time point is coded as POST (i.e. after). Therefore if the structure of the study was to measure resilience at baseline, 12 weeks of cognitive behavioural therapy was given to participants with no measurements taken until after all therapy sessions had finished, and resilience was measured directly following the intervention code 1's for T1, INTERVENTION, and T2_POST (because the second measure that was taken, T2, was taken post (directly following the intervention). Thus T1 = 1, INTERVENTION = 1, and T2_POST = 1. If however, a measure was not taken directly after, but a period of time passed, this is coded as FU (Follow-Up). See next section for follow-up coding.

Follow up (FU): In some studies, there is a delay in time following an event or intervention before resilience is measured, we will code these as FU, or follow-up. For example: If researchers measure resilience at baseline (T1 = 1), applied an intervention and no resilience measurement was taken during that intervention (e.g., a resilience training weekend seminar) (INTERVENTION = 1), resilience was measured directly following the seminar (T2_POST = 1), and resilience was measured again 6 weeks following the seminar (T3_FU = 1). The latter indicating the third measurement (T3) was a follow-up (FU) time point.

Times (T)	Events	Intervention and Intervention (T)	Post (T)	Follow-UP (T)
– T1	– EVENT	– INTERVENTION	– T1_POST	– T1_FU
– T2		(no measurement)	– T2_POST	– T2_FU
– T3		– T1_INT	– T3_POST	– T3_FU
– T4		– T2_INT	– T4_POST	– T4_FU
– T5		– T3_INT	– T5_POST	– T5_FU
		– T4_INT		– T6_FU
		– T5_INT		

— OTHER – resilience measurement cannot be described adequately using the above structure (Code OTHER = 1 where this structure does not apply, e.g., if resilience was measured once weekly for a year there are 52 time points and this structure would not apply, similarly, if participants were measured at follow-up annually for 10 years this structure would not apply. If researchers do not describe (or clearly describe) when resilience was measured, this structure cannot apply. In instance where this structure cannot be used to describe when resilience data were collected, code OTHER = 1 and make a note in the notes column. (**Important Note:** If OTHER = 1, leave all other columns in this section blank)

- **Country.** This is the country where the research took place. Use the guidelines below to select and code <u>only one</u> item from the list to follow.

Single country. If a study clearly indicates where (e.g., USA, Canada) then code the appropriate country with a 1. If research took place elsewhere, but in a location that is not a specific country (e.g., at sea) please use the code from either a) the stated place of residence of the participants (e.g., international cargo ship personnel from Australia, thus `Country_AUSTRALIA = 1`).

More than 1 country stated. For studies that contained samples from two or more countries, variables for collaborations between countries were generated (e.g. for a study that collected resilience data from the University of Western Ontario in Canada, and the University of Florida in the USA code `Country_CANADA_and_USA = 1`. For collaborations of more than 10 countries, code `Country_INTERNATIONAL_MULTICOUNTRY = 1`. If the regions given in the article are not on the given list code `Country_OTHER = 1` (the list was generated from the first 500 articles, so some new collaborations may be found) and use the notes column to indicate the new pair or list of countries. (**Reminder:** do not code more than one variable! also if `OTHER = 1`, leave all other columns in this section blank).

Country not stated. If the location where the research took place was not stated then the location of the authors affiliations (if all the same and it appears reasonable to assume the research took place also in that country then code the respective country with a 1. For example, if all researchers are affiliated with universities located in the USA and the topic of the study is the effects of the Deepwater Horizon Oil Spill and it is not explicitly stated that the research took place in the USA, from the information available it can be inferred that the research took place in the USA and coded as `Country_USA = 1`. Similarly, if all researchers are affiliated with the University of Toronto in Canada, and the participant sample is described as a sample of undergraduate psychology students it can be reasonably inferred that the research tool place in Canada, thus `Country_Canada = 1`.

If, however, authors are from more than one country e.g. three authors from Sweden, Denmark, and Norway, respectively, and if cannot be inferred from the other information available e.g. sample is described as "European young adults" code `Country_NotStated_AuthMixedAffil = 1.` Please code only one of the following choices with a 1.

Online. Presently many studies are online using services like Amazon's Mechanical Turk survey service to enroll study participants. If the region is stated (e.g., "we used MTurk to survey 450 New Zealanders") then code as the identified country. Here the code would be `Country_NEW_ZEALAND = 1`. If an online survey service is described but the region of the participants is not identified code as `Country_ONLINE_region not identified = 1`. If the study was conducted online using an online survey service and it is stated that "no limiters were used" or indicates participants were from multiple regions/countries code as `Country_ONLINE_International_NO_LIMITERS`. Note: for these codes to apply, participants must be recruited through the survey service.

— `Country_ANTARCTICA`

— Country_AUSTRALIA
— Country_BARBADOS
— Country_CANADA
— Country_GHANA
— Country_IRELAND (south, not in UK)
— Country_JAMAICA
— Country_LEBANON
— Country_MALAYSIA
— Country_NEW_ZEALAND
— Country_NIGERIA
— Country_SINGAPORE
— Country_SOUTH_AFRICA
— Country_UK (England, Scotland, Wales, Ireland – north)
— Country_USA

— Country_AUSTRALIA_and_CANADA
— Country_AUSTRALIA_and_NEW_ZEALAND
— Country_AUSTRALIA_and_UK
— Country_CANADA_and_USA
— Country_NIGERIA_and_SOUTH_AFRICA
— Country_SOUTH_AFRICA_and_UGANDA
— Coutnry_INTERNATIONAL_MULTICOUNTRY

— Country_ONLINE_RegionNotIdentified
— Country_ONLINE_NO_LIMITERS
— Country_ONLINE_USA_and_RegionNotIdentified
(used for 2 samples, 1 USA and 1 not identified)
— Country_NotStated_AuthMixedAffil
— Country_AT_SEA

If none of the above apply,
— Country_OTHER (and supply detail in the notes section, you may find some new countries that were not initially identified as variables)

- **Sample Count.** This variable is coded following the coding in the subsequent section to indicate how many distinct populations were recruited for participation in the study. Please return to this variable following the coding of the information for each sample in each study. Code a 1 for only one of the following variables:

— 1_POP
— 2_POP
— 3_POP
— 4_POP
— 5_POP

Sample Spreadsheet (SampleData.xlsx)

Definition: For the purposes of this systematic review, we are using the term sample to describe what criteria were used to recruit participants. You will use the data at the highest level (not subcategories derived after participants were recruited) to select which of these categories apply. For example, if a sample of university psychology students was selected, and a measure of resilience and a measure of test anxiety were administered, and following this students were separated into those with 'high test anxiety' scores and 'low test anxiety' scores, this is coded as 1 group: `UNI_Psychology_Students = 1`. We do not code this as 2 groups one with high scores of test anxiety and low scores of test anxiety because this information was not the basis for recruitment (i.e. the researchers did not know their anxiety scores until after they completed the study measures).

Header Variables and Sample Categories. For the following sections code as many as are applicable. Each grouping has a header, if a description under that header applies code both the header and the descriptor variable. E.g., if the sample of participants is university undergraduate psychology students code `Header_University = 1` and `Uni_PsychologyStudents = 1`.

— **`NSSA`** (Not Selected for Specific Attributes)
  Use this only when code no other categorical codes apply AND you do not see any criteria that might warrant the addition of a new categorical code.

— **`HEADER_University`**
— `UNI_Medical_or_Nursing_Students`
— `UNI_Psychology_Students`
— `UNI_Other_University_Students`
— `UNI_Faculty_or_University_Employees`
— `UNI_University_Related_Other`

— **`HEADER_Military`**
— `MILITARY_Active_Duty`
— `MILITARY_Veterans`
— `MILITARY_Cadets`
— `MILITARY_Military_Personnel_Other`

— **`HEADER_Selected_for_Specific_Work_or_Profession`**
— `WORK_First_Responders`
— `WORK_Healthcare_Professions`
  (Other than specified here, or mix of health professions, e.g. "doctors, nurses, and radiologists were surveyed")
— `WORK_Mental_Health_Professionals`
  (e.g. psychiatrists, psychologists, therapists, social workers that provide therapy)
— `WORK_Nurses`

— WORK_Physicians_Medical_Doctors
(code medial residents as medical students, code psychiatrists as mental health professionals as this is more specific)
— WORK_Nurses
— WORK_Teachers
(non-university, code university professors and other university faculty under university header, category faculty)
— WORK_Other
(provide brief note in note field to describe: e.g., miners, pilots, veterinarians, etc.)

— **HEADER_Caregivers**
— CARE_As_Parent_or_Pregnancy
— CARE_To_Family_Member_wPhysical_or_Mental_Illness
(if caregiver is providing care to their child who has been diagnosed with cancer, code here and also the parenting category. Do not code in the PH_MED_Cancer category as the study participant is not the cancer patient. Be certain to only code MED category conditions/diagnoses that the study participant is experiencing)
— CARE_Caregiver_Other

— **HEADER_PhysicalHealth_Medical**
— PH_MED_BrainInjury (ABI, TBI, Organic)
— PH_MED_Arthtitis
— PH_MED_Asthma
— PH_MED_Cancer
— PH_MED_Chronic_Pain_Fibromyalgia (and other pain conditions)
— PH_MED_COPD (Chronic Obstructive Pulmonary Disease)
— PH_MED_Diabetes
— PH_MED_HIV_andOr_Aids
— PH_MED_Heart_Condition
(Myocardial Infarction, Heart surgery pts, other heart conditions)
— PH_MED_MS (Multiple Sclerosis)
— PH_MED_Parkinsons_Disease
— PH_MED_Physical_Disability (including hearing or vision loss)
— PH_MED_SCI (Spinal Cord Injury)
— PH_MED_MedicalPts_Various_or_Other
(use this code if sample of study participants are receiving medical care but for varying conditions, if this variable is coded as 1 make use note column to describe (e.g., all participants were recruited for specific diagnosis that it is not listed here as a separate variable)
— PH_MED_CONTROL_GROUP_withoutMED_COND
(use this code only if a sample of participants without a medical condition are being compared to individuals with a medical condition, symptoms, or diagnosis. e.g. “there were two samples in this study, open heart surgery patients and a group of healthy controls”)

— **HEADER_Mental_Health**
— MH_Inpatients (individual receiving inpatient mental health services)
— MH_Outpatients (individual receiving inpatient mental health services)
— MH_Inpatient_Outpatient_status_not_stated
— MH_Sx_PTSD
— MH_Sx_Substance_Use
— MH_Sx_Anxierty
— MH_Sx_Depression_or_SuicideRel (e.g., suicide ideation)
— MH_Sx_Schizophrenia_or_OthPsychoticSymptoms
— MH_Sx_CONTROL_GROUP_without_SpecificMHSx
(use this code only if a sample of participants without a mental health symptoms/diagnoses were compared to individuals with a specified mental health difficulty, condition, symptoms, or diagnosis. e.g. “there were two samples in this study, psychiatric inpatients and a group of age matched individuals with no history of psychiatric symptoms”)
— MH_Sx_OTHER
(make note, but only if all participants selected for a specific cluster of symptoms or specific diagnoses prior to any administration of measures)

— **HEADER_StressorEvents**
(to clarify, these are events that individuals experienced that are generally assumed to cause some measure of stress or psychological distress in individuals who experience them, e.g.: loss of home due to hurricane, assault/abuse, bereavement, vehicle accident, combat. We chose this terminology because in the context of studying resilience individuals who experienced significant distress are sometimes studied in comparison to individuals who experience these events with no significant distress. Thus, while many of these events would be traumatic for many, not all individuals who experience these experience trauma. For example, if participants were recruited on the criteria that they are bereaved of a family member in the past 6 months, a participant who lost a great-grandmother aged 102 years and was able to express love and say goodbye would likely have a different experience of bereavement than parents who lost a child in a traffic accident, although this may not be the case. Thus we focus use of this category on events that are likely to have been selected by resilience researchers as having potential to cause distress but cautiously avoid use of the term trauma for this variable as this category focuses on participant recruitment based on a specific experience and not its impact to the participant. This noted, if participants are also receiving outpatient treatment for symptoms of post-traumatic stress following a combat experience, then the event would be coded here, and the treatment and conditions would be coded under the mental health category codes for such a situation would then be: HEADER_Mental_Health = 1, MH_Sx_PTSD = 1, MT_Outpatients = 1, HEADER_StressorEvents = 1, STRESS_Combat = 1)

— MH_STRESS_Bereavement
— MH_STRESS_Combat

— MH_STRESS_Holocaust_or_Genocide
— MH_STRESS_Earthquake
— MH_STRESS_Hurricane
— MH_STRESS_Industrial_Explosion
— MH_STRESS_Oil_Spill
— MH_STRESS_Motor_Vehicle_Accident
— MH_STRESS_Reported_Assault_or_Abuse
— MH_STRESS_other_event_or_more_than_1_Event
— MH_STRESS_CONTROL_GROUP_without_SpecificMHSx
(use this code only if a sample of participants who did not experience a specific stressor were compared to individuals with a stressor experience, “there were two samples in this study, individuals living in a coastal area impacted by Hurricane Katrina and individuals living in an adjacent coastal town that was not impacted by the hurricane”)

— **HEADER_SEX_GENDER_ORIENTATION**
Participants in this category were specifically selected for their biological sex, gender/gender identity, or sexual orientation. Also use this section to code samples that were limited specifically to any of these variables, even those not explicitly stated.

For example, if the study participants are were all recruited from a medium security men’s prison then the sample was limited to males, therefore code `SGSO_Male = 1`, as all participants would be biologically male or identifying as male. Similarly, if study participants were recruited because they were all in their first trimester of pregnancy, code `SGSO_Female = 1`. Note: variable names are shortened for the purposes of efficiency in statistical algorithms, but categories such as Male and Female include all individuals who identify as (or are identified by study authors as) a given sexual or gender identity.

— SGSO_MalesONLY
— SGSO_FemalesONLY
— SGSO_HeterosexualONLY
— SGSO_LGBTQIA_plusONLY
— SGSO_Cisgendered_IndividualsONLY
— SGSO_Transgender_IndividualsONLY
(There is overlap here, all samples coded in this category will also be coded in the SGSO_LGBTQIA_plusONLY =1 also. This additional specifier was added to observe how resilience research has been carried out in this specific population as individuals who have historically been excluded from much research.)

— **HEADER_ETHNICITY**
As these categories were developed based on the first 500 articles, it was noticed that while it is useful to include this variable as a specifier in that certain populations were selected for their ethnicity or cultural affiliation, that definitions

related to whom belongs in each category and terminology to describe each ethnicity vary substantially across authors, cultures, and regions. Thus, code `HEADER_ETHNICITY = 1` if participants were selected for their ethnicity and make a note of the term that the authors use to describe the population, but no other detail is needed. E.g., "we enrolled 200 individuals who self-identify as indigenous Canadians", "50 white males and 50 black males were recruited" (the latter here would be 2 samples, both selected for their identification as male and their ethnicity)

— **`HEADER_OTHER`**
Similar to the other variables on the sample spreadsheet additional categories were derived from observations made across the first 500 coded studies. There may be additional participant attributes to add to this list. If you find an attribute in a research article that was a unique basis for participant selection, please also code `OTHER_Other = 1` and make a note in the notes column (e.g., participants were all attendees of a specific conference or retreat and were selected on the basis that they attended)

— `OTHER_LimitedtoAge_65plus`
(or specifically identified as "senior citizens", "centenarians" or another identifier that would clearly indicate membership in this age group was the basis for selection).
Note: do not code solely on the basis of residence in a care facility/assisted living facility as these serve many age groups, and also do not code solely on the basis of the identifier "retiree" as individuals retire at many different ages)

— `OTHER_Athletics, Sport, or Exercise Participation`
Include all levels of sport: professional, collegiate, recreational, also include all types of sport: team sports, individual sports, and also include samples who were selected on the basis of exercise participation (e.g. "regular runners", "regular exercisers", individuals who train in gyms, etc.) Use the note field to specify the sport as we will use this to determine if additional subcategories are needed. Note: these are for physical sports based mostly on the participant's body, code sports such as: video game tournament players, motorcycle/NASCAR racers, and sports that relate to some form of technology as `OTHER_Other` (and specify)

— `OTHER_Limited_to_Couples_ONLY`

— `OTHER_Forensic_or_Prison_Settings`
Include all participants samples from correctional settings (e.g., prisons, jails, forensic mental health hospitals) – note for forensic mental health hospitals also code `MH_Inpatient = 1` in addition to coding this category as 1.

— `OTHER_Homelessness_or_ShelterResidents`

— `OTHER_Immigrants`

— `OTHER_Poverty_LowSES_impoverishedRegion`

(also code here if participants are described as recipients of ‘welfare’ or ‘social assistance’ or a program similarly described as providing some type of assistance to individuals in poverty – terminology will vary by country/region)

— `OTHER_Twins_or_Triplets`

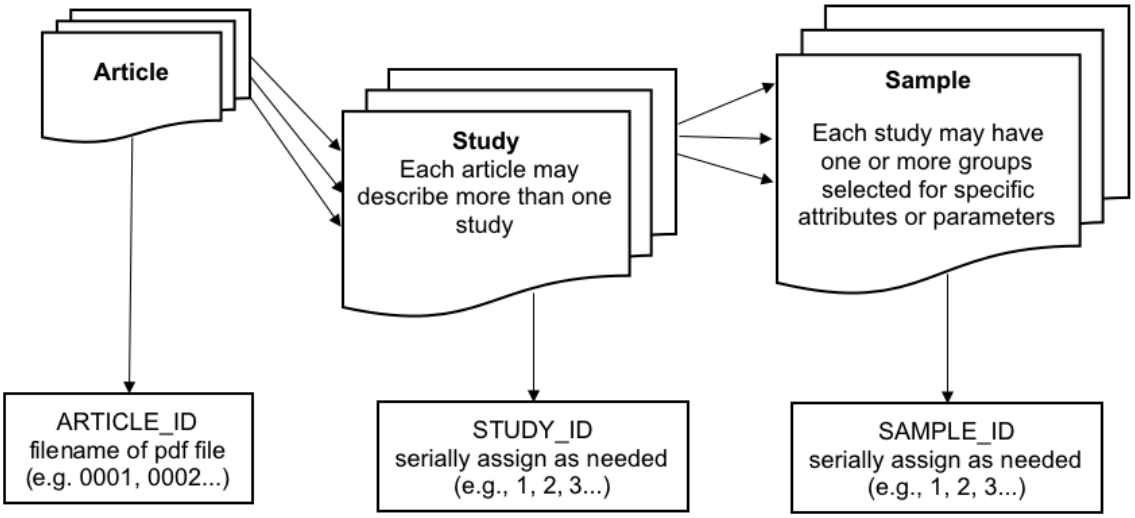
— `OTHER_Other`

(code any other specific criteria here that related to the entire sample of participants being selected based on this criteria, e.g., all individuals who attended specific retreat, specific conference, NASCAR racers, motorcycle racing club members). Note: if there is no clearly articulated characteristic upon which people were selected (e.g., convenience samples, sample from ‘an adjacent neighbourhood’, and so on code as NSSA = 1 (see first item in this section for detail on NSSA.

Appendix B: Coding Manual Study 2

Coding Manual

For each journal article, use the supplied spreadsheets (titled: StudyData.xlsx, and SampleData.xlsx) to code Rater 2 variables as follows. The relation between identification numbers (`article_ID`, `study_ID`, and `sample_ID`) are many-to-one in that each article may contain one or more studies, and one or more samples, but each study and each sample may relate/cascade from only one article.



Article Spreadsheet (ArticleData.xlsx)

The article spreadsheet has been provided for your reference and contains the article identification number – which will match the .pdf file name for each article to be coded (e.g., 0001.pdf) title of the journal article, publication date, journal name). Please enter the coding status of "1" in the column "COMPLETE" for each article you have coded.

Study Spreadsheet (StudyData.xlsx)

The study spreadsheet will be used to code the following variables, using one row to represent each study. Fill in the `article_ID` variable as above, and serially assign unique values incrementing by 1 (primary key values) to each study as needed. Please refer to provided reference tables to code each variable. Variables will be coded using a categorical binary structure unless otherwise indicated (1 = applies to this article, 0 = does not apply to this article.)

Meta-analytic information. Meta-analytic information is recorded per study (thus will go on the study level spreadsheet). Code a 1 in all of the following binary variables that apply:

INTERNAL CONSISTENCY

- HAS_IC_DATA (internal consistency data is present)
- IC_DATA_CronbachsAlpha
- IC_DATA_OtherStatistic (use if internal consistency reported but used other statistic and not Cronbach's alpha)
- T1_IC (IC value coded was at time 1)
- IC_T1_n (count for internal consistency statistic at T1)
- IC_T1_value (reported value for internal consistency at T1)
- IC_T2_n (count for internal consistency statistic at T2)
- IC_T2_value (reported value for internal consistency at T2)
- IC_T3_n (count for internal consistency statistic at T3)
- IC_T3_value (reported value for internal consistency at T3)

TEST-RETEST RELIABILITY

- T1_TR (TR value coded was at time 1)
- TR_T1_n (count for test-retest reliability statistic at T1)
- TR_T1_value (reported value for test-retest reliability at T1)
- TR_T2_n (count for test-retest reliability at T2)
- TR_T2_value (reported value for test-retest reliability at T2)
- TR_T3_n (count for itest-retest reliability at T3)
- TR_T3_value (reported value for test-retest reliability at T3)
- TR_STAT_ICC (select for interclass-correlation coefficient)
- TR_OTHER_STAT (select this for stat other than ICC)
- DETAIL: Specify name of statistic used if not ICC

SEX-GENDER INFORMATION FOR STUDY

- SEX_GEN_n (count total)
- SG_F (count of self-identified female participants)
- SG_M (count of self-identifiedmale participants)
- SG_TG (count of self-identified transgender participants)
- SG_O (count of other-identified participants)
- SG_MISSING_BLANK (count authors identified as missing SG data)
- SG_NOT_DESCRIBED (code "1" if no SG data given)
- SG_Ambiguous_Unclear (code "1" if SG data present but unclear, e.g. if data is reported in two places in the paper and does not agree, or information is not clearly described)

AGE DATA

- AGE_n (count total)
- AGE_MEAN

- AGE_SD (age standard deviation)
- AGE_Not_GIVEN (code "1" if age data not given)
- AGE_range_Only (code "1" if age data categorical, and no mean/sd given)
- AGE_Ambiguous_Unclear (code "1" if age data present but unclear, e.g. if data is reported in two places in the paper and does not agree, or information is not clearly described)

- **Resilience Measure Chronology.** Resilience measures may have been administered to study participants at one point in time or many. Using the following structure, code all timepoints with a 1 that are applicable to each study.

  Timepoints: T1 is used to describe "Time 1" (the first time/baseline time data was collected from participants using the resilience measure(s). Respectively `T2`, `T3`, `T4`, and so on, to describe subsequent timepoints.

  EVENT: If an event occurred between measurement timepoints code (circumstance that occurred not due to the experimenter – e.g., graduation from military boot camp) code event with a 1.

  INTERVENTION: If an intervention occurred (e.g., treatment, training, or specific experience provided by researchers as a part of the research) code intervention with a 1. If resilience measures were administered to the participants during the intervention (e.g., at T1/baseline, and at the mid-point of an intervention (3 weeks into a 6-week yoga class), that timepoint during the intervention would be coded, in this instance as `T2_INT = 1`, indicating that researchers made the second measurement `T2`, while the intervention was ongoing.

  POST: If resilience measure data was collected directly following an event or intervention this time point is coded as POST (i.e. after). Therefore if the structure of the study was to measure resilience at baseline, 12 weeks of cognitive behavioural therapy was given to participants with no measurements taken until after all therapy sessions had finished, and resilience was measured directly following the intervention code 1's for `T1`, `INTERVENTION`, and `T2_POST` (because the second measure that was taken, `T2`, was taken post (directly following the intervention). Thus `T1 = 1, INTERVENTION = 1,` and `T2_POST = 1.` If however, a measure was not taken directly after, but a period of time passed, this is coded as FU (Follow-Up). See next section for follow-up coding.

  Follow up (FU): In some studies, there is a delay in time following an event or intervention before resilience is measured, we will code these as FU, or follow-up. For example: If researchers measure resilience at baseline (`T1 = 1)`, applied an intervention and no resilience measurement was taken during that intervention (e.g., a resilience training weekend seminar) (`INTERVENTION = 1)`, resilience was measured directly following the seminar (`T2_POST = 1)`, and resilience was measured again 6 weeks following the seminar (`T3_FU = 1)`. The latter indicating the third measurement (`T3`) was a follow-up (`FU)` time point.

Times (T)	Events	Intervention and Intervention (T)	Post (T)	Follow-UP (T)
– T1	– EVENT	– INTERVENTION	– T1_POST	– T1_FU
– T2		(no measurement)	– T2_POST	– T2_FU
– T3		– T1_INT	– T3_POST	– T3_FU
– T4		– T2_INT	– T4_POST	– T4_FU
– T5		– T3_INT	– T5_POST	– T5_FU
		– T4_INT		– T6_FU
		– T5_INT		

— `OTHER` – resilience measurement cannot be described adequately using the above structure (Code `OTHER = 1` where this structure does not apply, e.g., if resilience was measured once weekly for a year there are 52 time points and this structure would not apply, similarly, if participants were measured at follow-up annually for 10 years this structure would not apply. If researchers do not describe (or clearly describe) when resilience was measured, this structure cannot apply. In instance where this structure cannot be used to describe when resilience data were collected, code `OTHER = 1` and make a note in the notes column. (**Important Note:** If `OTHER = 1`, leave all other columns in this section blank)

- **Country.** This is the country where the research took place. Use the guidelines below to select and code only one item from the list to follow.

  Single country. If a study clearly indicates where (e.g., USA, Canada) then code the appropriate country with a 1. If research took place elsewhere, but in a location that is not a specific country (e.g., at sea) please use the code from either a) the stated place of residence of the participants (e.g., international cargo ship personnel from Australia, thus `Country_AUSTRALIA = 1`) or b) make a note indicating details as clearly as possible and indicate "unclear – flag for rater consensus discussion".

  More than 1 country stated. For studies that contained samples from two or more countries, variables for collaborations between countries were generated (e.g. for a study that collected resilience data from the University of Western Ontario in Canada, and the University of Florida in the USA code `Country_CANADA_and_USA = 1`. For collaborations of more than 10 countries, code `Country_INTERNATIONAL_MULTICOUNTRY = 1`. If the regions given in the article are not on the given list code `Country_OTHER = 1` (the list was generated from the first 500 articles, so some new collaborations may be found) and use the notes column to indicate the new pair or list of countries. (**Reminder:** do not code more than one variable! also if `OTHER = 1`, leave all other columns in this section blank).

  Country not stated. If the location where the research took place was not stated then the location of the authors affiliations (if all the same and it appears reasonable to assume the research took place also in that country then code the respective country with a 1. For example, if all researchers are affiliated with universities located in the USA and the

topic of the study is the effects of the Deepwater Horizon Oil Spill and it is not explicitly stated that the research took place in the USA, from the information available it can be inferred that the research took place in the USA and coded as `Country_USA = 1`. Similarly, if all researchers are affiliated with the University of Toronto in Canada, and the participant sample is described as a sample of undergraduate psychology students it can be reasonably inferred that the research tool place in Canada, thus `Country_Canada = 1`.

If, however, authors are from more than one country e.g. three authors from Sweden, Denmark, and Norway, respectively, and if cannot be inferred from the other information available e.g. sample is described as "European young adults" code `Country_NotStated_AuthMixedAffil = 1.` Please code only one of the following choices with a 1.

Online. Presently many studies are online using services like Amazon's Mechanical Turk survey service to enroll study participants. If the region is stated (e.g., "we used MTurk to survey 450 New Zealanders") then code as the identified country. Here the code would be `Country_NEW_ZEALAND = 1`. If an online survey service is described but the region of the participants is not identified code as `Country_ONLINE_region not identified = 1`. If the study was conducted online using an online survey service and it is stated that "no limiters were used" or indicates participants were from multiple regions/countries code as `Country_ONLINE_International_NO_LIMITERS`. Note: for these codes to apply, participants must be recruited through the survey service.

— `Country_ANTARCTICA`
— `Country_AUSTRALIA`
— `Country_BARBADOS`
— `Country_CANADA`
— `Country_GHANA`
— `Country_IRELAND` (south, not in UK)
— `Country_JAMAICA`
— `Country_LEBANON`
— `Country_MALAYSIA`
— `Country_NEW_ZEALAND`
— `Country_NIGERIA`
— `Country_SINGAPORE`
— `Country_SOUTH_AFRICA`
— `Country_UK` (England, Scotland, Wales, Ireland – north)
— `Country_USA`

— `Country_AUSTRALIA_and_CANADA`
— `Country_AUSTRALIA_and_NEW_ZEALAND`
— `Country_AUSTRALIA_and_UK`
— `Country_CANADA_and_USA`
— `Country_NIGERIA_and_SOUTH_AFRICA`
— `Country_SOUTH_AFRICA_and_UGANDA`

— `Coutnry_INTERNATIONAL_MULTICOUNTRY`

— `Country_ONLINE_RegionNotIdentified`
— `Country_ONLINE_NO_LIMITERS`
— `Country_ONLINE_USA_and_RegionNotIdentified`
(used for 2 samples, 1 USA and 1 not identified)
— `Country_NotStated_AuthMixedAffil`
— `Country_AT_SEA`

If none of the above apply,
— `Country_OTHER` (and supply detail in the notes section, you may find some new countries that were not initially identified as variables)

- **Sample Count.** This variable is coded following the coding in the subsequent section to indicate how many distinct populations were recruited for participation in the study. Please return to this variable following the coding of the information for each sample in each study. Code a 1 for only one of the following variables:

— `1_POP`
— `2_POP`
— `3_POP`
— `4_POP`
— `5_POP`

Sample Spreadsheet (SampleData.xlsx)

Definition: For the purposes of this systematic review, we are using the term sample to describe what criteria were used to recruit participants. You will use the data at the highest level (not subcategories derived after participants were recruited) to select which of these categories apply. For example, if a sample of university psychology students was selected, and a measure of resilience and a measure of test anxiety were administered, and following this students were separated into those with 'high test anxiety' scores and 'low test anxiety' scores, this is coded as 1 group: `UNI_Psychology_Students = 1`. We do not code this as 2 groups one with high scores of test anxiety and low scores of test anxiety because this information was not the basis for recruitment (i.e. the researchers did not know their anxiety scores until after they completed the study measures).

Header Variables and Sample Categories. For the following sections code as many as are applicable. Each grouping has a header, if a description under that header applies code both the header and the descriptor variable. E.g., if the sample of participants is university undergraduate psychology students code `Header_University = 1` and `Uni_PsychologyStudents = 1`.

— **`NSSA`** (Not Selected for Specific Attributes)

Use this only when code no other categorical codes apply AND you do not see any criteria that might warrant the addition of a new categorical code.

— **HEADER_University**
  — UNI_Medical_or_Nursing_Students
  — UNI_Psychology_Students
  — UNI_Other_University_Students
  — UNI_Faculty_or_University_Employees
  — UNI_University_Related_Other

— **HEADER_Military**
  — MILITARY_Active_Duty
  — MILITARY_Veterans
  — MILITARY_Cadets
  — MILITARY_Military_Personnel_Other

— **HEADER_Selected_for_Specific_Work_or_Profession**
  — WORK_First_Responders
  — WORK_Healthcare_Professions
    (Other than specified here, or mix of health professions, e.g. "doctors, nurses, and radiologists were surveyed")
  — WORK_Mental_Health_Professionals
    (e.g. psychiatrists, psychologists, therapists, social workers that provide therapy)
  — WORK_Nurses
  — WORK_Physicians_Medical_Doctors
    (code medial residents as medical students, code psychiatrists as mental health professionals as this is more specific)
  — WORK_Nurses
  — WORK_Teachers
    (non-university, code university professors and other university faculty under university header, category faculty)
  — WORK_Other
    (provide brief note in note field to describe: e.g., miners, pilots, veterenarians, etc)

— **HEADER_Caregivers**
  — CARE_As_Parent_or_Pregnancy
  — CARE_To_Family_Member_wPhysical_or_Mental_Illness
  (if caregiver is providing care to their child who has been diagnosed with cancer, code here and also the parenting category. Do not code in the PH_MED_Cancer category as the study participant is not the cancer patient. Be certain to only code MED category conditions/diagnoses that the study participant is experiencing)
  — CARE_Caregiver_Other

- **HEADER_PhysicalHealth_Medical**
  - PH_MED_BrainInjury (ABI, TBI, Organic)
  - PH_MED_Arthtitis
  - PH_MED_Asthma
  - PH_MED_Cancer
  - PH_MED_Chronic_Pain_Fibromyalgia (and other pain conditions)
  - PH_MED_COPD (Chronic Obstructive Pulmonary Disease)
  - PH_MED_Diabetes
  - PH_MED_HIV_andOr_Aids
  - PH_MED_Heart_Condition
    (Myocardial Infarction, Heart surgery pts, other heart conditions)
  - PH_MED_MS (Multiple Sclerosis)
  - PH_MED_Parkinsons_Disease
  - PH_MED_Physical_Disability (including hearing or vision loss)
  - PH_MED_SCI (Spinal Cord Injury)
  - PH_MED_MedicalPts_Various_or_Other
    (use this code if sample of study participants are receiving medical care but for varying conditions, if this variable is coded as 1 make use note column to describe (e.g., all participants were recruited for specific diagnosis that it is not listed here as a separate variable)
  - PH_MED_CONTROL_GROUP_withoutMED_COND
    (use this code only if a sample of participants without a medical condition are being compared to individuals with a medical condition, symptoms, or diagnosis. e.g. "there were two samples in this study, open heart surgery patients and a group of healthy controls")
- **HEADER_Mental_Health**
  - MH_Inpatients (individual receiving inpatient mental health services)
  - MH_Outpatients (individual receiving inpatient mental health services)
  - MH_Inpatient_Outpatient_status_not_stated
  - MH_Sx_PTSD
  - MH_Sx_Substance_Use
  - MH_Sx_Anxiety
  - MH_Sx_Depression_or_SuicideRel (e.g., suicide ideation)
  - MH_Sx_Schizophrenia_or_OthPsychoticSymptoms
  - MH_Sx_CONTROL_GROUP_without_SpecificMHSx
    (use this code only if a sample of participants without a mental health symptoms/diagnoses were compared to individuals with a specified mental health difficulty, condition, symptoms, or diagnosis. e.g. "there were two samples in this study, psychiatric inpatients and a group of age matched individuals with no history of psychiatric symptoms")
- MH_Sx_OTHER
  (make note, but only if all participants selected for a specific cluster of symptoms or specific diagnoses prior to any administration of measures)

- **HEADER_StressorEvents**

(to clarify, these are events that individuals experienced that are generally assumed to cause some measure of stress or psychological distress in individuals who experience them, e.g.: loss of home due to hurricane, assault/abuse, bereavement, vehicle accident, combat. We chose this terminology because in the context of studying resilience individuals who experienced significant distress are sometimes studied in comparison to individuals who experience these events with no significant distress. Thus, while many of these events would be traumatic for many, not all individuals who experience these experience trauma. For example, if participants were recruited on the criteria that they are bereaved of a family member in the past 6 months, a participant who lost a great-grandmother aged 102 years and was able to express love and say goodbye would likely have a different experience of bereavement than parents who lost a child in a traffic accident, although this may not be the case. Thus we focus use of this category on events that are likely to have been selected by resilience researchers as having potential to cause distress but cautiously avoid use of the term trauma for this variable as this category focuses on participant recruitment based on a specific experience and not its impact to the participant. This noted, if participants are also receiving outpatient treatment for symptoms of post-traumatic stress following a combat experience, then the event would be coded here, and the treatment and conditions would be coded under the mental health category codes for such a situation would then be: `HEADER_Mental_Health = 1, MH_Sx_PTSD = 1, MT_Outpatients = 1, HEADER_StressorEvents = 1, STRESS_Combat = 1`)

- `MH_STRESS_Bereavement`
- `MH_STRESS_Combat`
- `MH_STRESS_Holocaust_or_Genocide`
- `MH_STRESS_Earthquake`
- `MH_STRESS_Hurricane`
- `MH_STRESS_Industrial_Explosion`
- `MH_STRESS_Oil_Spill`
- `MH_STRESS_Motor_Vehicle_Accident`
- `MH_STRESS_Reported_Assault_or_Abuse`
- `MH_STRESS_other_event_or_more_than_1_Event`
- `MH_STRESS_CONTROL_GROUP_without_SpecificMHSx`
  (use this code only if a sample of participants who did not experience a specific stressor were compared to individuals with a stressor experience, “there were two samples in this study, individuals living in a coastal area impacted by Hurricane Katrina and individuals living in an adjacent coastal town that was not impacted by the hurricane”)

- **`HEADER_SEX_GENDER_ORIENTATION`**
  Participants in this category were specifically selected for their biological sex, gender/gender identity, or sexual orientation. Also use this section to code samples that were limited specifically to any of these variables, even those not explicitly stated.

For example, if the study participants are were all recruited from a medium security men's prison then the sample was limited to males, therefore code `SGSO_Male = 1`, as all participants would be biologically male or identifying as male. Similarly, if study participants were recruited because they were all in their first trimester of pregnancy, code `SGSO_Female = 1`. Note: variable names are shortened for the purposes of efficiency in statistical algorithms, but categories such as Male and Female include all individuals who identify as (or are identified by study authors as) a given sexual or gender identity.

- `SGSO_MalesONLY`
- `SGSO_FemalesONLY`
- `SGSO_HeterosexualONLY`
- `SGSO_LGBTQIA_plusONLY`
- `SGSO_Cisgendered_IndividualsONLY`
- `SGSO_Transgender_IndividualsONLY`

  (There is overlap here, all samples coded in this category will also be coded in the SGSO_LGBTQIA_plusONLY =1 also. This additional specifier was added to observe how resilience research has been carried out in this specific population as individuals who have historically been excluded from much research.)

- **`HEADER_OTHER`**

  Similar to the other variables on the sample spreadsheet additional categories were derived from observations made across the first 500 coded studies. There may be additional participant attributes to add to this list. If you find an attribute in a research article that was a unique basis for participant selection, please also code `OTHER_Other = 1` and make a note in the notes column (e.g., participants were all attendees of a specific conference or retreat and were selected on the basis that they attended)

  - `OTHER_LimitedtoAge_65plus`

    (or specifically identified as "senior citizens", "centenarians" or another identifier that would clearly indicate membership in this age group was the basis for selection).

    Note: do not code solely on the basis of residence in a care facility/assisted living facility as these serve many age groups, and also do not code solely on the basis of the identifier "retiree" as individuals retire at many different ages)
  - `OTHER_Athletics, Sport, or Exercise Participation`

    Include all levels of sport: professional, collegiate, recreational, also include all types of sport: team sports, individual sports, and also include samples who were selected on the basis of exercise participation (e.g. "regular runners", "regular exercisers", individuals who train in gyms, etc.) Use the note field to specify the sport as we will use this to determine if additional subcategories are needed. Note: these are for physical sports based mostly on the participant's body, code sports such as: video game

tournament players, motorcycle/NASCAR racers, and sports that relate to some form of technology as `OTHER_Other` (and specify)

- `OTHER_Limited_to_Couples_ONLY`
- `OTHER_Forensic_or_Prison_Settings`
  Include all participants samples from correctional settings (e.g., prisons, jails, forensic mental health hospitals) – note for forensic mental health hospitals also code `MH_Inpatient = 1` in addition to coding this category as 1.
- `OTHER_Homelessness_or_ShelterResidents`
- `OTHER_Immigrants`
- `OTHER_Poverty_LowSES_impoverishedRegion`
  (also code here if participants are described as recipients of 'welfare' or 'social assistance' or a program similarly described as providing some type of assistance to individuals in poverty – terminology will vary by country/region)
- `OTHER_Twins_or_Triplets`
- `OTHER_Other`

(code any other specific criteria here that related to the entire sample of participants being selected based on this criteria, e.g., all individuals who attended specific retreat, specific conference, NASCAR racers, motorcycle racing club members). Note: if there is no clearly articulated characteristic upon which people were selected (e.g., convenience samples, sample from 'an adjacent neighbourhood', and so on code as NSSA = 1 (see first item in this section for detail on NSSA.

Correlation Spreadsheet (CorData.xlsx)
The correlation spreadsheet will Fill in the `article_ID, study_ID, and sample_ID` variables as above, and serially assign unique values incrementing by 1 (primary key values) to each study as needed.

CORRELATIONS

- `Pearson_r (code the r value)`
- `Pearson_r_n (count specific to this correlation)`
- `COR_Measure (indicate measure that is correlated with BRS, see reference list for appreviations for measures)`
- `COR_Var (indicate name of variable correlated with BRS, e.g., variables are values reported as correlated that do not use specific measures e.g., height, weight, years of education)`

Appendix C: References for Studies Included in Meta-Analyses in Study 2

Internal Consistency

Aery, A. (2019). *Cognitive Emotion Regulation and Resilience*. ProQuest Dissertations Publishing.

Arcury, T., Smyth, J. M., Johnson, J. A., Auer, B. J., Lehman, E., Talamo, G., & Sciamanna, C. N. (2018). Online positive affect journaling in the improvement of mental distress and well-being in general medical patients with elevated anxiety symptoms: A preliminary randomized controlled trial. *JMIR Mental Health*, *5*, e11290–e11290. https://doi.org/10.2196/11290

Autry, A. L. (2016). *The role of loneliness and resilience in psychosocial health for victims of cyberbullying in a college population*. ProQuest Dissertations Publishing.

Bajjani-Gebara, J., Hinds, P., Insel, K., Reed, P., Moore, K., & Badger, T. (2019). Well-being, self-transcendence, and resilience of parental caregivers of children in active cancer treatment: Where do we go from here? *Cancer Nursing*, *42*, 41–52. https://doi.org/10.1097/NCC.0000000000000662

Baker, D. A., Caswell, H. L., & Eccles, F. J. (2019). Self-compassion and depression, anxiety, and resilience in adults with epilepsy. *Epilepsy & Behavior*, *90*, 154–161. https://doi.org/10.1016/j.yebeh.2018.11.025

Bariola, E., Lyons, A., Leonard, W., Pitts, M., Badcock, P., & Couch, M. (2015). Demographic and psychosocial factors associated with psychological distress and resilience among transgender individuals. *American Journal of Public Health (1971)*, *105*, 2108–2116. https://doi.org/10.2105/AJPH.2015.302763

Batcho, K. I., & Shikh, S. (2016). Anticipatory nostalgia: Missing the present before it's

gone. *Personality and Individual Differences*, *98*, 75–84. https://doi.org/10.1016/j.paid.2016.03.088

Black, J. K., Balanos, G. M., & Whittaker, A. C. (2017). Resilience, work engagement and stress reactivity in a middle-aged manual worker population. *International Journal of Psychophysiology*, *116*, 9–15. https://doi.org/10.1016/j.ijpsycho.2017.02.013

Bonus, J. A., Peebles, A., Mares, M.-L., & Sarmiento, I. G. (2018). Look on the bright side (of media effects): Pokémon Go as a catalyst for positive life experiences. *Media Psychology*, *21*, 263–287. https://doi.org/10.1080/15213269.2017.1305280

Bore, M., Pittolo, C., Kirby, D., Dluzewska, T., & Marlin, S. (2016). Predictors of psychological distress and well-being in a sample of Australian undergraduate students. *Higher Education Research and Development*, *35*, 869–880. https://doi.org/10.1080/07294360.2016.1138452

Brown, G., Mikołajczak, G., Lyons, A., Power, J., Drummond, F., Cogle, A., Allan, B., Cooper, C., & O'Connor, S. (2018). Development and validation of PozQoL: a scale to assess quality of life of PLHIV. *BMC Public Health*, *18*, 527–527. https://doi.org/10.1186/s12889-018-5433-6

Burrow, A. L., & Hill, P. L. (2013). Derailed by diversity? Purpose buffers the relationship between ethnic composition on trains and passenger negative mood. *Personality & Social Psychology Bulletin*, *39*, 1610–1619. https://doi.org/10.1177/0146167213499377

Bursch, B., Lloyd, J., Mogil, C., Wijesekera, K., Miotto, K., Wu, M., Wilkinson, R., Klomhaus, A., Iverson, A., & Lester, P. (2017). Adaptation and evaluation of military resilience skills training for pediatric residents. *Journal of Medical Education and Curricular Development*, *4*, 1-7. https://doi.org/10.1177/2382120517741298

Bursch, B., Emerson, N. D., Arevian, A. C., Aralis, H., Galuska, L., Bushman, J., Sinclair, M., Grimley, K., Lester, P., & Bulut, Y. (2018). Feasibility of online mental wellness self-assessment and feedback for pediatric and neonatal critical care nurses. *Journal of Pediatric Nursing*, *43*, 62–68. https://doi.org/10.1016/j.pedn.2018.09.001

Campo, R. A., Bluth, K., Santacroce, S. J., Knapik, S., Tan, J., Gold, S., Philips, K., Gaylord, S., & Asher, G. N. (2017). A mindful self-compassion videoconference intervention for nationally recruited posttreatment young adult cancer survivors: feasibility, acceptability, and psychosocial outcomes. *Supportive Care in Cancer*, *25*, 1759–1768. https://doi.org/10.1007/s00520-017-3586-y

Can, N. (2018). *Caring for Beginning Counselors: The Relationship between Empathy, Supervisory Working Alliance, Resilience, Wellness, and Compassion Fatigue among Counselors-in-Training*. ProQuest Dissertations Publishing.

Carr, K., & Kellas, J. K. (2018). The role of family and marital communication in developing resilience to family-of-origin adversity. *Journal of Family Communication*, *18*, 68–84. https://doi.org/10.1080/15267431.2017.1369415

Cassidy, T. (2013). Benefit finding through caring: The cancer caregiver experience. *Psychology & Health*, *28*, 250–266. https://doi.org/10.1080/08870446.2012.717623

Christopher, M. S., Goerling, R. J., Rogers, B. S., Hunsinger, M., Baron, G., Bergman, A. L., & Zava, D. T. (2016). A pilot study evaluating the effectiveness of a mindfulness-based intervention on cortisol awakening response and health outcomes among law enforcement officers. *Journal of Police and Criminal Psychology*, *31*, 15–28. https://doi.org/10.1007/s11896-015-9161-x

Chrystal, M., Karl, J. A., & Fischer, R. (2019). The complexities of “minding the gap”:

Perceived discrepancies between values and behavior affect well-being. *Frontiers in Psychology*, *10*, 736–736. https://doi.org/10.3389/fpsyg.2019.00736

Crane, M., Phillips, J., & Karin, E. (2015). Trait perfectionism strengthens the negative effects of moral stressors occurring in veterinary practice. *Australian Veterinary Journal*, *93*(10), 354–360. https://doi.org/10.1111/avj.12366

Crane, M. F., & Searle, B. J. (2016). Building resilience through exposure to stressors: The effects of challenges versus hindrances. *Journal of Occupational Health Psychology*, *21*, 468–479. https://doi.org/10.1037/a0040064

Davis, M. A., & Cernas Ortiz, D. A. (2017). Revisiting the structural and nomological validity of the Zimbardo time perspective inventory. *Personality and Individual Differences*, *104*, 98–103. https://doi.org/10.1016/j.paid.2016.07.037

Drake, D. A., & Steege, L. M. B. (2016). Interpretation of hospital nurse fatigue using latent profile analysis. *Advances in Nursing Science*, *39*, 1–16. https://doi.org/10.1097/ANS.0000000000000130

Flett, J. A. M., Hayne, H., Riordan, B. C., Thompson, L. M., & Conner, T. S. (2019). Mobile mindfulness meditation: A randomised controlled trial of the effect of two popular apps on mental health. *Mindfulness*, *10*, 863–876. https://doi.org/10.1007/s12671-018-1050-9

Flett, J. A. M., Lie, C., Riordan, B. C., Thompson, L. M., Conner, T. S., & Hayne, H. (2017). Sharpen your pencils: Preliminary evidence that adult coloring reduces depressive symptoms and anxiety. *Creativity Research Journal*, *29*, 409–416. https://doi.org/10.1080/10400419.2017.1376505

Fullen, M. C., Richardson, V. E., & Granello, D. H. (2018). Comparing successful aging, resilience, and holistic wellness as predictors of the good life. *Educational*

Gerontology, *44*, 459–468. https://doi.org/10.1080/03601277.2018.1501230

Fuller, K. A., & Riggs, D. W. (2018). Family support and discrimination and their relationship to psychological distress and resilience amongst transgender people. *The International Journal of Transgenderism*, *19*, 379–388. https://doi.org/10.1080/15532739.2018.1500966

Gage, G. S. (2017). Social support and positive health practices in black late adolescents: The role of mediating variables. *Clinical Nursing Research*, *26*, 93–113. https://doi.org/10.1177/1054773815594579

Gloria, C. T., & Steinhardt, M. A. (2016). Relationships among positive emotions, coping, resilience and mental health. *Stress and Health*, *32*, 145–156. https://doi.org/10.1002/smi.2589

Grossman, M. R. (2017). *The Structure of Resilience: An Empirical Examination of Resilience Factors*. ProQuest Dissertations Publishing.

Haridas, S., Bhullar, N., & Dunstan, D. A. (2017). What's in character strengths? Profiling strengths of the heart and mind in a community sample. *Personality and Individual Differences*, *113*, 32–37. https://doi.org/10.1016/j.paid.2017.03.006

Harville, E. W., Xiong, X., Smith, B. W., Pridjian, G., Elkind-Hirsch, K., & Buekens, P. (2011). Combined effects of Hurricane Katrina and Hurricane Gustav on the mental health of mothers of small children. *Journal of Psychiatric and Mental Health Nursing*, *18*, 288–296. https://doi.org/10.1111/j.1365-2850.2010.01658.x

Heckenberg, R. A., Hale, M. W., Kent, S., & Wright, B. J. (2019). An online mindfulness-based program is effective in improving affect, over-commitment, optimism and mucosal immunity. *Physiology & Behavior*, *199*, 20–27. https://doi.org/10.1016/j.physbeh.2018.11.001

Heckman, B. W., Dahne, J., Germeroth, L. J., Mathew, A. R., Santa Ana, E. J., Saladin, M. E., & Carpenter, M. J. (2018). Does cessation fatigue predict smoking-cessation milestones? A longitudinal study of current and former smokers. *Journal of Consulting and Clinical Psychology*, *86*, 903–914. https://doi.org/10.1037/ccp0000338

Herbert, S. (2018). *Yoga Practice, Well-Being, and Resiliency*. ProQuest Dissertations Publishing.

Hourani, L. L., Williams, J., Lattimore, P. K., Morgan, J. K., Hopkinson, S. G., Jenkins, L., & Cartwright, J. (2018). Workplace victimization risk and protective factors for suicidal behavior among active duty military personnel. *Journal of Affective Disorders*, *236*, 45–51. https://doi.org/10.1016/j.jad.2018.04.095

Huberty, J., Green, J., Cacciatore, J., Buman, M. P., & Leiferman, J. (2018). Relationship between mindfulness and posttraumatic stress in women who experienced stillbirth. *Journal of Obstetric, Gynecologic, and Neonatal Nursing*, *47*, 760–770. https://doi.org/10.1016/j.jogn.2018.09.002

Ireland, M. J. (2013). Meditation and psychological health: Modeling theoretically derived predictors, processes, and outcomes. *Mindfulness*, *4*, 362–374. https://doi.org/10.1007/s12671-012-0136-z

Jackson, B., Compton, J., Whiddett, R., Anthony, D. R., & Dimmock, J. A. (2015). Preempting performance challenges: the effects of inoculation messaging on attacks to task self-efficacy. *PloS One*, *10*, e0124886–e0124886. https://doi.org/10.1371/journal.pone.0124886

Kaplan, J. B., Bergman, A. L., Christopher, M., Bowen, S., & Hunsinger, M. (2017). Role of resilience in mindfulness training for first responders. *Mindfulness*, *8*, 1373–1380. https://doi.org/10.1007/s12671-017-0713-2

Kortlever, J. T., Keulen, M. H., Teunis, T., Ring, D., Driscoll, M. D., Reichel, L. M., & Vagner, G. A. (2019). Does resiliency mediate the association of psychological adaptability with disability and pain in patients with an upper extremity injury or illness? *Journal of Psychosomatic Research*, *117*, 1–9. https://doi.org/10.1016/j.jpsychores.2018.12.002

Lenz, A. S., & Oliver, M. (2018). Predictive associations between family system characteristics and emergent protective factors. *The Family Journal*, *26*, 40–47. https://doi.org/10.1177/1066480717731343

Logie, C. H., Wang, Y., Lacombe-Duncan, A., Jones, N., Ahmed, U., Levermore, K., Neil, A., Ellis, T., Bryan, N., Marshall, A., & Newman, P. A. (2017). Factors associated with sex work involvement among transgender women in Jamaica: a cross-sectional study. *Journal of the International AIDS Society*, *20*, 1-14. https://doi.org/10.7448/IAS.20.01/21422

Lyons, A., Fletcher, G., & Bariola, E. (2016). Assessing the well-being benefits of belonging to resilient groups and communities: Development and testing of the Fletcher-Lyons Collective Resilience Scale (FLCRS). *Group Dynamics*, *20*, 65–77. https://doi.org/10.1037/gdn0000041

McArthur, M., Mansfield, C., Matthew, S., Zaki, S., Brand, C., Andrews, J., & Hazel, S. (2017). Resilience in veterinary students and the predictive role of mindfulness and self-compassion. *Journal of veterinary medical education*, *44*, 106–115. https://doi.org/10.3138/jvme.0116-027R1

McKay, S., Skues, J. L., & Williams, B. J. (2018). With risk may come reward: Sensation seeking supports resilience through effective coping. *Personality and Individual Differences*, *121*, 100–105. https://doi.org/10.1016/j.paid.2017.09.030

McNair, R. P., & Bush, R. (2016). Mental health help seeking patterns and associations among

Australian same sex attracted women, trans and gender diverse people: a survey-based study. *BMC Psychiatry*, *16*, 209–209. https://doi.org/10.1186/s12888-016-0916-4

Meintjes, A., & Hofmeyr, K. (2018). The impact of resilience and perceived organisational support on employee engagement in a competitive sales environment. *SA Journal of Human Resource Management*, *16*, 1–11. https://doi.org/10.4102/sajhrm.v16i0.953

Meyer, E. C., Kotte, A., Kimbrel, N. A., DeBeer, B. B., Elliott, T. R., Gulliver, S. B., & Morissette, S. B. (2019). Predictors of lower-than-expected posttraumatic symptom severity in war veterans: The influence of personality, self-reported trait resilience, and psychological flexibility. *Behaviour Research and Therapy*, *113*, 1–8. https://doi.org/10.1016/j.brat.2018.12.005

Milia, L., Waage, S., Pallesen, S., & Bjorvatn, B. (2013). Shift work disorder in a random population sample--prevalence and comorbidities. *PloS One*, *8*,1-7. e55306–e55306. https://doi.org/10.1371/journal.pone.0055306

Mongan, D., Shannon, C., Hanna, D., Boyd, A., & Mulholland, C. (2019). The association between specific types of childhood adversity and attenuated psychotic symptoms in a community sample. *Early Intervention in Psychiatry*, *13*, 281–289. https://doi.org/10.1111/eip.12478

Nicely, S. A. (2019). *Variations in Individualistic and Collectivistic Cultural Orientation and the Protective Factors that Contribute to Resilience: Comparisons from Jamaica, Rwanda, and the United States*. ProQuest Dissertations Publishing.

Nightingale, S., Spiby, H., Sheen, K., & Slade, P. (2018). Posttraumatic stress symptomatology following exposure to perceived traumatic perinatal events within the midwifery profession: The impact of trait emotional intelligence. *Journal of Advanced Nursing*, *74*,

2115–2125. https://doi.org/10.1111/jan.13719

Orom, H., Nelson, C. J., Underwood III, W., Homish, D. L., & Kapoor, D. A. (2015). Factors associated with emotional distress in newly diagnosed prostate cancer patients. *Psycho-Oncology (Chichester, England)*, *24*, 1416–1422. https://doi.org/10.1002/pon.3751

Pangallo, A., Zibarras, L., & Patterson, F. (2016). Measuring resilience in palliative care workers using the situational judgement test methodology. *Medical Education*, *50*, 1131–1142. https://doi.org/10.1111/medu.13072

Sánchez, J., Muller, V., Chan, F., Brooks, J. M., Iwanaga, K., Tu, W. M., Umucu, E., & Crespo-Jones, M. (2019). Personal and environmental contextual factors as mediators between functional disability and quality of life in adults with serious mental illness: a cross-sectional analysis. *Quality of Life Research*, *28*, 441–450. https://doi.org/10.1007/s11136-018-2006-1

Schroeder, D. A., Stephens, E., Colgan, D., Hunsinger, M., Rubin, D., & Christopher, M. S. (2018). A brief mindfulness-based intervention for primary care physicians: A pilot randomized controlled trial. *American Journal of Lifestyle Medicine*, *12*, 83–91. https://doi.org/10.1177/1559827616629121

Shakespeare-Finch, J., & Daley, E. (2017). Workplace belongingness, distress, and resilience in emergency service workers. *Psychological Trauma*, *9*, 32–35. https://doi.org/10.1037/tra0000108

Shoss, M. K., Jiang, L., & Probst, T. M. (2018). Bending without breaking: A two-study examination of employee resilience in the face of job insecurity. *Journal of Occupational Health Psychology*, *23*(1), 112–126. https://doi.org/10.1037/ocp0000060

Smith, L. B., Lenz, A. S., & Strohmer, D. (2017). Differential prediction of resilience among

individuals with and Wwithout a history of abuse. *Counseling and Values*, *62*, 106–122. https://doi.org/10.1002/cvj.12052

Smith, B., Tooley, E., Christopher, P., & Kay, V. (2010). Resilience as the ability to bounce back from stress: A neglected personal resource? *The Journal of Positive Psychology*, *5*, 166–176. https://doi.org/10.1080/17439760.2010.482186

Smith, B. W., Dalen, J., Wiggins, K., Tooley, E., Christopher, P., & Bernard, J. (2008). The Brief Resilience Scale: Assessing the ability to bounce back. *International Journal of Behavioral Medicine, 15,* 194-200. https://doi.org/10.1080/10705500802222972

Smith, B. W., Kay, V. S., Hoyt, T. V., & Bernard, M. L. (2009). Predicting the anticipated emotional and behavioral responses to an avian flu outbreak. *American Journal of Infection Control*, *37*, 371–380. https://doi.org/10.1016/j.ajic.2008.08.007

Smith, B. W., Tooley, E. M., Montague, E. Q., Robinson, A. E., Cosper, C. J., & Mullins, P. G. (2009). The role of resilience and purpose in life in habituation to heat and cold pain. *The Journal of Pain*, *10*, 493–500. https://doi.org/10.1016/j.jpain.2008.11.007

Smith, B. W., Guzman, A., & Erickson, K. (2018). The unconditional self-kindness scale: Assessing the ability to respond with kindness to threats to the self. *Mindfulness*, *9*, 1713–1722. https://doi.org/10.1007/s12671-018-0912-5

Spinosa, J., Christiansen, P., Dickson, J. M., Lorenzetti, V., & Hardman, C. A. (2019). From Socioeconomic disadvantage to obesity: The mediating role of psychological distress and emotional eating. *Obesity, 27*, 559–564. https://doi.org/10.1002/oby.22402

Szymanski, D. M., & Feltman, C. E. (2014). Experiencing and coping with sexually objectifying treatment: Internalization and resilience. *Sex Roles*, *71*, 159–170. https://doi.org/10.1007/s11199-014-0392-6

Tan, W. H., Sheffield, J., Khoo, S. K., Byrne, G., & Pachana, N. A. (2018). Influences on psychological well-being and ill-being in older women. *Australian Psychologist*, *53*, 203–212. https://doi.org/10.1111/ap.12297

Thogersen-Ntoumani, C., Black, J., Lindwall, M., Whittaker, A., & Balanos, G. M. (2017). Presenteeism, stress resilience, and physical activity in older manual workers: a person-centred analysis. *European Journal of Ageing*, *14*, 385–396. https://doi.org/10.1007/s10433-017-0418-3

Thompson, K. A., Bulls, H. W., Sibille, K. T., Bartley, E. J., Glover, T. L., Terry, E. L., Vaughn, I. A., Cardoso, J. S., Sotolongo, A., Staud, R., Hughes, L. B., Edberg, J. C., Redden, D. T., Bradley, L. A., Goodin, B. R., & Fillingim, R. B. (2018). Optimism and psychological resilience are beneficially associated with measures of clinical and experimental pain in adults with or at risk for knee osteoarthritis. *The Clinical Journal of Pain*, *34*, 1164–1172. https://doi.org/10.1097/AJP.0000000000000642

Thurston, I. B., Howell, K. H., Kamody, R. C., Maclin-Akinyemi, C., & Mandell, J. (2018). Resilience as a moderator between syndemics and depression in mothers living with HIV. *AIDS Care*, *30*, 1257–1264. https://doi.org/10.1080/09540121.2018.1446071

Thurston, I. B., Hardin, R., Kamody, R. C., Herbozo, S., & Kaufman, C. (2018). The moderating role of resilience on the relationship between perceived stress and binge eating symptoms among young adult women. *Eating Behaviors : an International Journal*, *29*, 114–119. https://doi.org/10.1016/j.eatbeh.2018.03.009

Umucu, E., Brooks, J. M., Lee, B., Iwanaga, K., Wu, J.-R., Chen, A., & Chan, F. (2018). Measuring dispositional optimism in student veterans: An item response theory analysis. *Military Psychology*, *30*, 590–597.

https://doi.org/10.1080/08995605.2018.1522161

Vatter, S., McDonald, K. R., Stanmore, E., Clare, L., & Leroi, I. (2018). Multidimensional care burden in Parkinson-related dementia. *Journal of Geriatric Psychiatry and Neurology*, *31*, 319–328. https://doi.org/10.1177/0891988718802104

Williamson, C. D. (2017). *The Impact of Perceived Leadership Behaviors on Subordinate Resilience among Working Professionals*. ProQuest Dissertations Publishing.

Wolfe, T., & Ray, S. (2015). The role of event centrality, coping and social support in resilience and posttraumatic growth among women and men. *The International Journal of Mental Health Promotion*, *17*, 78–96. https://doi.org/10.1080/13642529.2015.1008799

Woodford, M. R., Weber, G., Nicolazzo, Z., Hunt, R., Kulick, A., Coleman, T., Coulombe, S., & Renn, K. A. (2018). Depression and attempted suicide among LGBTQ college students: Fostering resilience to the effects of heterosexism and cisgenderism on campus. *Journal of College Student Development*, *59*, 421–438. https://doi.org/10.1353/csd.2018.0040

Convergent Validity

Grossman, M. R. (2017). *The Structure of Resilience: An Empirical Examination of Resilience Factors*. ProQuest Dissertations Publishing.

Shoss, M. K., Jiang, L., & Probst, T. M. (2018). Bending without breaking: A two-study examination of employee resilience in the face of job insecurity. *Journal of Occupational Health Psychology*, *23*(1), 112–126. https://doi.org/10.1037/ocp0000060

Smith, B. W., Dalen, J., Wiggins, K., Tooley, E., Christopher, P., & Bernard, J. (2008). The Brief Resilience Scale: Assessing the ability to bounce back. *International Journal of Behavioral Medicine, 15,* 194-200. https://doi.org/10.1080/10705500802222972

Anxiety

Autry, A. L. (2016). *The role of loneliness and resilience in psychosocial health for victims of cyberbullying in a college population*. ProQuest Dissertations Publishing.

Bajjani-Gebara, J., Hinds, P., Insel, K., Reed, P., Moore, K., & Badger, T. (2019). Well-being, self-transcendence, and resilience of parental caregivers of children in active cancer treatment: Where do we go from here? *Cancer Nursing*, *42*, 41–52. https://doi.org/10.1097/NCC.0000000000000662

Bonus, J. A., Peebles, A., Mares, M.-L., & Sarmiento, I. G. (2018). Look on the bright side (of media effects): Pokémon Go as a catalyst for positive life experiences. *Media Psychology*, *21*, 263–287. https://doi.org/10.1080/15213269.2017.1305280

Crane, M., Phillips, J., & Karin, E. (2015). Trait perfectionism strengthens the negative effects of moral stressors occurring in veterinary practice. *Australian Veterinary Journal*, *93*(10), 354–360. https://doi.org/10.1111/avj.12366

Flett, J. A. M., Hayne, H., Riordan, B. C., Thompson, L. M., & Conner, T. S. (2019). Mobile mindfulness meditation: A randomised controlled trial of the effect of two popular apps on mental health. *Mindfulness*, *10*, 863–876. https://doi.org/10.1007/s12671-018-1050-9

Flett, J. A. M., Lie, C., Riordan, B. C., Thompson, L. M., Conner, T. S., & Hayne, H. (2017). Sharpen your pencils: Preliminary evidence that adult coloring reduces depressive symptoms and anxiety. *Creativity Research Journal*, *29*, 409–416. https://doi.org/10.1080/10400419.2017.1376505

Gloria, C. T., & Steinhardt, M. A. (2016). Relationships among positive emotions, coping, resilience and mental health. *Stress and Health*, *32*, 145–156. https://doi.org/10.1002/smi.2589

McNair, R. P., & Bush, R. (2016). Mental health help seeking patterns and associations among Australian same sex attracted women, trans and gender diverse people: a survey-based study. *BMC Psychiatry*, *16*, 209–209. https://doi.org/10.1186/s12888-016-0916-4

Smith, B. W., Dalen, J., Wiggins, K., Tooley, E., Christopher, P., & Bernard, J. (2008). The Brief Resilience Scale: Assessing the ability to bounce back. *International Journal of Behavioral Medicine, 15,* 194-200. https://doi.org/10.1080/10705500802222972

Flett, J. A. M., Hayne, H., Riordan, B. C., Thompson, L. M., & Conner, T. S. (2019). Mobile mindfulness meditation: A randomised controlled trial of the effect of two popular apps on mental health. *Mindfulness*, *10*, 863–876. https://doi.org/10.1007/s12671-018-1050-9

Flett, J. A. M., Lie, C., Riordan, B. C., Thompson, L. M., Conner, T. S., & Hayne, H. (2017). Sharpen your pencils: Preliminary evidence that adult coloring reduces depressive symptoms and anxiety. *Creativity Research Journal*, *29*, 409–416. https://doi.org/10.1080/10400419.2017.1376505

Gloria, C. T., & Steinhardt, M. A. (2016). Relationships among positive emotions, coping, resilience and mental health. *Stress and Health*, *32*, 145–156. https://doi.org/10.1002/smi.2589

Tan, W. H., Sheffield, J., Khoo, S. K., Byrne, G., & Pachana, N. A. (2018). Influences on psychological well-being and ill-being in older women. *Australian Psychologist*, *53*, 203–212. https://doi.org/10.1111/ap.12297

Umucu, E., Brooks, J. M., Lee, B., Iwanaga, K., Wu, J.-R., Chen, A., & Chan, F. (2018). Measuring dispositional optimism in student veterans: An item response theory analysis. *Military Psychology*, *30*, 590–597.

Vatter, S., McDonald, K. R., Stanmore, E., Clare, L., & Leroi, I. (2018). Multidimensional care

burden in Parkinson-related dementia. *Journal of Geriatric Psychiatry and Neurology*, *31*, 319–328. https://doi.org/10.1177/0891988718802104

Depression

Autry, A. L. (2016). *The role of loneliness and resilience in psychosocial health for victims of cyberbullying in a college population*. ProQuest Dissertations Publishing.

Bajjani-Gebara, J., Hinds, P., Insel, K., Reed, P., Moore, K., & Badger, T. (2019). Well-being, self-transcendence, and resilience of parental caregivers of children in active cancer treatment: Where do we go from here? *Cancer Nursing*, *42*, 41–52. https://doi.org/10.1097/NCC.0000000000000662

Bonus, J. A., Peebles, A., Mares, M.-L., & Sarmiento, I. G. (2018). Look on the bright side (of media effects): Pokémon Go as a catalyst for positive life experiences. *Media Psychology*, *21*, 263–287. https://doi.org/10.1080/15213269.2017.1305280

Haridas, S., Bhullar, N., & Dunstan, D. A. (2017). What's in character strengths? Profiling strengths of the heart and mind in a community sample. *Personality and Individual Differences*, *113*, 32–37. https://doi.org/10.1016/j.paid.2017.03.006

Meyer, E. C., Kotte, A., Kimbrel, N. A., DeBeer, B. B., Elliott, T. R., Gulliver, S. B., & Morissette, S. B. (2019). Predictors of lower-than-expected posttraumatic symptom severity in war veterans: The influence of personality, self-reported trait resilience, and psychological flexibility. *Behaviour Research and Therapy*, *113*, 1–8. https://doi.org/10.1016/j.brat.2018.12.005

McNair, R. P., & Bush, R. (2016). Mental health help seeking patterns and associations among Australian same sex attracted women, trans and gender diverse people: a survey-based study. *BMC Psychiatry*, *16*, 209–209. https://doi.org/10.1186/s12888-016-0916-4

Newton-John, T. R. O., Mason, C., & Hunter, M. (2014). The Role of Resilience in Adjustment and Coping With Chronic Pain. *Rehabilitation Psychology, 59*, 360–365. https://doi.org/10.1037/a0037023

Smith, B. W., Dalen, J., Wiggins, K., Tooley, E., Christopher, P., & Bernard, J. (2008). The Brief Resilience Scale: Assessing the ability to bounce back. *International Journal of Behavioral Medicine, 15,* 194-200. https://doi.org/10.1080/10705500802222972

Tan, W. H., Sheffield, J., Khoo, S. K., Byrne, G., & Pachana, N. A. (2018). Influences on psychological well-being and ill-being in older women. *Australian Psychologist*, *53*, 203–212. https://doi.org/10.1111/ap.12297

Umucu, E., Brooks, J. M., Lee, B., Iwanaga, K., Wu, J.-R., Chen, A., & Chan, F. (2018). Measuring dispositional optimism in student veterans: An item response theory analysis. *Military Psychology*, *30*, 590–597.

Vatter, S., McDonald, K. R., Stanmore, E., Clare, L., & Leroi, I. (2018). Multidimensional care burden in Parkinson-related dementia. *Journal of Geriatric Psychiatry and Neurology*, *31*, 319–328. https://doi.org/10.1177/0891988718802104

Optimism

Carr, K., & Kellas, J. K. (2018). The role of family and marital communication in developing resilience to family-of-origin adversity. *Journal of Family Communication*, *18*, 68–84. https://doi.org/10.1080/15267431.2017.1369415

Cassidy, T. (2013). Benefit finding through caring: The cancer caregiver experience. *Psychology & Health*, *28*, 250–266. https://doi.org/10.1080/08870446.2012.717623

Smith, B., Tooley, E., Christopher, P., & Kay, V. (2010). Resilience as the ability to bounce back from stress: A neglected personal resource? *The Journal of Positive Psychology*, *5*, 166–

176. https://doi.org/10.1080/17439760.2010.482186

Smith, B. W., Dalen, J., Wiggins, K., Tooley, E., Christopher, P., & Bernard, J. (2008). The Brief Resilience Scale: Assessing the ability to bounce back. *International Journal of Behavioral Medicine, 15,* 194-200. https://doi.org/10.1080/10705500802222972

Smith, B. W., Kay, V. S., Hoyt, T. V., & Bernard, M. L. (2009). Predicting the anticipated emotional and behavioral responses to an avian flu outbreak. *American Journal of Infection Control*, *37*, 371–380. https://doi.org/10.1016/j.ajic.2008.08.007

Tan, W. H., Sheffield, J., Khoo, S. K., Byrne, G., & Pachana, N. A. (2018). Influences on psychological well-being and ill-being in older women. *Australian Psychologist*, *53*, 203–212. https://doi.org/10.1111/ap.12297

Thompson, K. A., Bulls, H. W., Sibille, K. T., Bartley, E. J., Glover, T. L., Terry, E. L., Vaughn, I. A., Cardoso, J. S., Sotolongo, A., Staud, R., Hughes, L. B., Edberg, J. C., Redden, D. T., Bradley, L. A., Goodin, B. R., & Fillingim, R. B. (2018). Optimism and psychological resilience are beneficially associated with measures of clinical and experimental pain in adults with or at risk for knee osteoarthritis. *The Clinical Journal of Pain*, *34*, 1164–1172. https://doi.org/10.1097/AJP.0000000000000642

Umucu, E., Brooks, J. M., Lee, B., Iwanaga, K., Wu, J.-R., Chen, A., & Chan, F. (2018). Measuring dispositional optimism in student veterans: An item response theory analysis. *Military Psychology*, *30*, 590–597.

Social Support

Carr, K., & Kellas, J. K. (2018). The role of family and marital communication in developing resilience to family-of-origin adversity. *Journal of Family Communication*, *18*, 68–84. https://doi.org/10.1080/15267431.2017.1369415

Cassidy, T. (2013). Benefit finding through caring: The cancer caregiver experience. *Psychology & Health*, *28*, 250–266. https://doi.org/10.1080/08870446.2012.717623

Gage, G. S. (2017). Social support and positive health practices in black late adolescents: The role of mediating variables. *Clinical Nursing Research*, *26*, 93–113. https://doi.org/10.1177/1054773815594579

Newton-John, T. R. O., Mason, C., & Hunter, M. (2014). The Role of Resilience in Adjustment and Coping With Chronic Pain. *Rehabilitation Psychology, 59*, 360–365. https://doi.org/10.1037/a0037023

Nightingale, S., Spiby, H., Sheen, K., & Slade, P. (2018). Posttraumatic stress symptomatology following exposure to perceived traumatic perinatal events within the midwifery profession: The impact of trait emotional intelligence. *Journal of Advanced Nursing*, *74*, 2115–2125. https://doi.org/10.1111/jan.13719

Sánchez, J., Muller, V., Chan, F., Brooks, J. M., Iwanaga, K., Tu, W. M., Umucu, E., & Crespo-Jones, M. (2019). Personal and environmental contextual factors as mediators between functional disability and quality of life in adults with serious mental illness: a cross-sectional analysis. *Quality of Life Research*, *28*, 441–450. https://doi.org/10.1007/s11136-018-2006-1

Smith, B., Tooley, E., Christopher, P., & Kay, V. (2010). Resilience as the ability to bounce back from stress: A neglected personal resource? *The Journal of Positive Psychology*, *5*, 166–176. https://doi.org/10.1080/17439760.2010.482186

Smith, B. W., Dalen, J., Wiggins, K., Tooley, E., Christopher, P., & Bernard, J. (2008). The Brief Resilience Scale: Assessing the ability to bounce back. *International Journal of Behavioral Medicine, 15,* 194-200. https://doi.org/10.1080/10705500802222972

Smith, B. W., Kay, V. S., Hoyt, T. V., & Bernard, M. L. (2009). Predicting the anticipated emotional and behavioral responses to an avian flu outbreak. *American Journal of Infection Control*, *37*, 371–380. https://doi.org/10.1016/j.ajic.2008.08.007

Tan, W. H., Sheffield, J., Khoo, S. K., Byrne, G., & Pachana, N. A. (2018). Influences on psychological well-being and ill-being in older women. *Australian Psychologist*, *53*, 203–212. https://doi.org/10.1111/ap.12297

Wolfe, T., & Ray, S. (2015). The role of event centrality, coping and social support in resilience and posttraumatic growth among women and men. *The International Journal of Mental Health Promotion*, *17*, 78–96. https://doi.org/10.1080/13642529.2015.1008799

Zumsteg, D. M. (2016). *The impact of social support and resilience on surgical outcomes*. ProQuest Dissertations Publishing.

Stress

Autry, A. L. (2016). *The role of loneliness and resilience in psychosocial health for victims of cyberbullying in a college population*. ProQuest Dissertations Publishing.

Crane, M., Phillips, J., & Karin, E. (2015). Trait perfectionism strengthens the negative effects of moral stressors occurring in veterinary practice. *Australian Veterinary Journal*, *93*(10), 354–360. https://doi.org/10.1111/avj.12366

Crane, M. F., & Searle, B. J. (2016). Building resilience through exposure to stressors: The effects of challenges versus hindrances. *Journal of Occupational Health Psychology*, *21*, 468–479. https://doi.org/10.1037/a0040064

Flett, J. A. M., Lie, C., Riordan, B. C., Thompson, L. M., Conner, T. S., & Hayne, H. (2017). Sharpen your pencils: Preliminary evidence that adult coloring reduces depressive symptoms and anxiety. *Creativity Research Journal*, *29*, 409–416.

https://doi.org/10.1080/10400419.2017.1376505

Fullen, M. C., Richardson, V. E., & Granello, D. H. (2018). Comparing successful aging, resilience, and holistic wellness as predictors of the good life. *Educational Gerontology*, *44*, 459–468. https://doi.org/10.1080/03601277.2018.1501230

Gloria, C. T., & Steinhardt, M. A. (2016). Relationships among positive emotions, coping, resilience and mental health. *Stress and Health*, *32*, 145–156. https://doi.org/10.1002/smi.2589

Kemper, K. J., & Rao, N. (2017). Brief Online Focused Attention Meditation Training: Immediate Impact. *Journal of Evidence-Based Complementary & Alternative Medicine*, *2*, 395–400. https://doi.org/10.1177/2156587216663565

McKay, S., Skues, J. L., & Williams, B. J. (2018). With risk may come reward: Sensation seeking supports resilience through effective coping. *Personality and Individual Differences*, *121*, 100–105. https://doi.org/10.1016/j.paid.2017.09.030

McNair, R. P., & Bush, R. (2016). Mental health help seeking patterns and associations among Australian same sex attracted women, trans and gender diverse people: a survey-based study. *BMC Psychiatry*, *16*, 209–209. https://doi.org/10.1186/s12888-016-0916-4

Smith, B., Tooley, E., Christopher, P., & Kay, V. (2010). Resilience as the ability to bounce back from stress: A neglected personal resource? *The Journal of Positive Psychology*, *5*, 166–176. https://doi.org/10.1080/17439760.2010.482186

Smith, B. W., Dalen, J., Wiggins, K., Tooley, E., Christopher, P., & Bernard, J. (2008). The Brief Resilience Scale: Assessing the ability to bounce back. *International Journal of Behavioral Medicine, 15,* 194-200. https://doi.org/10.1080/10705500802222972

Tan, W. H., Sheffield, J., Khoo, S. K., Byrne, G., & Pachana, N. A. (2018). Influences on

psychological well-being and ill-being in older women. *Australian Psychologist*, *53*, 203–212. https://doi.org/10.1111/ap.12297

Thurston, I. B., Howell, K. H., Kamody, R. C., Maclin-Akinyemi, C., & Mandell, J. (2018). Resilience as a moderator between syndemics and depression in mothers living with HIV. *AIDS Care*, *30*, 1257–1264. https://doi.org/10.1080/09540121.2018.1446071