Research Report

Improving Noncognitive Constructs for Career Readiness and Success: A Theory of Change for Postsecondary, Workplace, and Research Applications

ETS RR–22-04

Kevin M. Williams
Michelle P. Martin-Raugh
Jennifer E. Lentini

December 2022

ETS Research Report Series

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Report series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Report series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

ETS Research Report Series ISSN 2330-8516

RESEARCH REPORT

Improving Noncognitive Constructs for Career Readiness and Success: A Theory of Change for Postsecondary, Workplace, and Research Applications

Kevin M. Williams, Michelle P. Martin-Raugh, & Jennifer E. Lentini

ETS , Princeton, NJ

Researchers and practitioners in postsecondary and workplace settings recognize the value of noncognitive constructs in predicting academic and vocational success but also perceive that many students or employees are lacking in these areas. In turn, there is increased interest in interventions designed to enhance these constructs. We provide an empirically informed theory of change (ToC) that describes the inputs, mechanisms, and outputs of noncognitive construct interventions (NCIs). The components that inform this ToC include specific relevant constructs that are amenable to intervention, intervention content and mechanisms of change, methodological considerations, moderators of program efficacy, recommendations for evaluating NCIs, and suggested outcomes. In turn, NCIs should provide benefits to individuals, institutions, and society at large and also advance our scientific understanding of this important phenomenon.

Keywords Noncongnitive constructs; academic success; vocational success; training

doi:10.1002/ets2.12346

Preparing individuals to successfully enter the workforce and to maintain effective performance upon entry are fundamental duties of educational and organizational institutions. Nearly half of all jobs in the United States require some amount of postsecondary education, and virtually all remaining jobs require relevant work experience, on-the-job training, or both (Bureau of Labor Statistics, 2018, Table 1.7). However, employers consistently regard recent graduates as deficient in various career readiness competencies (National Association of Colleges and Employers [NACE], 2018). In the United States alone, organizations spend approximately 171 billion dollars annually on employee learning and development interventions (Green & McGill, 2011). The positive impact of successful training programs may be felt not only at the individual and organizational level, but it may even influence economic growth at the national level (Aguinis & Kraiger, 2009). These trends highlight continuing opportunities for both higher education and workplace institutions to improve the efficacy and efficiency of career readiness training.

Interestingly, and perhaps unsurprisingly, the majority of competencies that employers find lacking in job candidates are those that are usually not an explicit focus of postsecondary courses. For instance, employers typically rate competencies such as social skills, teamwork/collaboration, professionalism/work ethic, leadership, and global/multicultural fluency as at least somewhat essential for the workplace (e.g., Capelli, 2012; NACE, 2018; Schanzenbach et al., 2016) but also underdeveloped in recent college graduates (NACE, 2018). One commonality shared across these competencies is that they can be characterized as noncognitive constructs. These constructs are distinct from other individual differences (i.e., technical knowledge or skills, physical abilities, job experience) and cognitive constructs (i.e., problem solving, literacy, and traditional definitions of intelligence). Noncognitive constructs—often referred to as "soft skills" or "21st century skills" (e.g., Heckman & Kautz, 2012)—have been described as "demonstrable personality, motivational, attitudinal, self-regulatory, and learning approach constructs for which there are differences among people, which standardized tests of cognitive ability are not primarily designed to measure, and the behavioral expression of which is considered useful" (Klieger et al., 2015, p. 3; see also Duckworth & Yeager, 2015; Kautz et al., 2014; Kyllonen, 2012). Additional examples of noncognitive constructs include personality traits, socioemotional skills, communication skills, motivation, self-efficacy, and interests, among others. Support for these constructs' occupational relevance goes beyond employer surveys, as various meta-analyses and reviews have corroborated their predictive validity in workplace settings (e.g., Barrick

Corresponding author: K. Williams, E-mail: kmwilliams@ets.org

et al., 2002; Barrick & Mount, 1991; Judge & Bono, 2001; Roberts et al., 2007; Schmidt & Hunter, 1998; Spector et al., 2006; Van Iddekinge et al., 2011, 2012). Moreover, these studies often suggest that noncognitive skills predict job performance independently of cognitive constructs (e.g., provide incremental validity; Schmidt & Hunter, 1998) and to a similar magnitude (e.g., Roberts et al., 2007).

Improving Noncognitive Constructs Through Intervention: A Theory of Change

Given the relevance of noncognitive constructs to the workplace, it is important to understand how they can be enhanced. One of the appealing features of noncognitive constructs is that they may be as malleable as cognitive skills, if not more so (Aguinis & Kraiger, 2009; Boyatzis, 2008; Heckman, 2000; Heckman & Kautz, 2012). Indeed, several meta-analyses and reviews have demonstrated that various noncognitive constructs may be altered to varying degrees through intervention (e.g., Arthur Jr. et al., 2003; Barth & Lannen, 2011; Bezrukova et al., 2016; Bowen & Neill, 2013; Delise et al., 2010; Lacerenza et al., 2017; Lipsey & Wilson, 1993; Salas et al., 2008). A recent review by Martin-Raugh et al. (2020) synthesized 39 meta-analyses and systematic reviews on the malleability of three broad categories of noncognitive skills, drawing upon a diverse set of disciplines including industrial/organizational (I/O) psychology, educational psychology, and medicine, among others. We will be using the findings from Martin-Raugh et al. (2020) to develop a theory of change (ToC) to inform empirically supported, actionable recommendations for delivering and evaluating noncognitive construct interventions (NCIs) in higher education, occupational, and research settings.

The concept of ToCs date back to at least the 1970s (e.g., Weiss, 1972) and are used to articulate the inputs, mechanisms, and outcomes of change efforts. Put differently, ToCs describe what is being changed, why or how it changes, and the impact of these changes. ToCs arose around a similar period as the related concept of theories of action (ToAs; e.g., Argyris & Schon, 1974; Wholey, 1979), which "constitute a framework that individuals use to guide, interpret, or justify" courses of action (Malen et al., 2002, p. 113). ToCs and ToAs each illustrate a logical sequence of steps in service of specific outcomes, though ToCs tend to emphasize mechanisms of change whereas ToAs are more concerned with implementation or delivery (George, 2019). The ToA versus ToC comparison is similar to Weiss's (1997) distinction between implementation theory and program theory, respectfully. In other words, a ToA may be viewed as a specific application of a ToC. In the context of noncognitive interventions, a ToC may describe changes at the construct level (teamwork) whereas a ToA would focus on a particular intervention (crew resource management training; e.g., Salas et al., 2001).

ToCs incorporate the concept of a logic model (e.g., McLaughlin & Jordan, 1999), which is a graphical cause-and-effect depiction describing particular construct changes. By including and elaborating upon these depictions, ToCs may be used in program evaluation as an accessible method for describing the intended results of change efforts (e.g., McLaughlin & Jordan, 1999). Although there is some stylistic variation in ToC graphical depictions, they generally include three primary elements: First, the ToC lists the inputs or targets of the intervention program. Second, the ToC describes the processes that motivate and explain longitudinal changes, which may also be described as the mechanisms of action. Finally, the effects of these actions or mechanisms are described with intermediate and ultimate effects sometimes categorized separately. The causal pathways between these elements are also included in the ToC figure. An additional element — unintended consequences (see Geisinger, 2010) — is also occasionally discussed within ToCs.

ToCs' status as an essential foundational component of planning, implementing, and evaluating longitudinal interventions (e.g., Ployhart & Vandenberg, 2010; see review by Msila & Setlhako, 2013) is represented in their application across a diverse set of high-stakes initiatives, including urban education reform (Connell & Klem, 2000), complex psychological and medical interventions (De Silva et al., 2014), and development finance (Jackson, 2013). Similarly, a recent ToA example was provided by Bennett (2010), who illustrated a specific application of formative and summative assessment for improving K – 12 student learning. These examples illustrate that both ToCs and ToAs may serve as important research bases for informing intervention development, including situations where evaluation results are unavailable (Leusner & Lyon, 2008).

Our ToC — NCI ToC (Figure 1) — illustrates the inputs, mechanisms, and intended outcomes of NCIs as informed by Martin-Raugh et al. (2020). Noncognitive constructs relevant to academic and workplace success may be altered through specific interventions (Path 1) or more general mechanisms of change (Path 2). These general mechanisms may be present in specific existing interventions and may also inform the development of new training programs (Path 3). When implemented properly, the positive causal impact of these interventions should manifest in various outcomes (Path 4). The remainder of this article will detail the NCI ToC elements.

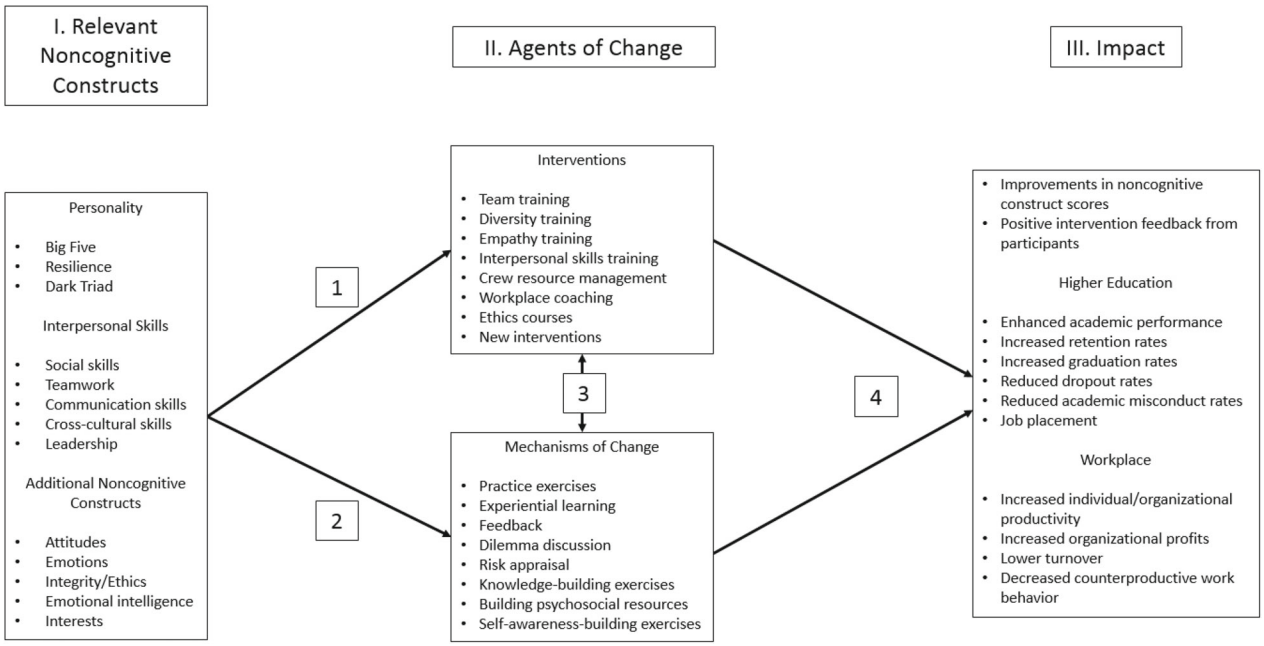

Figure 1 Noncognitive construct intervention theory of change (NCI ToC). *Note*. List of relevant noncognitive constructs is based on Martin-Raugh et al. (2020) and is not intended to be exhaustive.

Relevant Noncognitive Constructs

The noncognitive constructs targeted by academic or workplace interventions are a key concern. These constructs are typically identified by reviewing the characteristics that organizations value in their employees, predict job performance, and may be perceived as deficient in typical employee candidate pools (NACE, 2018). As such, a primary goal of the Martin-Raugh et al. (2020) literature review was to locate relevant constructs whose malleability was supported by substantial empirical evidence. Martin-Raugh et al. identified three broad categories for which some degree of longitudinal change was corroborated in the form of meta-analyses or systematic reviews: personality, social skills (communication skills, interpersonal skills, leadership, teamwork, cross-cultural skills), and additional noncognitive constructs (attitudes, self-concept, emotion, motivation).

In reviewing this evidence, a critical distinction must first be made regarding changes due to effortful intervention versus naturally occurring changes observed in the absence of intervention. Although it may be tempting to view naturally occurring changes as evidence that these constructs are amenable to effortful change, we would argue against this conclusion. Instead, just one alternative explanation is that it is equally plausible—if not more so—that naturally occurring changes represent an opposing force that intervention programs must overcome. Similarly, conflating natural and effortful change may unfairly discourage efforts to deliberately improve constructs that are highly stable over the natural life course. Furthermore, it is possible that proposed mechanisms of natural change (e.g., genetics, significant or traumatic life experiences; see Roberts et al., 2006) cannot be translated into workplace interventions, which are the focus of our NCI ToC. Therefore, we assert that the most valuable conclusion to be drawn from evidence of naturally occurring change is in emphasizing that studies of effortful change should employ an experimental design. That is, the inclusion of a control group in which the intervention is not applied is imperative in distinguishing intervention effects from natural longitudinal trajectories. This strategy should facilitate research that supports causal inference for deliberate interventions. It should also be noted that we maintain this assertion regardless of whether naturally occurring changes arise from factors that are biological, environmental, or some combination. Researchers may, however, be interested in noting any significant life events that may have coincidentally occurred during training in order to control for these variables. In turn, in informing academic and workplace interventions, as well as areas where further research is needed, the NCI ToC's program components focus on the research evidence involving effortful change.

Based on the Martin-Raugh et al. (2020) review, the strongest empirical support for effortful malleability was located for social skills, teamwork, attitudes, and emotion with effect sizes typically in the moderate range. Mixed or limited

support was reported for cross-cultural skills, self-concept, motivation, communication skills, leadership, integrity, and some personality constructs (e.g., resilience). It should also be noted that within these constructs, there is some variation regarding specific subsets of constructs. For instance, attitudes that are more strongly held by an individual are more resistant to change. Similarly, when discussing emotion malleability, there is some debate regarding whether negative or positive affect is easier to manipulate, though there is evidence that guilt is easier to induce than other emotions (e.g., fear, worry, anxiety). There was no evidence located for vocational interests, emotional intelligence, or personality constructs such as the Big Five (see McCrae & Costa Jr., 2008) or Dark Triad (narcissism, Machiavellianism, psychopathy; Paulhus & Williams, 2002). However, these findings largely reflect a lack of research rather than conclusive evidence of construct immutability. Therefore, future efficacy studies and eventually meta-analytic research on these constructs, if feasible, would be invaluable. We also note that malleability evidence may exist in primary studies or outside academic or occupational settings, but was beyond the scope of Martin-Raugh et al. (2020).

Agents of Change

Once workplace-relevant noncognitive constructs have been identified for intervention, a logical next step is to identify specific programs whose efficacy is empirically supported (Figure 1, Path 1). Martin-Raugh et al. (2020) noted that programs such as team training, small group training, crew resource management training, interpersonal skills training, and leadership training demonstrated efficacy in improving various social skills. Attitudes related to several forms of counterproductive work behavior (CWB) such as sexual assault, substance abuse, and racial prejudice may be addressed through programs such as diversity training, education programs, and integrity or ethics courses. Additional support was located for programs involving workplace coaching, emotion regulation, and empathy training in improving various noncognitive constructs.

Although these reviews and meta-analyses were illuminating, they often grouped interventions into broad categories (Martin-Raugh et al., 2020). Therefore, although training efficacy was not typically evaluated at the level of specific intervention programs, some research examined the strategies employed by intervention programs (Figure 1, Path 2). In turn, the value inherent in understanding these strategies lies not only in implying the causal mechanisms of specific interventions, but in informing the development of new programs as well (Figure 1, Path 3). As previously mentioned, because these mechanisms are intended to explain effortful change, proposed agents of naturally occurring change (e.g., genetics, life events) are not considered. Instead, one commonality across many of the most effective NCIs is that they seek to produce change through practice or experiential learning and associated feedback with examples including communication skills (e.g., Barth & Lannen, 2011), attitudes (e.g., Bezrukova et al., 2016), self-concept (e.g., Bowen & Neill, 2013), emotion (e.g., Teding van Berkhout & Malouff, 2016), and integrity (Schlaefli et al., 1985). Various other specific examples were also noted; for example, some effective emotion regulation courses operated by heightening risk appraisal associated with problematic workplace behaviors. Specifically, these programs tend to be more effective when risk perceptions (i.e., beliefs about personal vulnerability to harm) and perceived severity (i.e., beliefs about the seriousness of the hazard's negative consequences) are increased. Alternatively, integrity courses demonstrated efficacy when incorporating personality development theories and dilemma discussion. Constructs such as leadership are developed with methods intended to increase knowledge, self-awareness, and interpersonal skills (e.g., Lacerenza et al., 2017). Resilience-building programs adopt an entirely different approach: They are designed to improve individuals' psychosocial resources such as self-efficacy and optimism, which serve as protective factors against stressors (see Vanhove et al., 2016). For other constructs, the mechanisms of change are less clear; for instance, Klein et al. (2009) noted that although teamwork interventions typically target problem-solving, interpersonal relations, and goal-setting, "no one is quite sure how and why these interventions work" (p. 182). Overall, these mechanisms suggest various strategies for successfully stimulating noncognitive construct change, along with avenues for future research and new NCI development.

Impact

Improvements in Noncognitive Construct Scores

Assuming appropriate implementation, NCIs' impact may be observed in myriad outcomes (Figure 1, Path 4). Direct changes in noncognitive construct scores may be the most significant outcome of NCI efforts. This is particularly true from

a theoretical or research perspective where such results may describe the general malleability of noncognitive constructs. A constructs' amenability or resistance to effortful change may be considered one of its fundamental definitional characteristics. Overall, a common theme across NCIs is that they are generally effective in improving noncognitive construct scores (Martin-Raugh et al., 2020; see also the Relevant Noncognitive Constructs section). Importantly, these changes have been observed via self-report, other-report, or performance-based assessments (Martin-Raugh et al., 2020). However, these changes are typically short-term as few studies included in meta-analyses and reviews evaluate long-term changes. Of those that do, changes sustained over longer periods are rare (Martin-Raugh et al., 2020). Comparisons among various assessment options are also relatively understudied. As an example, Martin-Raugh et al. (2020) noted conflicting results for the use of self- or other-reports versus behavioral observations, with the former supported for some constructs (communication skills) and no differences observed in others (teamwork). These results suggest that although much is known about the malleability of noncognitive constructs, some critical research avenues have yet to be explored sufficiently.

Alternative Outcome Measurement

Conversely, NCI evaluators may be interested in outcomes beyond direct changes in the noncognitive construct. It is possible that direct changes in the construct do not guarantee changes in other related outcomes. In turn, the NCI evaluation must be designed in a manner that will accommodate these outcomes. Although not always described as such, many of the outcomes used in noncognitive construct change research align with those described by Kirkpatrick (e.g., 1996). Kirkpatrick's model is widely used in organizational settings and describes four levels of program evaluation metrics—reaction, learning, behavior, and results—that progress from short-term to long-term.

The first level—reaction—assess trainees' level of satisfaction with the intervention. In NCI research, this outcome is operationalized in a similar fashion across workplace and higher education settings and is sometimes referred to as affective outcomes (see Martin-Raugh et al., 2020). Aside from participants' evaluations of the intervention, these outcomes may capture participants' attitudes related to the noncognitive construct of interest (cross-cultural skills, leadership, teamwork, motivation). Positive reactions are typically desired, but critical responses may also be informative. Reactions outcomes are commonly used in NCI evaluations due to their ease of assessment. However, in evaluating an intervention's efficacy, participant reactions are among the most criticized outcome measures given their limited associations with other, more valuable outcomes (Martin-Raugh et al., 2020). Alternatively, reactions outcomes may be more informative if separated into affective and utility subcategories, with the former describing participants' enjoyment of the intervention whereas the latter captures perceptions of usefulness and transferability to the workplace (Alliger et al., 1997). Ultimately, reactions to training may be most instructive in modifying interventions to increase participant enjoyment and perceived relevance. Participant feedback may also be critical in identifying and addressing potential usability and accessibility issues. In turn, these alterations may be beneficial in encouraging new employees or students to participate and in improving intervention efficacy. Practically speaking, an intervention cannot be evaluated if it is too unappealing to recruit participants.

The second outcome—learning—refers to pre–post changes in relevant constructs associated with training. The aforementioned direct changes in noncognitive construct scores represent a clear parallel between NCI outcomes and Kirkpatrick's (1996) learning outcome. For some noncognitive constructs (personality, cross-cultural skills, leadership, teamwork, attitudes, emotion, motivation), direct changes in these assessments are occasionally augmented with additional outcomes such as construct-relevant self-efficacy or knowledge (Martin-Raugh et al., 2020). In the case of knowledge, these outcomes are sometimes described as cognitive outcomes in the NCI literature. There is some evidence (e.g., attitude change research) that knowledge outcomes are more amenable to change than other outcome types, but further investigation is warranted.

Kirkpatrick's (1996) third and fourth outcomes—behavior and results—represent changes in on-the-job behavior (also referred to as transfer of training) and various objective performance outcomes such as productivity or errors, respectively. Although NCI research typically does not distinguish between these two groups of outcomes in the same way as traditional workplace intervention researchers, obvious differences between workplace and postsecondary settings exist. These outcomes also differ from performance-based assessments of noncognitive constructs, which may involve responses to hypothetical or simulated situations. In the workplace, NCI behavior and results outcome examples include ratings from clients/customers or patients or from objective data such as productivity. Financial outcomes permit organizations to calculate the return on investment from NCI efforts (e.g., Cascio & Boudreau, 2008). Some outcomes and their associated

interventions are of varying importance across industries. Communication skills, unsurprisingly, are often emphasized in industries or occupations such as healthcare, sales, customer service, or other positions that involve critical teamwork contexts such as aviation crews (see Martin-Raugh et al., 2020). Alternatively, postsecondary outcomes include academic performance such as course grades and degree attainment. Retention and graduation provide an example of short-term versus long-term impact: The former is typically defined as continued enrollment 1 year after entering college, whereas the latter is assessed several years later depending on the type of degree, program, or institution (e.g., 2-year vs. 4-year degrees; see McFarland et al., 2017). When the assessment of behavioral outcomes is not feasible, behavioral intentions are occasionally measured instead.

Despite differences in the specific performance-based outcomes across workplace and postsecondary settings, it is possible that these outcomes differ in context only. These similarities suggest common noncognitive antecedents, which may be targeted through comparable interventions. For instance, NCIs designed to improve interpersonal skills should cause more frequent and effective social interactions in both workplace and postsecondary settings. Similarly, parallels may be drawn between CWB and academic misconduct, suggesting that interventions addressing constructs such as ethics or the Dark Triad may be equally relevant across contexts. Therefore, many of the causal mechanisms underlying successful interventions may be similar across settings, differing primarily in appropriately contextualized presentations.

In practice, similar to learning outcomes, behavior and results impact may be assessed through various methods, including self-reports, observer ratings, or objective measures (e.g., sales figures; see Martin-Raugh et al., 2020). Overall, these outcomes have been applied in research involving a diverse set of noncognitive constructs, including personality, communication skills, cross-cultural skills, leadership, attitudes, emotion, motivation, and teamwork (Martin-Raugh et al., 2020). However, although behavior and results outcomes may be of most value to various stakeholders, there are also some challenges in gathering these data. Collecting these types of outcomes may be less practical than reactions or learning outcomes. Furthermore, as outcomes become more distal, they are likely more susceptible to influences beyond the NCI.

It should also be emphasized that despite its popularity in industrial/organizational psychology, the Kirkpatrick (1996) framework is not without criticisms (e.g., Bates, 2004; Giangreco et al., 2010; Holton III & Naquin, 2004), which have led some researchers to propose alternative evaluation approaches (e.g., Alliger et al., 1997; Beer, 1990; Brinkerhoff & Dressler, 2015; Holton III & Naquin, 2004). For example, Brinkerhoff and Dressler (2015) state that the Kirkpatrick model—among other criticisms—does not sufficiently describe why or how training efforts are (or are not) effective. In response, the authors propose a model that potentially addresses these concerns, presented in a format similar to a ToC. Further research in which these alternatives are applied in NCIs would be beneficial.

Quantifying Impact

After selecting appropriate evaluation outcomes, metrics for quantifying noncognitive construct change or intervention efficacy must be considered. Naturally, these considerations are codependent with the statistical methodology selected for evaluating change. NCI efficacy research tends to emphasize effect sizes rather than statistical significance (Martin-Raugh et al., 2020). Effect size calculation is not only necessary for the meta-analytic combination of primary studies, but also provides established interpretation guidelines (e.g., Cohen, 1988) and avoids many of the problematic issues inherent in statistical significance testing (e.g., Kline, 2013). As such, although the criterion for appraising a NCI as "successful" is largely arbitrary, most meta-analytic estimates report moderate effect sizes (e.g., $d \approx 0.50 - 0.79$ in a desirable direction; Martin-Raugh et al., 2020). A small effect size ($d \approx 0.20 - 0.49$ in a desirable direction) would appear to be a reasonable minimum requirement. Of note, these guidelines utilize Cohen's (1988) largely theoretical interpretation benchmarks; conversely, alternative data-driven guidelines for effect size interpretation may also be considered (e.g., Bosco et al., 2015; Gignac & Szodorai, 2016; Paterson et al., 2015). Regardless, describing the precision of these estimates through confidence intervals is also straightforward.

Intermediate Versus Ultimate Impact

Unlike the graphical depictions of other ToCs, intermediate and ultimate outcomes in the NCI ToC are combined into a single category. This description largely reflects the notion that most intermediate and ultimate outcomes differ only by the temporal period in which they are assessed (i.e., quantitatively as opposed to qualitatively). The specific time period distinguishing short-term versus long-term impact is largely arbitrary, as many definitions have been observed in the literature (Martin-Raugh et al., 2020). NCI setting may influence this decision: studies tracking noncognitive change over

several years appear to be more common in research contexts, whereas postsecondary and workplace institutions are likely less willing or able to monitor progress over such long durations in operational applications. Instead, given that organizations often conduct performance evaluations annually, 1 year may represent a practical demarcation point. Regardless, assessing ultimate impact is particularly important for NCIs as noncognitive construct change can be a slow process and short-term changes that do not persist could justify retraining efforts. Another practical consideration is that the risk of NCI participant attrition increases over time, which disproportionately threatens assessment of ultimate impact and must be factored into initial sample sizes. However, it may also be beneficial to consider participant attrition to be an outcome measure itself, as it may reflect poor academic or workplace performance (i.e., dropout or termination, respectively) or deficits in specific noncognitive constructs such as persistence or motivation. Martin-Raugh et al.'s (2020) review noted that studies examining the long-term effects of NCIs are few and tend to generate mixed results (e.g., teamwork, attitudes), suggesting further research is needed.

Beyond temporal distinctions, it is possible that the dynamic between intermediate and ultimate impact includes causal links or other dependencies. For example, short-term direct changes in noncognitive construct may explain longer term behavioral changes. Alternatively, intermediate and ultimate impact may occur in stages, in that short-term changes are prerequisites for long-term effects. Using longitudinal data, statistical techniques such as mediation analysis through structural equation modeling may be used to evaluate these possibilities (e.g., Cole & Maxwell, 2003).

Unintended Consequences

The impact component of the NCI ToC describes positive outcomes of effective NCIs. Conversely, unintended consequences are undesirable outcomes that may arise from the NCIs or action mechanisms. To clarify, these potential consequences are observed when the NCIs are executed as outlined. That is, unintended consequences are not the result of interventions conducted in a manner contraindicated by the NCI ToC recommendations. For instance, it is possible that moderator effects may cause learners who already score highly on targeted constructs (e.g., conscientiousness) to benefit more than individuals with low scores—a phenomenon known as the Matthew effect (Merton, 1968). Thus, NCIs may eventually widen the noncognitive construct gap between higher and lower skilled individuals. However, we would argue that this widening would likely only occur if moderating variables such as preintervention levels of relevant noncognitive constructs are not considered during intervention planning and implementation. For example, training for individuals weaker in target constructs could be delivered in a greater frequency and duration (i.e., dosage). In other words, this outcome should only occur if the NCI ToC recommendations are not abided, suggesting that this issue does not qualify as an unintended consequence.

Conversely, one troublesome unintended consequence of NCIs is that they may encourage participants to exaggerate or fabricate assessment responses or behavior in a manner that would suggest noncognitive improvement that did not in fact occur. This possibility likely increases if participants are incentivized to demonstrate such improvement. This issue is similar to those discussed in various summative assessment contexts such as college admissions, workplace selection, or professional accreditation (e.g., Birkeland et al., 2006). In other words, social desirability, impression management, coaching, or other influences may threaten the integrity of formative assessment results. Thus, improvements in noncognitive construct scores may not reflect actual improvements in the construct of interest. In research contexts, the analogous and equally serious consequence would involve falsification of empirical data supporting the effectiveness of NCIs. These concerns highlight the importance of further research and development into methods for deterring and detecting response manipulation, as well as assessment approaches that are less susceptible to faking, such as performance-based measures as opposed to self-report measures.

A second unintended consequence of NCIs is that some participants may use their improved noncognitive proficiencies for destructive purposes. For instance, some individuals may use newfound interpersonal skills to unethically manipulate customers, clients, or coworkers. A similar concern had been raised in the context of psychopathy interventions in forensic settings, for instance (Rice et al., 1992). Although these concerns have been subsequently discredited (e.g., D'Silva et al., 2004), less is known about this topic in postsecondary and workplace contexts. Similarly, there is precedence that interventions may have a counterproductive impact on certain noncognitive constructs; for example, diversity training programs occasionally reinforce some participants' stereotypical attitudes toward specific racial groups (see Bezrukova et al., 2016). These concerns highlight the need to identify participant characteristics—such as the Dark Triad, low integrity, or poor ethics—that may interact with NCIs in damaging ways. Further research in these areas, as

well as integrating measures to screen for these characteristics prior to NCI implementation, should aid in mitigating these consequences.

Discussion

Postsecondary and workplace interventions aimed at improving noncognitive constructs complement those targeting more traditional areas such as knowledge, technical skills, or cognitive constructs. The NCI ToC is intended to provide a research basis for planning and evaluating these interventions. The NCI ToC may be considered a "living document," in that the results from relevant research and field studies should provide opportunities to test and possibly revise the ToC. Thus, several future directions for both research and practice are available.

Future Research: Understudied Constructs

Despite their established conceptual or empirical links to workplace behavior, Martin-Raugh et al. (2020) noted research gaps involving the malleability of several noncognitive constructs. For example, there appears to be little to no research involving the effortful malleability of personality constructs such as the Big Five and the Dark Triad. Some research has examined these constructs' longitudinal trajectories in the context of naturalistic changes (e.g., Roberts et al., 2006) or clinical intervention (e.g., Roberts et al., 2017); we have discussed their inappropriateness for informing academic or workplace interventions. These degree to which these constructs may be improved through workplace-relevant interventions is amenable to empirical investigation. Consequently, research disseminations may be considered another example of an impact outcome.

Translating the NCI ToC into ToAs

A primary distinction between ToCs and ToAs—or program theories and implementation theories, respectively (Weiss, 1997) —involves their level of specificity. The NCI ToC's description of general mechanisms of noncognitive construct change may be used to inform ToAs for individual intervention programs in academic and workplace settings. In turn, the focus of ToAs shifts to implementation concerns such as construct measurement, training program delivery methods, intervention duration and intensity (i.e., "dosage"), and moderators of training efficacy. Intervention setting represents another central topic: Although these strategies are intended to improve workplace-relevant noncognitive constructs, the interventions themselves may be conducted in academic or workplace contexts. We briefly discussed strategies for quantifying change (i.e., statistical methodology), though these may also differ according to the construct of interest and the specific intervention. Therefore, these may be elaborated within a ToA. ToAs may also be more appropriate for investigating causal links between intermediate and ultimate impact specific to a particular intervention. Implementation theories may also describe unintended consequences at a more detailed level, thus informing specific approaches for minimizing and addressing such concerns. Overall, results from research involving these ToAs may have a reciprocal effect in informing potential modifications to NCI ToC.

Concluding Comments

As employers and researchers increasingly recognize noncognitive constructs' value in predicting workplace performance, efforts to improve these constructs become more significant. More broadly, these efforts signal a shift from simple predictive validity research toward intervention efficacy studies. In turn, the practical value of specific workplace-relevant noncognitive constructs may be situated on a spectrum from selection applications (less malleable) to training and development settings (more malleable). In either case, a premium is placed on the collection of longitudinal data. Within noncognitive constructs, this research should allow stakeholders to differentiate between less effective intervention strategies and more successful programs. Although some unintended NCI outcomes are of obvious concern, they are largely speculative, and methods for minimizing and detecting these outcomes are provided. Furthermore, the intended consequences appear to be far more numerous, including the positive development of various noncognitive constructs and improvement in occupational performance. In turn, NCIs should have a positive impact not only on individuals' psychological attributes and socioeconomic mobility, but also for the advancement of science and for society as a whole (Aguinis & Kraiger, 2009).

References

Aguinis, H., & Kraiger, K. (2009). Benefits of training and development for individuals and teams, organizations, and society. *Annual Review of Psychology*, *60*, 451–474. https://doi.org/10.1146/annurev.psych.60.110707.163505

Alliger, G. M., Tannenbaum, S. I., Bennett, W., Jr., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations of training criteria. *Personnel Psychology*, *50*(2), 341–358. https://doi.org/10.1111/j.1744-6570.1997.tb00911.x

Argyris, C., & Schon, D. A. (1974). *Theory in practice: Increasing professional effectiveness*. Jossey-Bass.

Arthur, W., Jr., Bennett, W., Jr., Edens, P. S., & Bell, S. T. (2003). Effectiveness of training in organizations: A meta-analysis of design and evaluation features. *Journal of Applied Psychology*, *88*, 234–245. https://doi.org/10.1037/0021-9010.88.2.234

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*, 1–26. https://doi.org/10.1111/j.1744-6570.1991.tb00688.x

Barrick, M. R., Stewart, G. L., & Piotrowski, M. (2002). Personality and job performance: Test of the mediating effects of motivation among sales representatives. *Journal of Applied Psychology*, *87*, 43–51. https://doi.org/10.1037/0021-9010.87.1.43

Barth, J., & Lannen, P. (2011). Efficacy of communication skills training courses in oncology: A systematic review and meta-analysis. *Annals of Oncology*, *22*(5), 1030–1040. https://doi.org/10.1093/annonc/mdq441

Bates, R. (2004). A critical analysis of evaluation practice: The Kirkpatrick model and the principle of beneficence. *Evaluation and Program Planning*, *27*(3), 341–347. https://doi.org/10.1016/j.evalprogplan.2004.04.011

Beer, V. (1990). Surveymania in HRD program evaluation. *Performance Improvement Quarterly*, *3*(3), 2–11. https://doi.org/10.1111/j.1937-8327.1990.tb00463.x

Bennett, R. E. (2010). Cognitively based assessment of, for, and as learning (CBAL): A preliminary theory of action for summative and formative assessment. *Measurement*, *8*(2–3), 70–91. https://doi.org/10.1080/15366367.2010.508686

Bezrukova, K., Spell, C. S., Perry, J. L., & Jehn, K. A. (2016). A meta-analytical integration of over 40 years of research on diversity training evaluation. *Psychological Bulletin*, *142*(11), 1227–1274. https://doi.org/10.1037/bul0000067

Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment*, *14*(4), 317–335. https://doi.org/10.1111/j.1468-2389.2006.00354.x

Bosco, F. A., Aguinis, H., Singh, K., Field, J. G., & Pierce, C. A. (2015). Correlational effect size benchmarks. *Journal of Applied Psychology*, *100*(2), 431–449. https://doi.org/10.1037/a0038047

Bowen, D. J., & Neill, J. T. (2013). A meta-analysis of adventure therapy outcomes and moderators. *Open Psychology Journal*, *6*, 28–53. https://doi.org/10.2174/1874350120130802001

Boyatzis, R. (2008). Competencies in the 21st century. *Journal of Management Development*, *27*(1), 5–12. https://doi.org/10.1108/02621710810840730

Brinkerhoff, R. O., & Dressler, D. (2015). Using evaluation to build organizational performance and learning capability: A strategy and a method. *Performance Improvement*, *54*(7), 37–44. https://doi.org/10.1002/pfi.21497

Bureau of Labor Statistics. (2018). Employment projections. https://www.bls.gov/emp/tables.htm

Capelli, P. (2012). *Why good people can't get jobs: The skills gap and what companies can do about it*. Wharton Digital Press.

Cascio, W. F., & Boudreau, J. W. (2008). *Investing in people: Financial impact of human resource initiatives*. FT Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Cole, D. A., & Maxwell, S. E. (2003). Testing mediation models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology*, *112*(4), 558–577. https://doi.org/10.1037/0021-843X.112.4.558

Connell, J. P., & Klem, A. M. (2000). You *can* get there from here: Using a theory of change approach to plan urban education reform. *Journal of Educational and Psychological Consultation*, *11*, 93–120.

De Silva, M. J., Breuer, E., Lee, L., Asher, L., Chowdhary, N., Lund, C., & Patel, V. (2014). Theory of change: A theory-driven approach to enhance the Medical Research Council's framework for complex interventions. *Trials*, *15*, 267. https://doi.org/10.1186/1745-6215-15-267

Delise, L. A., Gorman, C. A., Brooks, A. M., Rentsch, J. R., & Steele-Johnson, D. (2010). The effects of team training on team outcomes: A meta-analysis. *Performance Improvement Quarterly*, *22*, 53–80. https://doi.org/10.1002/piq.20068

D'Silva, K., Duggan, C., & McCarthy, L. (2004). Does treatment really make psychopaths worse? A review of the evidence. *Journal of Personality Disorders*, *18*(2), 163–177. https://doi.org/10.1521/pedi.18.2.163.32775

Duckworth, A. L., & Yeager, D. S. (2015). Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, *44*(4), 237–251. https://doi.org/10.3102/0013189X15584327

Geisinger, K. F. (2010). Consequences and validity: An uneasy relationship. *NCME Newsletter*, *18*, 7–8.

George, R. (2019). What is a theory of action? TetraTech International Development. https://intdev.tetratechasiapacific.com/what-is-a-theory-of-action/

Giangreco, A., Carugati, A., & Sebastiano, A. (2010). Are we doing the right thing? Food for thought on training evaluation and its context. *Personnel Review*, *39*(2), 162 – 177. https://doi.org/10.1108/00483481011017390

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, *102*, 74 – 78. https://doi.org/10.1016/j.paid.2016.06.069

Green, M., & McGill, E. (2011). *State of the industry*. American Society for Training and Development.

Heckman, J. J. (2000). Policies to foster human capital. *Research in Economics*, *54*(1), 3 – 56. https://doi.org/10.1006/reec.1999.0225

Heckman, J. J., & Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics*, *19*(4), 451 – 464. https://doi.org/10.1016/j.labeco.2012.05.014

Holton, E. F., III, & Naquin, S. S. (2004). New metrics for employee development. *Performance Improvement Quarterly*, *17*(1), 56 – 80. https://doi.org/10.1111/j.1937-8327.2004.tb00302.x

Jackson, E. T. (2013). Interrogating the theory of change: Evaluating impact investing where it matters most. *Journal of Sustainable Finance & Investment*, *3*(2), 95 – 110. https://doi.org/10.1080/20430795.2013.776257

Judge, T. A., & Bono, J. E. (2001). Relationship of core self-evaluations traits – Self-esteem, generalized self-efficacy, locus of control, and emotional stability – With job satisfaction and job performance: A meta-analysis. *Journal of Applied Psychology*, *86*(1), 80 – 92. https://doi.org/10.1037/0021-9010.86.1.80

Kautz, T., Heckman, J. J., Diris, R., Ter Weel, B., & Borghans, L. (2014). *Fostering and measuring skills: Improving cognitive and noncognitive skills to promote lifetime success* (NBER Working Paper No. 20749). National Bureau of Economic Research. https://doi.org/10.3386/w20749

Kirkpatrick, D. L. (1996). Great ideas revisited: Techniques for evaluating training programs – Revisiting Kirkpatrick's four level model. *Training and Development*, *50*, 54 – 59.

Klein, C., DiazGranados, D., Salas, E., Le, H., Burke, C. S., Lyons, R., & Goodwin, G. F. (2009). Does team building work? *Small Group Research*, *40*(2), 181 – 222. https://doi.org/10.1177/1046496408328821

Klieger, D., Ezzo, C., Bochenek, J., & Cline, F. (2015, April). *The predictive validity of non-cognitive skills for graduate and professional student success: Some initial findings* [Paper presentation]. Validity and Fairness Issues in Assessing Higher Education Students Symposium, Annual Meeting of the American Educational Research Association, Chicago, IL, United States.

Kline, R. B. (2013). *Beyond significance testing: Statistics reform in the behavioral sciences* (2nd ed.). American Psychological Association. https://doi.org/10.1037/14136-000

Kyllonen, P. C. (2012, May). *Measurement of 21st century skills within the Common Core State Standards* [Paper presentation]. The Invitational Research Symposium on Technology Enhanced Assessments, Princeton, NJ, United States.

Lacerenza, C. N., Reyes, D. L., Marlow, S. L., Joseph, D. L., & Salas, E. (2017). Leadership training design, delivery, and implementation: A meta-analysis. *Journal of Applied Psychology*, *102*(12), 1686 – 1718. https://doi.org/10.1037/apl0000241

Leusner, D., & Lyon, C. (2008). *Providing a research basis for ETS products* (ETS Research Memorandum RM-08-01). ETS.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment. *American Psychologist*, *48*(12), 1181 – 1209.

Malen, B., Croninger, R., Muncey, D., & Redmond-Jones, D. (2002). Reconstituting schools: "Testing" the "theory of change". *Educational Evaluation and Policy Analysis*, *24*, 113 – 132. https://doi.org/10.3102/01623737024002113

Martin-Raugh, M. P., Williams, K. M., & Lentini, J. E. (2020). *The malleability of workplace-relevant noncognitive constructs: Empirical evidence from 39 meta-analyses and reviews* (Research Report No. RR-20-23). ETS. https://doi.org/10.1002/ets2.12306

McCrae, R. R., & Costa, P. T., Jr. (2008). Empirical and theoretical status of the Five-Factor Model of personality traits. In G. Boyle, G. Matthews, & D. Saklofske (Eds.), *Sage handbook of personality theory and assessment* (Vol. *1*, pp. 273 – 294). Sage Publications. https://doi.org/10.4135/9781849200462.n13

McFarland, J., Hussar, B., de Brey, C., Snyder, T., Wang, X., Wilkinson-Flicker, S., Gebrekristos, S., Zhang, J., Rathbun, A., Barmer, A., Bullock Mann, F., & Hinz, S. (2017). *The condition of education 2017* (NCES 2017-144). National Center for Education Statistics. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2017144

McLaughlin, J. A., & Jordan, G. B. (1999). Logic models: A tool for telling your program's performance story. *Evaluation and Program Planning*, *22*(1), 65 – 72. https://doi.org/10.1016/S0149-7189(98)00042-1

Merton, R. K. (1968). The Matthew effect in science. *Science*, *159*(3810), 56 – 63.

Msila, V., & Setlhako, A. (2013). Evaluation of programs: Reading Carol H. Weiss. *Universal Journal of Education Research*, *1*(4), 323 – 327. https://doi.org/10.13189/ujer.2013.010408

National Association of Colleges and Employers. (2018). Job outlook 2019.

Paterson, T. A., Harms, P. D., Steel, P., & Credé, M. (2015). An assessment of the magnitude of effect sizes: Evidence from 30 years of meta-analysis in management. *Journal of Leadership & Organizational Studies*, *23*(1), 66 – 81. https://doi.org/10.1177/1548051815614321

Paulhus, D. L., & Williams, K. M. (2002). The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, *36*(6), 556 – 563. https://doi.org/10.1016/S0092-6566(02)00505-6

Ployhart, R. E., & Vandenberg, R. J. (2010). Longitudinal research: The theory, design, and analysis of change. *Journal of Management*, *36*(1), 94–120. https://doi.org/10.1177/0149206309352110

Rice, M. E., Harris, G. T., & Cormier, C. A. (1992). An evaluation of maximum-security therapeutic community for psychopaths and other mentally disordered offenders. *Law and Human Behavior*, *16*(4), 399–412. https://doi.org/10.1007/BF02352266

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, *2*(4), 313–345. https://doi.org/10.1111/j.1745-6916.2007.00047.x

Roberts, B. W., Luo, J., Briley, D. A., Chow, P. I., Su, R., & Hill, P. L. (2017). A systematic review of personality trait change through intervention. *Psychological Bulletin*, *143*(2), 117–141.

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin*, *132*(1), 1–25. https://doi.org/10.1037/0033-2909.132.1.1

Salas, E., Burke, S., Bowers, C. A., & Wilson, K. A. (2001). Team training in the skies: Does crew resource management (CRM) training work? *Human Factors*, *43*(4), 641–674. https://doi.org/10.1518/001872001775870386

Salas, E., DiazGranados, D., Weaver, S. J., & King, H. (2008). Does team training work? Principles for health care. *Academic Emergency Medicine*, *15*(11), 1002–1009. https://doi.org/10.1111/j.1553-2712.2008.00254.x

Schanzenbach, D. W., Nunn, R., Bauer, L., Mumford, M., & Breitweiser, A. (2016). *Seven facts on noncognitive skills from education to the labor market*. Brookings.

Schlaefli, A., Rest, J. R., & Thoma, S. J. (1985). Does moral education improve moral judgment? A meta-analysis of intervention studies using the Defining Issues Test. *Review of Educational Research*, *55*(3), 319–352. https://doi.org/10.3102/00346543055003319

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*(2), 262–274. https://doi.org/10.1037/0033-2909.124.2.262

Spector, P. E., Fox, S., & Domagalski, T. (2006). Emotions, violence, and counterproductive work behavior. In E. K. Kelloway, J. Barling, & J. J. Hurrell, Jr. (Eds.), *Handbook of workplace violence* (pp. 29–46). Sage. https://doi.org/10.4135/9781412976947.n3

Teding van Berkhout, E., & Malouff, J. M. (2016). The efficacy of empathy training: A meta-analysis of randomized controlled trials. *Journal of Counseling Psychology*, *63*(1), 32–41. https://doi.org/10.1037/cou0000093

Van Iddekinge, C. H., Roth, P. L., Putka, D. J., & Lanivich, S. E. (2011). Are you interested? A meta-analysis of relations between vocational interests and employee performance and turnover. *Journal of Applied Psychology*, *96*(6), 1167–1194. https://doi.org/10.1037/a0024343

Van Iddekinge, C. H., Roth, P. L., Raymark, P. H., & Odle-Dusseau, H. N. (2012). The criterion-related validity of integrity tests: An updated meta-analysis. *Journal of Applied Psychology*, *97*(3), 499–530. https://doi.org/10.1037/a0021196

Vanhove, A. J., Herian, M. N., Perez, A. L. U., Harms, P. D., & Lester, P. B. (2016). Can resilience be developed at work? A meta-analytic review of resilience-building programme effectiveness. *Journal of Occupational and Organizational Psychology*, *89*(2), 278–307. https://doi.org/10.1111/joop.12123

Weiss, C. H. (1972). *Evaluation research: Methods of assessing program effectiveness*. Prentice-Hall.

Weiss, C. H. (1997). Theory-based evaluation: Past, present, and future. *New Directions for Evaluation*, *76*, 68–81. https://doi.org/10.1002/ev.1086

Wholey, J. S. (1979). *Evaluation: Promise and performance*. Urban Institute.

Suggested citation:

Williams, K. M., Martin-Raugh, M. P., & Lentini, J. E. (2022). *Improving noncognitive constructs for career readiness and success: A theory of change for postsecondary, workplace, and research applications* (Research Report No. RR-22-04). ETS. https://doi.org/10.1002/ets2.12346

Action Editor: John Sabatini and Sooyeon Kim

Reviewers: Caroline Wylie and Jonathan Schmidgall