Chapter 10 - Improving the Design and Use of Meta-Analyses of Career Interventions

Betsy Jane Becker

Florida State University

In this chapter I examine research syntheses and meta-analyses in the area of career interventions and propose a few ideas for improving on the excellent work done to date. I begin with a brief overview of meta-analysis and systematic reviews and consider the kinds of questions that have been asked about career interventions. I also briefly discuss the role of theories in meta-analysis. Beginning with a theoretical model enables the reviewer to assess not only what is well studied, but also what issues or effects are in need of further research. I then introduce a framework for exploring the generalizability of results of meta-analyses. The approach is also useful for planning of future studies. Within each section I draw on meta-analyses in the area of career interventions to illustrate my points.

What is Meta-Analysis?

The term meta-analysis was coined by Glass to mean "the analysis of analyses…. a rigorous alternative to … casual, narrative discussions of research studies" (1976, p. 3). Meta-analyses draw together series of studies on a well-defined topic, and use quantitative techniques to analyze and understand the diversity of effects found in those studies. Meta-analyses and more generally systematic reviews (Petticrew & Roberts, 2006) involve versions of the same steps we follow in primary research. Systematic reviews can inform policy and practice alike. Because they are by definition less haphazard than traditional narrative reviews, they also have the potential to be more thorough and less biased.

Similarities between the processes and steps in primary research and those required in meta-analyses have been well described by Cooper (1982) and others (e.g., Cook, Sackett, & Spitzer, 1995). Cooper (1982) laid out five, then later seven steps (e.g., Cooper, 2016) in this process. They are as follows:

1. problem formulation,

2. literature search,

3. extracting information from studies (coding),

4. evaluating quality of the studies,

5. analyzing study results,

6. interpreting the evidence, and

7. public presentation.

At each step the goal is to be thorough, systematic, and clear about the process so that other reviewers can follow the same steps and arrive at essentially the same

point when their review is done. In short, applying replicable procedures in the first four steps should enable the reviewer to collect studies and to characterize the features of studies thought to relate to the outcomes of interest. If the reviewer uses quantitative analyses at Step 5, then the review is often called a meta-analysis or quantitative synthesis.

Problem Formulation: A Focus on Variation and Process

What Works? When, Where, and for Whom?

Many meta-analyses of intervention literatures begin with a very practical goal: to find out "what works." Often this is not even stated. Oliver and Spokane (1981, 1983; Spokane & Oliver, 1983) as well as Baker and Popowicz (1983) were among the earliest advocates of meta-analysis in the area of career counseling, though they and others (e.g., Fretz, 1981) discuss prior reviews of a narrative nature. Spokane and Oliver (1983) did not lay out specific research questions, simply stating that they intended to "examine the outcomes of … vocational treatments and draw some conclusions that will be of help to researchers and practitioners" (1983, p. 100). They reported a very large treatment effect of 0.85 standard deviation units.

Obvious additional questions arise, such as for whom do interventions work, what features are associated with successful interventions, and the like. Spokane and Oliver found differences between group, individual, and other treatments but did not study other moderators. Later Oliver and Spokane (1988) noted that intensity and treatment type were confounded in their studies, raising questions about the simplicity of their earlier finding. Spokane and Oliver (1983) and also Oliver and Spokane (1983) strongly argued for using more rigorous methods in future studies and gave a methodological checklist for authors to follow. They also conveyed the importance of complete reporting in primary studies, noting the problems that occur when extracting information for a review (Step 3 above). These early authors faced incompleteness and errors when trying to find information not only on intervention study outcomes, but about the nature of the career interventions, the control conditions, and their participants. Recent reviews continue to complain about poor reporting (Liu, Huang, & Wang, 2014).

Putting the questions reviewers have asked into a statistical framework, we find that these meta-analysts looked for *main effects* of treatments and possible *interactions* with moderators. If the treatment effect interacts with study itself, the results do not all agree, and we say they are not homogeneous. It has become common to test the null hypothesis of homogeneity, both across all studies and within subsets of studies. However, it is less common to estimate and interpret the degree of variability present in the study collection. Doing so can help us understand how widely the "typical" results can be generalized.

Identifying and testing moderators of the size of the intervention effect is equivalent to asking whether the size of the effect depends on (interacts with) those moderators—the article characteristics, sample characteristics, or treatment features of most interest. Fretz (1981) presaged this idea by calling for career researchers to focus on (client) attribute-treatment interactions in career-intervention research (drawing on Cronbach and Snow's 1977 work on aptitude treatment interactions in instruction). Likewise Evans and Burck's (1992) simply stated questions reflect this progression: "What statistical statement can be made about the overall effect size produced by the career education interventions of the 67 studies? What is the relationship between study characteristics and study effect size results?" (p. 65). Over time, series of career studies and meta-analyses of career-intervention studies have addressed several iterations of these basic questions about effect magnitude and what it depends on, with meta-analysts adding to the knowledge base by focusing on different features of the studies at hand.

Moderators also provide potential explanations of variation in effects across studies. This is reflected to a degree in statements about how much variance is explained by certain predictors. For instance, Brown and Ryan Krane (2000) stated, "specific intervention components … accounted for between 2% and 38% unique variance in effect sizes" (p. 744). However, we usually do not know the amount of between studies variance that was explained in the metric of the effect size.

Tools for Answering How and Why Questions

As useful as it is to know about treatment main effects, it is equally important to learn about how and why career interventions work as they do. Questions may concern the components of and processes behind career interventions (e.g., Whiston, Brecheisen, & Stephens, 2003) and the broader contexts of associated variables that may impact intervention effectiveness, such as age, race/ethnicity, or participant/intervention "fit" (Nichols, 2009). Such inquiries become more salient once interventions have been found to "work" on average. Investigations of this kind are more unusual in career-intervention meta-analyses, but they are possible. I briefly describe two approaches that may be useful to the career field.

Response-surface models. Certainly investigations of program components appear in the career field, such as in meta-analyses by Whiston, Sexton, and Lasoff (1998), Ryan (1999), Brown and Ryan Krane (2000), and Brown et al. (2003). Brown and Ryan Krane listed five "critical ingredients" and found that the number of those critical elements used in an intervention related to the size of the treatment effect. These ingredients have been investigated in a variety of ways since then.

How can we improve the analyses of program components done to date? Rubin (1992) argued that meta-analytic "response-surface" (i.e., regression) models can be used to predict the effect that an optimal study would have produced. For example, we could predict the effect size for a study of a group-counseling intervention with all five critical components identified by Ryan (1999) and with other features such as a high level of intensity, college-aged or adult participants, and so forth. Other possible projections can be obtained for comparison.

While valuable, even a response-surface model that predicts the results of a high quality "five-ingredient intervention" cannot capture all of the potential benefits, or problems, of a real intervention. Thus, findings like Ryan's list of ingredients or other features included in a response-surface model that is based on the between studies evidence in a meta-analysis can lead to designs for new studies. Those new studies can be organized so features vary within study for stronger evidence on each

intervention component. One might imagine a five-way factorial design in which participants either receive or do not receive each of the five critical ingredients named by Ryan (1999). Additional important factors such as demographics could be used as blocking variables. As Matt and Cook (2009) point out, between studies evidence from meta-analyses is always subject to the threat of confounded moderators, and is weaker than within-study evidence.

Path models. A second modeling approach to how and why questions is exemplified by Sheu and colleagues' (2010) work in developing meta-analytic models of the roles of various psychological antecedents of career choices. Tracey and Rounds (1993) applied similar ideas to studies of Holland and Gati's model. Becker (1992, 1995, 2009) and later others (Cheung & Chan, 2005; Viswesvaran & Ones, 1995) have written about the benefits of incorporating theoretically based models into research synthesis. A key benefit is that laying out possible path models forces the meta-analyst to check whether all paths have been studied. Missing correlations between model components suggest relations needing further study.

Such phenomena as counselor-client rapport and the working alliance (Whiston & Rose, 2015) and process variables such as expectations for counseling (Heppner & Heppner, 2003) are potential constructs to be investigated in such a meta-analytic model. Also though few meta-analysts have done so, one can include treatment variables as components in path models. As with Rubin's response-surface approach, this kind of meta-analysis requires prior knowledge, or prior theory, about the factors important to the outcome of interest.

A Framework for Generalization (MUTOS)

A potential weakness of the typical meta-analysis process is that studies within any research domain seem to be created more or less through happenstance. This is not true for any particular researcher's collection of studies, which typically represent a program or stream of related work, but taken as a whole, primary research domains and thus meta-analyses in the social sciences are less structured than in some other fields. Registries of primary studies are rare in social science, so the would-be reviewer often does not know what studies exist until

collecting data for a review. Two exceptions to this include the U.S. Institute for Education Sciences What Works Clearinghouse registry of trials in education located at http://ies.ed.gov/ncee/wwc/references/registries/RCTSearch/RCTSearch.aspx and Baldwin and Del Re's (2016) creation of an open access data base of effect sizes in the area of psychotherapy (e.g., http://shinyserver.byu.edu/family_therapy/).

Because of the lack of structure and planning in various research domains, a meta-analyst may define a problem of interest only to find that few studies have examined it. Alternatively, because existing studies have no overarching organizational scheme, the meta-analyst might find great diversity (yet also often duplication) in methods, populations, measures, and so on. In the career-intervention domain, primary researchers may have used commonly accepted career development measures, such as the Self-Directed Search and Strong Campbell Interest Inventory, researcher developed and administered questionnaires (e.g., Shevlin & Millar, 2006), or interviews like those used by Thoresen, Hosford, and Krumboltz (1970) to elicit self-reports of information seeking behavior. Participants may be young or old, female or male, or of any socioeconomic background or ethnicity, but these features of participants may not be well described, and analyses may not have explored their import for effectiveness. I next describe an approach that can help identify and understand the impacts of this diversity in research literatures.

An organizing framework can help us to take stock of what is known. I put forward a framework useful for problem formulation, assessment of existing literatures, and planning future studies of career interventions. It is drawn from the work of Cronbach (1982) in the area of program evaluation and has been applied to meta-analysis by Becker (1996) and Matt and Cook (2009). Cronbach's UTOS framework refers to populations of units (Us), treatments (Ts), observing operations (Os), and settings (Ss). These labels refer to possible features of studies. Aloe suggested adding methods (Ms) to Cronbach's scheme to capture the role of the diversity of approaches to study design that appear in most meta-analyses; thus, we have MUTOS (Becker & Aloe, 2016).

In Cronbach's original approach, populations of Us (individuals or groups of participants), Ts, and Os are the target of inferences. In a particular setting a researcher samples instances from those populations of Us, Ts, and Os in order to generalize to the sampled domain UTOS. S is not sampled because Cronbach viewed S as fixed for a given study. Assuming a study is conducted without serious compromises, assessing whether results are generalizable involves assessing whether and how well that study's results can be extrapolated to what Cronbach calls the "domain of application," or *UTOS. *UTOS can refer to any number of domains—from all instances of interest (e.g., all career interventions), to a narrower domain such as an intervention in a single university career center using a particular measure of a given outcome. *UTOS is not directly observed.

MUTOS in Meta-Analysis

Using MUTOS in the context of meta-analysis can help make the domains of interest explicit, a key aspect of problem formulation and data collection. As noted above, meta-analysts typically try to identify moderators that lead to differences in study results; these may be characteristics of Us, Ts, Os, and Ss. Becker (1996) and Shadish, Cook, and Campbell (2002) argue that analysis of moderators in meta-analysis is akin to evaluating the importance of the specific Us, Ts, Os and Ss that appear across studies.

Because study designs and methods also typically vary in a meta-analysis, the M component represents the range of possible methods beyond what is reflected in Cronbach's O component, which typically includes only diversity in the measures used. If studies use different sampling strategies, analyses, different degrees of attrition, or ways to assign participants to conditions (in treatment-control studies), they have different Ms. By specifying the populations in MUTOS, the meta-analyst outlines inclusion rules for the synthesis. Also, MUTOS suggests what aspects of studies should be evaluated and determined to be relevant (or irrelevant) to the effect of interest, here, to the effects of career interventions.

The use of this organizational framework in meta-analyses differs in one important way from Cronbach's original application to primary studies. Cronbach argued that one should sample the Us, Ts and Os with the target of generalizability (*UTOS) in mind. However, most meta-analyses involve purposive and, ideally, exhaustive collections of studies on a

topic—collections not generated through an explicit sampling design on MUTOS. Acknowledging that random samples are rare in the meta-analysis context, Matt and Cook (2009) argue that inferences in meta-analysis depend on different warrants, not on sampling theory alone. Their work describes in detail those other warrants for generalization.

Campbell's (1969) multiple operationism, the process by which a theoretical construct such as a treatment or outcome is exemplified in multiple ways, is also linked to Cronbach's (1982) idea of sampling from populations of Ts or Os. Each realization of a theoretical entity (here, each M, U, T, O, or S) can differ in terms of theoretically irrelevant characteristics. The incorporation of such "heterogeneous irrelevancies" (Cook, 1991) in the operationalizations of constructs is a key benefit of multiple operationism. Because all Ms, Ts, Os, and so forth are imperfect representations of the constructs they stand for, using multiple instantiations of each construct builds stronger understandings of the phenomena we are studying and increases the potential for generalizability. Brown (2017) has called for future work to identify core constructs, arguing that many measures used in the field may actually be measuring the same thing, despite having different names. The most direct way to accomplish Brown's goal would be to study the array of proposed operations together in a new primary study, as he has done in several cases (Brown, 2015). However, similarities between the effects based on different operationalizations found in a meta-analysis can also suggest communalities in operations not yet examined together in any single study. Whether the potential for generalization across operations is realized in any meta-analysis depends on whether findings appear similar across the many versions of MUTOS.

By including a diversity of Us, Ts, and ways of measuring each construct (different Os), the meta-analyst can better gauge what is common and fundamental to the effect of interest and what is not. Meta-analysis facilitates exploring the impact of multiple operations through examination of a broader range of operations than would appear in any single study (e.g., treatments of different lengths and intensities, measures with different numbers and kinds of items). This fact argues against the idea of total uniformity in the measures used in a field: What if we select one measure, and it somehow does not tap a critical feature of the construct we want to

examine? However, having different operationalizations within and across studies can also add variability and uncertainty to the results.

Using MUTOS to Evaluate Generalizability in Meta-Analysis

How can we use MUTOS with meta-analytic data to assess generalizability? Becker (1996) proposed an analogue to generalizability theory (Cronbach, Gleser, Nanda, & Rajaratnam, 1972) for meta-analysis that can be used to estimate and model sources of error. Both generalizability and decision studies are possible. Becker and Aloe (2016) expanded on those ideas to propose the following steps for a simple generalizability analysis:

1. Specify desired target of inference by defining MUTOS.

2. Classify study features using MUTOS.

3. Evaluate diversity in Ms, Us, Ts, Os, and Ss.

4. Assess overall heterogeneity of effects.

5. Evaluate empirical variation in results using MUTOS.

6. Assess connections to desired domain of application, *UTOS.

Specify desired target of inference by defining MUTOS. Ideally the meta-analyst begins by using theory and the nature of the problems that guide the review to set out the desired target population, UTOS. Describing the ideal participant populations, types of interventions, and outcomes of interest gives form to inclusion rules for data collection. M is not included here because one makes inferences to the real world, not to a collection of future studies using particular methods. However, rules about what study designs are acceptable would also be set. This step is not illustrated in the example below; it would be completed before the meta-analysis is begun.

The use of MUTOS at the problem-formulation stage connects well to the ideas described by Whiston (2017), who argues that the field of career-interventions would benefit from more "setting-specific" meta-analyses.

Whiston argues that broadly defined meta-analyses allow reviewers to gloss over important category distinctions, an argument made early in the history of meta-analysis by Presby (1978). As Whiston notes, the inclusion of varied collections of diverse contexts and populations requires coding of critical moderators—a task that can be challenging even when coders have strong knowledge of the relevant content.

Classify study features using MUTOS. After studies are collected, part of the data extraction involves coding the levels of MUTOS observed across studies. The key difference from a typical meta-analysis is conceptual—that is, we now explicitly note whether the literature contains information on the range of possible Us, Ts, and so forth, which were outlined earlier. Finding that researchers have not examined certain populations or specific outcomes provides a first opportunity to describe "what is not known."

Evaluate diversity in Ms, Us, Ts, Os, and Ss. Evaluating the numbers and kinds of instantiations of M, U, T, O, and S in the collection of studies is next. We qualitatively assess how much those instantiations differ, and how well they represent conditions of interest for generalization. Are the participants all upper-level undergraduate students, or are more general samples in the mix? Do the measures used appear to be very similar? Are most or all well-known measures included? These assessments, based on professional judgments, provide information on the potential for dispersion in study results. More diversity in Us, Ts, Os, and Ss means more possible heterogeneity in results, if those features matter. It also means greater potential to generalize widely (e.g., across many different kinds of units, treatments, etc.).

Assess overall heterogeneity of effects. The importance of MUTOS components is evaluated in comparison to the degree of variability in the full collection of study results. The meta-analyst can test homogeneity of effects across all studies, and provide indices of dispersion such as I^2 (Higgins & Thompson, 2002), Birge's ratio (Birge, 1932), and between studies variance estimates. These overall variance estimates are used with estimated means to compute plausible-values intervals and to graph hypothetical population distributions of effects.

Evaluate empirical variation in results using MUTOS. Against a background of how much the effect sizes vary across studies, we then analyze the study features associated with the MUTOS components to reveal whether each feature relates statistically to the observed study outcomes. This step may sound identical to the moderator analyses used in many career-intervention meta-analyses, but it involves an added step: For each factor of interest, we will estimate the degree of variation that is present, in the scale of the effect size. This, along with tests of significance, plausible-values intervals, and graphic displays, tells us the extent to which each component limits or allows generalizability. We may also assess the joint impacts of all features associated with each MUTOS component.

Assess connections to desired domain of application, *UTOS. Finally, with this knowledge in hand, the meta-analyst (or a reader) can make connections to more particular domains of application, *UTOS. This entails consideration of the contexts to which one wants to generalize. If you are working in an industrial re-training setting and all available information is based on college students, it may be hard to safely generalize the observed findings to your situation. If a widely diverse set of scales has been used in the studies at hand, but they all agree in the results found, it is easy to generalize to situations where any one of those scales has been applied. This work of selecting *UTOS domains would be done after all analyses were complete. I do not illustrate this step in detail below.

Evaluating Variation in Measures for Decision Making Self-Efficacy

The steps above are described in more detail in work by Becker and Aloe (2016); here I demonstrate a hypothetical application and limited analysis using a subset of data from Ryan's (1999) meta-analysis of career interventions. For this example I selected eight standardized-mean-difference effect sizes for the outcome of decision making self-efficacy, shown in Table 1. The overall results are described briefly to provide a context for the use of MUTOS. In practice the overall assessment of heterogeneity would follow the initial three steps of the process outlined above.

Table 1: Studies of career decision making self-efficacy from Ryan (1999)

Authors	Year	n^T	n^C	g	Scale
Garis & Harris-Bowlsbey	1984	47	15	1.05	SACP
Healy & Mourton	1984	74	99	-0.48	SECMC
Fukuyama et al.	1988	39	38	0.53	CDMSES
Tempestini, Horan & Good	1988	55	11	0.08	SACP
Garis & Niles	1990	87	25	0.40	SACP
Niles & Garis	1990	51	17	0.37	SACP
Luzzo & Taylor	1993	44	44	0.66	CDMSES
Luzzo, Funk & Strang	1996	30	30	0.20	CDMSES

Note: Measures included the Career Decision Making Self-Efficacy Scale (CDMSES), the Self-Assessment of Confidence and Progress in Educational/Career Planning (SACP), and the Self-Estimated Career Management Competencies scale (SECMC). Sample sizes for treated and control samples are n^T and n^C, and g is the uncorrected standardized mean difference from Ryan (1999).
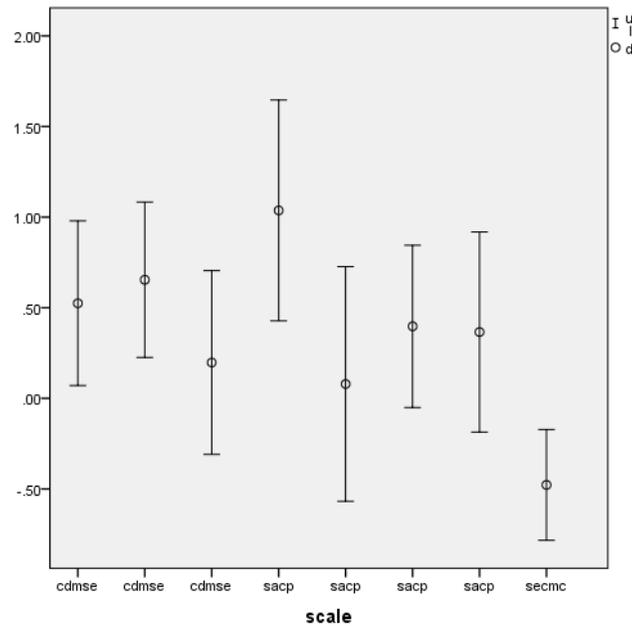

Figure 1. Forest plot of effect sizes from Ryan (1999), sorted by scale used.

Assess Overall Heterogeneity of Effects

Ryan computed the standard test of homogeneity for these studies and found significant between studies variation ($Q(7) = 33.66$, $p < .05$). Even though this test is significant, and the value of I^2 suggests that 79% of variability in these effects reflects true differences, the forest plot of the effects and their 95% confidence limits in Figure 1 shows overlapping intervals, reflecting a good deal of agreement among the effect sizes.

Ryan did not report a measure of between studies variation; for her data, the weighted Dersimonian and Laird (1986) variance is 0.216. This corresponds to a standard deviation for the full population of effects of 0.464—almost half of a standard-deviation unit. If the effects came from a normal distribution of population effects, 95% of those true values would cover a range of 1.81. Centering the distribution on the appropriate mean effect, which for these data is a random-effects mean of 0.33, 95% of all true intervention effects for decision making self-efficacy fall between -0.58 and 1.22. Figure 2 shows this hypothetical distribution in

the upper portion of the plot with a solid line. The vertical lines in the middle of the plot represent the 95% random-effects confidence interval for the mean of all effects; this covers from just below zero (-0.04) to 0.70, a rather wide range of uncertainty about where the mean effect size is actually situated.
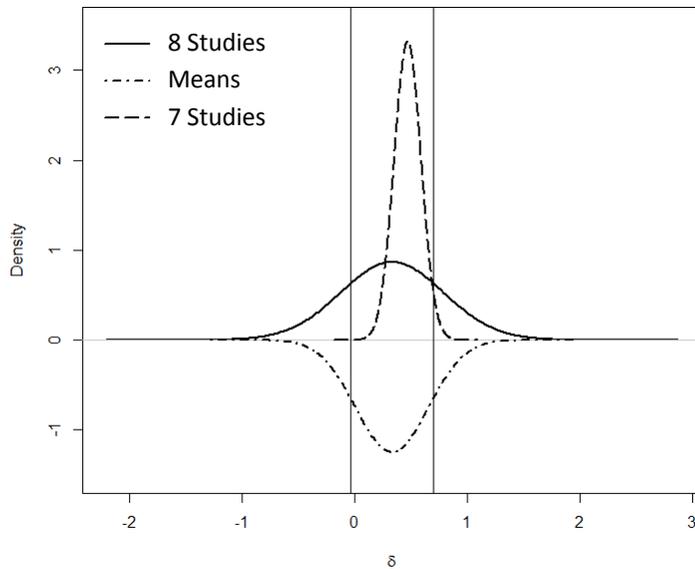

Figure 2. Distributions of population effects (above) and of mean population effects by type of measure used (below). Vertical lines mark the 95% confidence interval for the random effects mean of the eight effects.

Classify Study Features Using MUTOS

To fully apply the MUTOS framework to this data set would require a full reading of all eight articles and consideration of each component. For this demonstration I examine one feature that represents the O component: the decision making self-efficacy measure used.

Evaluate Diversity in Ms, Us, Ts, Os, and Ss

Because I am not familiar with the array of available career decision making self-efficacy measures available, I sought input on whether the set of three measures used in the studies located by Ryan (1999) is diverse. Personal communications from colleagues knowledgeable in the field (Lenz, Osborn, and Sampson) indicated that only the CDMSES was a familiar instrument. Also other versions of included scales exist (e.g., Betz, Klein, & Taylor, 1996), and Lent et al. (2016) have used a newly developed scale tapping this construct. Lastly, only the Career Decision Self-Efficacy

scale appears in the National Career Development Association's *Counselor's Guide to Career Assessment Instruments* (Wood & Hays, 2013).

The included instruments should be inspected for a better evaluation of their diversity, and content experts could judge other aspects such as their reactivity, formats, and so forth. For this example, given that other instruments have been identified, my assessment is that the measures used in the studies in Ryan's synthesis do not cover the full range of possible measures of decision making self-efficacy.

Evaluate Empirical Variation in Results Using MUTOS

In Figure 1 the effect sizes are sorted according to the scale used. It is immediately apparent that the study by Healy and Mourton (1984), the only study which used the Self-Estimated Career Management Competencies scale, shows a rather divergent effect. A chi-square test of differences in means between scales shows that the scale used relates to the size of the effect (Q_B (2) = 26.85, p < .0001). In addition, the residual test (Q_E (5) = 6.82, p > .05) shows no remaining variability; thus, a fixed-effects model is appropriate. The effects based on each distinct instrument can be considered equal.

Rather than focus on the specific group means for the MUTOS analysis of this O feature, we estimate the degree of variation in the means, using techniques similar to those used above. This reveals a somewhat smaller variance estimate, of 0.102, for between groups variability due to measures used. This variance is just under half the size of the full population variance, with a standard deviation of 0.320. The upside-down distribution in Figure 2 shows the spread of the true means by scale type. The range within which 95% of these means fall runs from -0.30 to 0.96, which is still a fairly broad range; thus, the upper and lower distributions based on all eight studies are both fairly wide. The measure used is important to the size of the effects, and it would not be safe to generalize across all measures of career decision making self-efficacy to make our inferences.

The fact that the Healy and Mourton effect stands out, and has a negative value, suggests closer inspection of their measure. Healy and Mourton's brief description of the Self-Estimated Career Management Competencies (SECMC) scale says that respondents rate their "Career Decision Making, World of Work Information, and Knowledge of Preferred Occupation on four-point scales (1 = upper 25 percent, 2 = upper 50 percent, 3 = upper 75 percent, 4 = lower 25 percent" (1984, p. 58). Placing one's self in a percentile range requires a rating relative to some other distribution of individuals, presumably relative to one's peers. However, not only are the categories confusing because they are not mutually exclusive (e.g., Category 1 is included in the ranges of Categories 2 and 3), but it is not clear what group provides the reference distribution. In addition, it seems like it would be extremely difficult to make such judgments. How does one know how well one's peers make decisions about careers (or how much they know about the world of work or their preferred occupations)?

The SECMC scale clearly behaved differently from the other two. When the effect from Healy and Mourton is omitted, the remaining results become homogeneous ($Q(6) = 6.84$, $p = .34$) with a mean of 0.47 ($SE = .12$). Also the between studies variance overall decreases to .01, which means 95% of all population effects are projected to fall between 0.28 and 0.66. This distribution is shown as the narrower distribution in the upper half of Figure 2, graphed with a dashed line. Virtually all effects are expected to be positive and moderate in size. If the Healy and Mourton effect is omitted we can be nearly certain that any intervention that examines career decision making will have a positive impact. Effects based on the Career Decision Making Self-Efficacy Scale and the Self-Estimated Career Management Competencies scale are comparable, and findings can be generalized across those two instruments. In addition, the mean of intervention effects can now be described as moderate in size. Most of the between scales (and between studies) variance was due to one unusual result.

Assess Connections to Desired Domain of Application, *UTOS

The last step is to assess the relevance of the results to a desired domain or situation. In this case, with all studies included, 76% of effects in the population are likely to be positive, but the mean effect is not clearly above zero. Having set aside the unusual scale used by Healy and Mourton, we can easily generalize the findings of this set of seven studies across the other two measures used. The effect of career interventions on decision making self-efficacy is sizeable, at about a half of a standard deviation, and nearly all effects are expected to be positive.

A real application of this approach would also consider a range of other features of the studies. In this data set I also examined the role of publication date and found it did not relate to the size of the effect. Other features would be of more interest, such as the compositions of the samples of participants, the nature of the treatments used, and so on. A complete analysis might find that other features also relate to the size of the treatment effect or are confounded with the scale used.

Conclusions

In this paper I have discussed three ways to incorporate theories and prior expectations about the generalizability of results into a meta-analysis. This is in line with Sampson and colleagues' (2014) call for more attention to the role of theory in building an evidence base for career interventions. Prideaux, Creed, Muller, and Patton (2000) pointed out that some primary studies present theories as putative foundations for their work but do not explicitly link those theories to the interventions they study. Almost half of the studies they reviewed had no clear connection to theory at all. It is possible to build meta-analytic explanations purely empirically, such as with the response-surface modeling approach proposed by Rubin. It uses traditional metaregression methods, but predicts what combinations of study features lead to "optimal" results. Such models can be based on purely empirical findings about coded study moderators. However even this approach benefits from the use of theoretically grounded predictors. Without a basis in theory, we risk building intervention approaches on chance and forego opportunities to improve both theory and practice (Sampson et al., 2014).

A second approach based strongly on theory is meta-analytic path modeling. Approaches based on structural equation modeling in meta-analysis have to

date been largely based on correlational studies, but information about effective treatments can also be incorporated into this framework. Specifying a model before beginning the meta-analysis enables the reviewer to find missing connections, leading to potential future studies. Also, the data from such syntheses reveals which connections are well understood and which have been examined by fewer studies. Meta-analytic path models can also identify mediating variables (e.g., Whiteside & Becker, 2000), which is not possible in syntheses of simple bivariate relations.

The third approach is based on Cronbach's theory of generalizability. Specifying moderators of interest a priori and assessing the diversity of instances for each one is another way to evaluate whether pieces of evidence are missing from the meta-analysis—this is part of assessing what we don't know. The MUTOS approach focuses us on heterogeneity in the effects themselves and in the moderators studied for each MUTOS component. Estimating the amount of variability resulting from differences in these components allows us to make judgments about how widely or narrowly our results can safely be generalized.

In my example using Ryan's (1999) data, a simple, single result could not be safely generalized across the measures of career decision making self-efficacy found by Ryan. In addition, the set of measures did not include all known measures; thus, it was not particularly diverse. A mean effect was estimated across all measures and all studies, but with great uncertainty. Whiston and Rose (2015, p. 55) have called for the standardization of measures used in career studies, but that practice would limit our capacity to assess generalizability across measures. A middle ground might be preferable, where studies include one or two standard measures along with other unique measures chosen by the researcher. Information on the relations among these instruments, along with impact data from the interventions used, will inform us about commonalities among the measures different researchers are using. If all measures are deemed similar and appropriate to retain in a synthesis or in practice, more studies may be needed to get more precise estimates of the mean effect size for a treatment (e.g., Becker, 1996). Our focus on MUTOS and on variance estimation allows us to make that determination.

Let me present a final challenge: Could the field of career-intervention studies consider the creation of a shared database of effect-size information about career interventions, like that proposed by Baldwin and Del Re? Such an effort would allow ongoing assessment of what is known and not known, could be guided and informed by theory, and should provide up to date guidance for practice. Analytic strategies like those described here, along with such a detailed data source of study results, would improve future meta-analyses as well as provide ideas and guidance for design of future studies that target needed information. It is a collaboration worth considering.

References

* Studies included in the example are marked with an asterisk.

Baker, S. B., & Popowicz, C. L. (1983). Meta-analysis as a strategy for evaluating effects of career education interventions. *Vocational Guidance Quarterly, 31*(3), 178–186.

Baldwin, S. A., & Del Re, A. C. (2016). Open access meta-analysis for psychotherapy research. *Journal of Counseling Psychology, 63*(3), 249–260.

Becker, B. J. (1992). Using results from replicated studies to estimate linear models. *Journal of Educational Statistics, 17*(4), 341–362.

Becker, B. J. (1995). "Using results from replicated studies to estimate linear models": Correction. *Journal of Educational and Behavioral Statistics, 20*(1), 100–102.

Becker, B. J. (1996). The generalizability of empirical research results. In C. P. Benbow & D. Lubinski (Eds.), *Intellectual talent: Psychometric and social issues* (pp. 362–383). Baltimore: Johns Hopkins Press.

Becker, B. J. (2009). Model based meta-analysis. In H. M. Cooper, L. V. Hedges & J. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (Vol. 2nd ed., pp. 377–396). New York, NY: Russell Sage Foundation.

Becker, B. J., & Aloe, A. M. (2016). *Generalizing meta-analytic data using cronbach's (M)UTOS framework.* Unpublished manuscript.

Betz, N. E., Klein, K. L., & Taylor, K. M. (1996). Evaluation of a short form of the Career Decision-Making Self-Efficacy scale. *Journal of Career Assessment, 4*, 47–58.

Birge, R. T. (1932). The calculation of errors by the method of least squares. *Reviews of Modern Physics, 40*, 207–227.

Brown, S. D. (2015). On statistical wizardry, construct proliferation, and other challenges for our science. *The Counseling Psychologist, 43*, 614–628.

Brown, S. D. (2017). Meta-analysis and evidence-based career practice: Current status and future directions. In J. P. Sampson, E. Bullock-Yowell, V. C. Dozier, D. S. Osborn, & J. G. Lenz (Eds.), *Integrating theory, research, and practice in vocational psychology: Current status and future directions* (pp. 82). Tallahassee, FL: Florida State University. Retrieved from https://doi.org/10.17125/svp2016.ch8

Brown, S. D., & Ryan Krane, N. E. (2000). Four (or five) sessions and a cloud of dust: Old assumptions and new observations about career counseling. In S. D. Brown and R. W. Lent (Eds.), *Handbook of counseling psychology* (pp. 740–766). New York: Wiley.

Brown, S. D., Ryan Krane, N. E., Brecheisen, J., Castelino, P., Budisin, I., Miller, M., & Edens, L. (2003). Critical ingredients of career choice interventions: More analyses and new hypotheses. *Journal of Vocational Behavior, 62*(3), 411–428.

Campbell, D. T. (1969). Definitional versus multiple operationism. *Et al., 2*, 14–17.

Cheung, M. W.-L., & Chan, W. (2005). Meta-analytic structural equation modeling: A two-stage approach. *Psychological Methods, 10*(1), 40–64.

Cook, T. D. (1991). Meta-analysis: Its potential for causal description and causal explanation within program evaluation. In G. Albrecht & H. Otto (Eds.), *Social prevention and the social sciences: Theoretical controversies, research problems, and evaluation strategies*. Berlin, Germany and New York, NY: Walter de Gruyter.

Cook, D. J., Sackett, D. L., & Spitzer, W. O. (1995). Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam Consultation on meta-analysis. *Journal of Clinical Epidemiology, 48*(1), 167–171.

Cooper, H. M. (1982). Scientific guidelines for conducting research reviews. *Review of Educational Research, 52*(2), 291–302.

Cooper, H. M. (2016). *Research synthesis and meta-analysis: A step-by-step approach*. Thousand Oaks, CA: Sage.

Cronbach, L. J. (1982). *Designing evaluations of educational and social programs*. San Francisco: Jossey-Bass.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional materials*. New York: Wiley.

Dersimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7*, 177–187.

Evans, J. H., Jr., & Burck, H. D. (1992). The effects of career education interventions on academic achievement: A meta-analysis. *Journal of Counseling & Development, 71*(1), 63–68.

Fretz, B. R. (1981). Evaluating the effectiveness of career interventions. *Journal of Counseling Psychology, 28*(1), 77–90. dx.doi.org/10.1037/0022-0167.28.1.77

*Fukuyama, M. A., Probert, B. S., Neimeyer, G. J., Nevill, D. D., & Metzler, A. E. (1988). Effects of DISCOVER on career self-efficacy and decision making of undergraduates. *The Career Development Quarterly, 37*, 56–62.

*Garis, J. W., & Harris-Bowlsbey, J. (1984). Discover and the counselor: Their effects upon college student career planning progress. *ACT Research Report #85*.

*Garis, J. W., & Niles, S. G. (1990). The separate and combined effects of SIGI or DISCOVER and a career planning course on undecided university students. *The Career Development Quarterly, 38*, 261–274.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*, 3–8.

*Healy, C. C., & Mourton, D. L. (1984). The effects of an abbreviated Self-Directed Search on the career decision competencies of community college students. *The Vocational Guidance Quarterly*, 33, 55–62.

Heppner, M. J., & Heppner, P. P. (2003). Identifying process variables in career counseling: A research agenda. *Journal of Vocational Behavior, 62,* 429–452.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21,* 1539–1558.

Lent, R. W., Ezeofor, I., Morrison, M. A., Penn, L. T., & Ireland, G. W. (2016). Applying the social cognitive model of career self-management to career exploration and decision-making. *Journal of Vocational Behavior, 93,* 47–57.

Liu, S., Huang, J. L., & Wang, M. (2014). Effectiveness of job search interventions: A meta-analytic review. *Psychological Bulletin, 140,* 1009–1041.

*Luzzo, D. A., Funk, D. P., & Strang, J. (1996). Attributional retraining increases career decision-making self efficacy. *The Career Development Quarterly, 44,* 378–386.

*Luzzo, D. A. & Taylor, M. (1993-1994). Effects of verbal persuasion on the career self-efficacy of college freshmen. *CACD Journal, 14,* 31–34.

Matt, G. E., & Cook, T. D. (2009). Threats to the validity of generalized inferences. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 537–560). New York: Russell Sage Foundation.

Nichols, G. C. (2009). *Career interventions in context: A meta-analysis examining the influence of environmental factors on outcome* (Order No. 3367792). Available from ProQuest Dissertations & Theses Global. (305029604). Retrieved from https://login.proxy.lib.fsu.edu/login?url=http://search.proquest.com/docview/305029604?accountid=4840

*Niles, S., & Garis, J. W. (1990). The effects of a career planning course and a computer-assisted career guidance program (SIGI PLUS) on undecided university students. *Journal of Career Development, 16,* 237–248.

Oliver, L. W., & Spokane, A. R. (1981, August). *Integrative analysis of career counseling outcome research: Some preliminary findings.* Paper presented at the annual meeting of the American Psychological Association, Los Angeles, CA.

Oliver, L. W., & Spokane, A. R. (1983). Research integration: Approaches, problems, and recommendations for research reporting. *Journal of Counseling Psychology, 30*(2), 252–257.

Oliver, L. W., & Spokane, A. R. (1988). Career-intervention outcome: What contributes to client gain? *Journal of Counseling Psychology, 35*(4), 447–462.

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide.* Malden, MA: Blackwell.

Presby, S. (1978). Overly broad categories obscure important differences between therapies. *American Psychologist, 33*(5), 514–515.

Prideaux, L., Creed, P. A., Muller, J., & Patton, W. (2000). A review of career interventions from an educational perspective: Have investigations shed any light? *Swiss Journal of Psychology, 59*(4), 227–239.

Rubin, D. B. (1992). Meta-analysis: Literature synthesis or effect-size surface estimation? *Journal of Educational Statistics, 17*(4), 363–374.

Ryan, N. E. (1999). *Career counseling and career choice goal attainment: A meta-analytically derived model for career counseling practice* (Order No. 9930585). Available from ProQuest Dissertations & Theses Global. (304513406). Retrieved from https://login.proxy.lib.fsu.edu/login?url=http://search.proquest.com/docview/304513406?accountid=4840

Sampson, J. P., Jr., Hou, P.-C., Kronholz, J. F., Dozier, V. C., McClain, M.-C., Buzzetta, M., Pawley, E. K., Finklea, J. T., Peterson, G. W., Lenz, J. G., Reardon, R. C., Osborn, D. S., Hayden, S. C. W., Colvin, G. P. and Kennelly, E. L. (2014). A content analysis of career development theory, research, and practice—2013. *The Career Development Quarterly, 62*(4), 290–326.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton Mifflin.

Sheu, H.-B., Lent, R. W., Brown, S. D., Miller, M. J., Hennessy, K. D.,& Duffy, R. D. (2010). Testing the choice model of social cognitive career theory across Holland themes: A meta-analytic path analysis. *Journal of Vocational Behavior, 76*(2), 252–264. http://doi.org/10.1016/j.jvb.2009.10.015

Shevlin, M., & Millar, R. (2006). Career education: An application of latent growth curve modelling to career information-seeking behaviour of school pupils. *British Journal of Educational Psychology, 76,* 141–153.

Spokane, A. R., & Oliver, L. W. (1983). Outcomes of vocational intervention. In W. B. Walsh & S. H. Osipow

(Eds.), *Handbook of vocational psychology: Applications* (Vol. 2, pp. 99–136).

*Tempestini, M. A., Horan, J. J., & Good, R. (1988). The effect of two commercial software programs on indices of career development. *Computers in Human Behavior, 4*, 167–173.

Thoresen, C. E., Hosford, R. E., & Krumboltz, J. D. (1970). Determining effective models for counseling clients of varying competencies. *Journal of Counseling Psychology, 17*(4), 389–375.

Tracey, T. J., & Rounds, J. B. (1993). Evaluating Holland's and Gati's vocational-interest models: A structural meta-analysis. *Psychological Bulletin, 113*(2), 229–246.

Viswesvaran, C., & Ones, D. S. (1995). Theory testing: Combining psychometric meta-analysis and structural equations modeling. *Personnel Psychology, 48*, 865–885.

Whiston, S. C. (2017). Conducting general versus population and setting-specific meta analyses. In J. P. Sampson, E. Bullock-Yowell, V. C. Dozier, D. S. Osborn, & J. G. Lenz (Eds.), *Integrating theory, research, and practice in vocational psychology: Current status and future directions* (pp. 90). Tallahassee, FL: Florida State University. Retrieved from https://doi.org/10.17125/svp2016.ch9

Whiston, S. C., & Rose, C. S. (2015). Career counseling process and outcome. In P. J. Hartung, M. L. Savickas, & W. B. Walsh, (Eds.) *APA Handbook of career intervention: Vol. 1. Foundations* (pp 43–60). Washington, D.C.: American Psychological Association.

Whiston, S. C., Sexton, T. L., & Lasoff, D. L. (1998). Career-intervention outcome: A replication and extension of Oliver and Spokane (1988). *Journal of Counseling Psychology, 45*(2), 150–165.

Whiston, S. C., Brecheisen, B. K., & Stephens, J. (2003). Does treatment modality affect career counseling effectiveness? *Journal of Vocational Behavior, 62*(3), 390–410.

Whiteside, M. F., & Becker, B. J. (2000). The young child's post-divorce adjustment: A meta-analysis with implications for parenting arrangements. *Journal of Family Psychology, 14*, 1–22.

Wood, C., & Hays, D. (Eds.). (2013). *A counselor's guide to career assessment instruments* (6th ed.). Broken Arrow, OK: National Career Development Association.