Social Science Research and Policymaking: Meta-Analysis and Paradox

Steven I. Miller, Marcel Fredericks, Frank J. Perino

Abstract

The purpose of this article is to explore some of the non-obvious characteristics of the social science research-social policy (SSRSP) paradigm. We examine some of the underlying assumptions of the readily accepted claim that social science research can lead to the creation of rational social policy. We begin by using the framework of meta-analysis as one of the most powerful means of informing policy by way of empirical research findings. This approach is critiqued and found wanting in several ways. Several conceptual and definitional issues connected to the term "policy" are explored as well. A central argument is that even the best social science research is no guarantee of enlightened policymaking because the very (inductive) basis of empirical research militates against the possibility of going from research findings to policy. This claim is explored within the context of a central paradox. This paradox is explored in some depth. Finally, within the SSRSP claim, we analyze related issues such as the possibility of utilizing Mixed Methods and the politics of policymaking. We conclude that the SSRSP framework is, at best, a subjective one which ironically is needed, but one which is constrained by the very methods that is uses to formulate policy.

The purpose of this article is to explore a number of claims linking the supposed relation of social science research to social policy. While such investigations are not wholly new (Miller and Fredericks 2000; Miller and Safer 1993), we wish to give the topic a perspective that is rather distinct. The issue is important in a very basic way: the justification for the conduct of social science research (especially in terms of increasing methodological sophistication) can only be made if such efforts are tied to the development of social policy in some way. This is rather a strong claim but careful reflection must always bring social science research to social policy; otherwise, what would be the point of conducting it? One may, of course, argue that the further development of some particular methodological technique (e.g., hierarchical analysis) should be pursued for its own sake as a contribution, for instance, to the development of statistical theory. While such a position is plausible, the purpose of doing what we call "social science research" has always been viewed as being directed toward social "practices", or more broadly, the formulation of some type of policy. This pragmatic connection between research and practice

is not unusual in the social and behavioral sciences, and in some sense is the norm.

Our argument concerning the social science research and social policy link will, however, be somewhat different. Basically, we are suggesting that this link (hereafter SSRSP) is beset with so many problems that the entire issue needs to be rethought. The problems we will discuss concerning SSRSP fall into two broad categories: those dealing with conceptual clarity and those dealing with the possibility of applying sophisticated statistical methods (e.g., meta-analysis) to rational policymaking. Our argument, more specifically, is that the SSRSP link presents a basic paradox in that we usually believe that policy to be "workable" or "good," depends on the application of the best (and usually most recent) research methods and findings, and yet the very nature of these methods often precludes the possibility of arriving at rational policy decisions. We are not arguing that social policy is devoid of appeals to social science research, but only that such linkage is flawed because of a lack of insight into both the limitations of the policymaking process as well as (in many cases) the impossibility of translating findings based on sophisticated research methods into workable policies.

We will begin by analyzing some conceptual and logical aspects associated with the term "policy." This will be followed by assessing the issues of using sophisticated research techniques as a basis for attempting to make educational policy. Here, our principal analysis will be the examination of meta-analytic techniques. Lastly, we will have something to say about SSRSP and the possibility of employing so-called Mixed Methods strategies. We begin with the conceptual analysis.

What is Social Policy?

This seemingly simple term is fraught with all sorts of conceptual difficulties. The major one is simply trying to establish an adequate definition. For example, Miller (1973) attempts the following definition among several others.

A policy is a set of consistent statements arrived at by rational persons in an institutional setting which are directed towards:

  a)   the solution of a problematic situation, or
  b)   the creation of a new set of actions related to and directed at a specified

group of actors – who are part of a given institutional setting – with the intent of changing or modifying previously accepted behaviors.

Miller goes on to criticize this and other definitions of policy, but his intention is to alert the social policy community how even such a supposedly neutral term is notoriously difficult to define. For example, there is the difficult concept of "rational" that is often implicit in what we call policy. We assume that a certain class of actions is made by people who know what they are doing. However, their actions must presuppose some notion of policy itself before an assessment can be made of what is rational or not. This begs the question, of course, of what constitutes a rational action (a point we will return to later), but it is important to note that the very notion of "rationality" must be linked to the empirical methods used to determine its rationality. This unfortunately often leads to an undesired circularity, where rationality becomes methods and methods are the defining criterion of rationality.

Perhaps the need for conceptual clarity of what social policy can be is better illustrated through the often used tripartite division of policy formulation, policy implementation and policy evaluation. These conditions are believed to be not only defining characteristics of the term policy, but taken cumulatively, the necessary and sufficient conditions for policymaking as a process. This process seems on the surface to be unproblematic – to be rational if you will – but conceptual problems lurk. The three terms are in need of definition themselves, but let us assume their common sense meanings as being adequate and concentrate on them as constituting the "process" of policymaking. Let us also label the terms (1), (2), and (3), respectively, to avoid having to repeat them. One issue that first comes to mind is whether the three terms are sequential, i.e., (2) has to follow (1), and (3) has to follow (2). On this views, policy making is a series of steps that proceed in a certain direction, and a rational policy is one that assumes all three in that particular order. This interpretation seems reasonable if it assumes that all three are in a sense "independent", meaning, for instance, that (3) can stand independently as part of the process. The other possibility is that a given condition may be a "subset" of other ones. For instance, if we definite (1) adequately (2) and (3) are concretely/operationally defined as part of what (1) is about. That is, if we know what it means to "formulate" a policy, we also are committed to a given way to implement and evaluate it. But if this is the case, then the three elements are not independent. There is nothing wrong in this but it does show that the formulation stage is not only crucial, but it may determine in an a priori way the structure and function

of the remaining parts of the process. Now, if the formulation stage of social policymaking is driven by the methods of social science research (e.g., meta-analysis), then it sets the parameters on how the policy ought to be formulated, and this in turn, sets how (2) and (3) will proceed. Thus we have, by way of statistical analogy, only one "degree of (conceptual) freedom" for policymaking which may or may not be suitable depending on how the "problem" calling for the policy is conceptualized.

Keeping with this theme of requiring some level of conceptual analysis in social policymaking, let's take a look at (3) alone. Policymaking is a purposeful activity. That is, each policy is formulated and implemented with some "purpose" in mind that is expressed in terms of how it is evaluated. The evaluation of a policy is the norm or benchmark of whether or not it has "worked" in some sense. But it is exactly here that most of the ambiguity surrounding the nature of the policymaking process occurs. There is, first of all, the recognition that (3) is dependent on what is decided in (1) and (2) in a variety of ways. For instance, policy implementation is dependent on design and sampling strategies which dictate how (3) will be carried out. Should the design be quantitative or qualitative and what sampling strategies are appropriate conditions to the possibilities for deciding the "success" of the policy?

While such considerations are rather self-evident, there are often other implicit dimensions that are not usually recognized in the SSRSP paradigm. For example, a policy's "success" or "failure" is assumed to be a rationally based process that is open to public scrutiny. This may be so, but the "rationality" of (3) must be structured by how (1) and (2) are determined in the first place. Conceptually, this implies that we have some type of system of "rules" which direct us to making rational policy choices. But we don't; the closest to what could be described as "justification rules" are rather vague assumptions about what constitutes such rules. These "rules" seem to be mostly closely related to what we (the social science research community) determine to be acceptable methodological principles. Thus, if we want to establish a policy concerning the assignment of school homework, to use an educational example (Cooper 1989), the rules would involve a confirmatory (statistical) quantitative model in which we would specify a probabilistically structured design and corresponding quantitative methods to test a hypothesis. However, other rules may be invoked for the same issue – perhaps we come to believe that a qualitative or mixed methodological strategy constitutes the normatively appropriate rules for the study of this particular policy issue. The point is that even with (1), no matter how characterized, the parameters for the policymaking process are

set. We can, of course, say that (3) is the "test" of (1), and so we don't know what the outcomes will be – they may be what we expect or they may run counter to expectations. However, there is always not only a presumption of "workability" (i.e., success) for the policy implied in (3), but the success of the policy, if not guaranteed, then must be closely approximated by the very fact that EVIDENCE must first be garnered in the first place by engaging in (1). That is, the type and amount of evidence that is presumed to be necessary for developing a policy produce the very conditions that we then attempt to determine in (3). For instance, if we wish to construct a policy on the utility/non-utility of assigning school homework, we make a priori choices concerning two things: the choice of methodology and the validity of the pre-existing findings based on this methodology.

These two conditions are closely tied with one another and the desired outcome for (1) is that the findings, given the methodology, will reflect a "positive" outcome. The stronger such outcomes, the more "rational" the formulation aspect seems to be. However, the strength of this connection already implicitly influences the expectations in (3); we are, in effect, trying to evaluate a policy that we already have good expectations of succeeding. Of course, we could alter the conditions in (2) – that is change or alter the implementation phase from what has been done in assessing the methods – finding the link in (1). But if the policy formulation link is strong, a policymaker would not want to deliberately alter it in (2) as a "test" of its efficacy. Moreover, even if such an attempt were made, let us say by way of a meta-analysis, the very studies making up the meta-analysis could be very different in design, although their final metric may be standardized.

Thus, there are a host of conceptual issues that are implicitly present in SSRSP but that are seldom recognized nor addressed if recognized. At root here is the old "fact-value" problem (a variety of the "naturalistic fallacy") (Audi 1999, 582-583), whereby we are attempting to formulate an eventual "ought" (i.e., putting a social policy into practice) based on an empirical "is". The problem, of course, is that the results of policymaking need not logically follow from the "facts" that have been gathered. This consideration has never stopped policymakers from pushing their particular agenda, but what it suggests is that the process is heavily value-laden and that this cannot be reconciled in any fast and easy fashion by appealing to the "facts" (i.e., findings) of empirical inquiry. In other words, (1) does not necessarily imply (2) and (3), but if so thought, what will result is tautological reasoning – which, of course, is what we want to avoid. Beyond such considerations, there are other and somewhat more

technical issues associated with policymaking, and it is to these that we now turn, emphasizing the popular strategy of meta-analysis.

Meta-analysis and Social Policymaking

SSRSP is becoming increasingly dependent, as are other social as well as medical sciences, on the use of such techniques as meta-analysis. The motivation is twofold: the actual pragmatic formulation of policies that can be predicted to "work" and the incorporation of rather sophisticated statistical techniques which now can be claimed (in the quest for scientific justification) as properly belonging to a field of enquiry. Both of these aims are desirable, and there is little doubt that in some instances these techniques have resulted in forms of rational policymaking. This is not surprising since meta-analysis provides a forum by which disparate empirical studies can be reduced to a common metric, and so if policy formulation is desired on some topic, issue or problem, what better way to proceed then to show that some "effect" or direction can be shown to be better than others?

We will not be entering into the technical aspects of meta-analysis (see, for instance, such standard works as Glass, McGaw and Smith 1981; Light and Pillemer 1984; Rosenthal 1991) and will assume the reader knows of the general ways such analysis is conducted, although some technical concepts may be mentioned and explained as needed. We will rely principally on the concise presentation of Cooper (1998) and mention from his volume an example or two of actual meta-analysis projects.

In keeping with the first section, our primary objective will be to show that the SSRSP link has a number of problems which are not settled by the use of meta-analysis. By way of a very quick review, the intent of meta-analysis is to search the empirical literature for studies on a topic of interest, e.g., do teacher characteristics relate to student achievement? The studies admitted into the final meta-analysis must not only be empirical but also possess enough statistical information (e.g., means, standard deviations, statistical tests, sample size, statistical significance, etc.) so that an overall conclusion or "synthesis" may be attempted. Cooper (1998), for instance, not only documents the conceptual issues (e.g., conducting a literature search) that proceed a meta-analysis, but also explains the basic data analysis techniques that enter into such an analysis. For instance, there is the technique of combining significance levels where the

data from the various studies are transformed into Z scores, added and then summed over the total number of findings giving an overall estimate (Cooper, 120-124). There are a vast array of such techniques with increasing levels of statistical complexity and pros and cons associated with their use that Cooper documents.

However, no matter how sophisticated these methods become, the intent of meta-analysis remains the same: synthesize existing findings into one (or a few) overall results and draw some type of inference as to "workability." Now, while meta-analysis can be (and is) a powerful method of synthesis, it has a variety of problems when it is used as a basis for making social policy. The interesting thing is that while vital, these problems are either unknown or ignored. But ignoring them threatens, or at least impairs, the whole area of social policymaking and by implication the field of social science research itself. We will now examine a few of these issues and their consequences for policy making.

One issue that permeates the SSRSP relation is the lack of what we will call procedural rules. That is, there are no explicit guidelines for knowing when and how to apply empirical findings to any of the three stages of the policy-making process. Another way of stating this is that we do not have a "logic-of-choice" which governs what constitutes "rational" policymaking. By way of illustration, one of the basic techniques of meta-analysis is what Cooper (1998, 116-121) refers to as "vote counting methods." These methods have a variety of advantages and disadvantages, both conceptual and empirical which Cooper amply documents, but the basic idea is to assess the number of "positive" and "negative" studies on a given topic that can be transformed into a Z-statistic. Apparently, there are two ways the researcher may proceed: one is to separate the positive from the negative findings and choose the positive for further analysis, or combine both types of findings with the same sign, sum Z-scores, and use the overall finding as a test for the null hypothesis that the number of findings in each direction are the same.

The overall procedure appears to be rational, but it still begs the question of the logic of choice. For example, who should make the decision as to which vote counting procedure to use? But more importantly, what criteria should be appealed to which then lead to a rational decision? The idea is that supposedly the very criteria that are part of the statistical method will be sufficient for a clear cut decision. This is plausible but what should the rule be, e.g., "Rule for Voting Count Procedure: If and only if the combined procedure produces a finding that rejects the null hypothesis at p=.05 will we go ahead with policy formulation." Such a stipulation could, of course, be adopted (by whom would

be a related question), but it would also need to omit the possibility of using the separation method as either being "weaker" or illogical in some sense. However, such an approach would require some type of retrospective manner of analysis whereby it could be shown that combining versus splitting had produced more "consistent" results. But we can quickly see where such an approach would lead. The paucity of procedural rules leads to the paradox of policymaking previously mentioned in which we need data to proceed to make rational policy decisions, but the very evidence we appeal to does not provide the clear guidance we require. The lack of procedural rules incorporating a logic of choice simply compounds the problem, although, ironically, our methods (e.g., meta-analysis) are technically sophisticated with a defensible "internal" logic of their own.

Again, our intent here is to create an awareness (and possibly dialogue) of the central problems of policymaking that often are left unexplored in social science policymaking because of the belief that increased statistical sophistication will solve these problems. Leaving aside some of the technical aspects of meta-analysis for the moment, the same point as to what creates this lack of awareness might be illustrated by a more general consideration. For example, the vote counting method is really an expression (often implicit) of a cultural norm that is rooted in the logic of enumerative induction. The logic of induction, of course, undergirds all empirical operations and techniques but enumerative induction is a specific type that relates to the SSRSP link. In enumerative induction, we have a process which is a logical counterpart of the vote counting procedure in meta-analysis. Enumerative induction is about "counting" what we consider to be "positive instances" for our hypothesis, claim, or point of view (Manicas and Kruger 1976, 243-248). Thus, we may only wish to "count" those studies that already indicate statistically significant relationships as being "relevant" to what should count as acceptable evidence for, eventually, basing policy. However, with any notion of an inductive inference providing sufficient evidence for our hypothesis, there are a host of problems with increasing degrees of technical difficulty (see, Achinstein 2001; Salmon 1966). On the simpler side, there is the general problem of induction that no evidence is completely sufficient (or possibly necessary) for establishing the hypothesis beyond doubt (our use of "levels" of statistical significance is simply a way of providing us some comfort that the problem of induction will not keep us from doing research).

Should such considerations stop the pursuit of the SSRSP link? Of course not, but rather the point is that there cannot be exact parameters for mak-

ing good policy since the methods we use (no matter the degree of technical sophistication) simply cannot do what we ask of them. A further example to show this is what we will call the "problem of the independent variable." That is, the choice of studies in a meta-analysis must not only reflect, if you will, the "dependent" variable of interest (e.g., school achievement) but also the range and scope of the independent variables for a given study. There are two issues here: one is the "wideness" of the search to be undertaken, and the other is the conceptual relevance of the constructs within (and, of course then, between) the studies chosen. Cooper (1998, 20-21) is well aware of these problems and goes on to say:

> The only recommendation that can be made with regard to conceptual relevance is that the synthesist should begin the literature with the broadest conceptual definition in mind. In determining the acceptance of operations for inclusion within the broad concept, the synthesist again should remain as open-minded as possible. At later stages – notably, during data evaluation – it is possible to exclude particular operations due to their lack of relevance.

This view, however, does not eliminate other features of the independent variable problem within meta-analysis. In a perfect situation, all the independent variables would be the same in terms of conceptual definition and validity. For example, if all studies on school achievement and "social class" used the same definition and measurement approach, e.g., some notion of SES, then the studies could vary by design but still be useful in determining an overall effect. Thus, if there were construct validity for the independent variables, no real problem would exist, and we could even come up with a procedural rule. But, unfortunately, such ideal situations seldom exist. It is not enough, also, to claim that an independent variable, such as social class, has equivalent meaning no matter how it is operationalized. At a very general level it does, in the sense that it refers to a way of identifying the social phenomenon of stratification. On the other hand, a Marxist may have a very different conception of social class than a Structural Functionalist.

The point of this comparison is to reiterate the need to look at the independent variables as closely as the dependent variable in meta-analyses that are policy directed. In the previous example, this would have to be done by insuring some type of reliability measure among the various SES conceptions. If this were done, it would still leave open the question of how strong such an estimate would have to be. Likewise, one could look at the dependent variable measures and assume, if the correlations are statistically significant, the independent variables must themselves be highly correlated. However, such an inference might

not be warranted since different studies may not only use different statistical methods, but they may also vary on what (and how many) other intervening or extraneous variables are controlled. For example, one study may control for "ability," while another for "self-concept"; if the independent variables are conceptualized in similar ways and the controls are equally effective, then, perhaps, the independent is appropriate to assess for both studies. If, however, the independent variables are conceptualized differently, but produce similar effects, or if both are conceptualized similarly, but produce different effects, then the problem of inclusion remains for meta-analysis incorporation. While such issues may be important for meta-analysis, more emphasis has been placed on the technical and statistical aspects surrounding this procedure. A prominent one is known as "effect size" determination. In the following section, we wish to examine this methodological strategy and its potential importance to the SSRSP connection.

Effect Sizes and Social Policy

It would seem that computing an effect size would be one of the most compelling ways to develop policy based on empirical findings. The so-called effect size was developed by Cohen (1988), and basically consists of an estimate, the d-Index, which assesses the difference between two means such as the t-test (other statistics can be used, also), divided by the sum of their respective standard deviations over 2. The d-Index measures the difference in standard deviation units between, let us say, the experimental treatment and the control group. If, hypothetically, the index was .40, this would mean that the experimental group produced a four-tenths standard deviation increase over the control group (Cooper 1998, 128). Cohen's index is developed so that values of 0-.2 indicate "small" effects, .2-.4 "medium effects," and over .8 "strong effects" (see Gravetter and Wallnau 2004, 262-269). Effect sizes are connected with the "power" of a test (i.e., probability of committing a Type II error), and a variety of associated measures (see Cooper 1998, 128-136 for a discussion of various extensions of the d-Index and such measures as U3 and the odds ratio). Our basic concern, however, is how such a meta-analysis can contribute to rational policymaking?

While there are several technical strategies that can be used to compute effect sizes, the real issue for policymakers is whether such findings can be used for

policy formulation. Cooper, in attempting to analyze some of the issues associated with the "practicality" of using effect sizes, outlines a number of factors that should be considered. For example, given the conceptual development of an area, what effect size should the researcher accept as being important? Additionally, there are issues concerning (already mentioned) how similar variables must be across studies, as well as the type(s) of statistical designs, the "historical" fact that earlier studies may reflect different conditions than recent ones, which variables are or are not controlled, and so on (Cooper 1998, 173-178). Even with these several caveats, Cooper still believes meta-analysis is a viable means of applying findings to make "practical" decisions.

We do not necessarily disagree with his overall conclusions, but there are obstacles in saying meta-analysis provide the necessary guidelines for making rational policy choices. The problem lies in what we called previously the "logic-of-choice" issue. This can be illustrated in a table (6.5, 179) that Cooper uses to compare findings concerning the homework vs. no homework issue with additional findings from other meta-analyses. The original finding for the above was $d = .21$, which by Cohen's standard is not large, but on the other hand may be important given what substantive area is being explored. While we cannot reproduce Cooper's entire table here, we will illustrate some representative effect sizes listing the independent variable and the actual "d". These meta-analyses were directed toward examining "influence on achievement." For instance, there is one finding for Direct Instruction with $d = .60$ (Pflaum et al., 1980); the Amount of Television Watching with $d = .10$ (Williams et al., 1982); Co-operative versus Competitive Learning with $d = .78$ (Johnson et al., 1981); and Advanced Organizers with $d = .23$ (Luiten, Ames and Aerson 1980). There are many additional findings that Cooper presents in the table, but these will do to illustrate the point we are trying to make.

Let us assume the policymaker is concerned with a persistent pattern of low achievement in a large urban school district as reflected on State mandated tests. She needs to do something but what would be the "best" way to formulate a policy? Her initial strategy (based on observational data) is to require additional amounts of homework, and she bases this decision on the $d = .21$ meta-analysis previously mentioned. Although initially persuaded (by herself or others) to begin the policy implementation stage with such a finding, she becomes aware of the other meta-analyses illustrated above. Assuming both the original and the newly discovered analyses meet the technical requirements, what decision should she make? That is, what should constitute her logic-of-choice as a rational policy maker? We may also assume, which is often the actual

case that she simply can't choose all those findings she believes are plausible and implement them – one choice, and only one, has to be made.

The logic-of-choice issue is then tied closely with the one for the establishment of procedural rules. Thus, in the above example, a procedural rule might indicate choosing the largest "d" as the one to begin the policy implementation phase. Here it would be Cooperative vs. Competitive Learning, d = .78. However, the policymaker may either not have faith in this particular finding or not be able to implement it for practical or other reasons. If this is the case, another procedural rule might specify that (using the present example of a meta-analysis of meta-analyses) the policymaker rank order the results and choose from the list the one that is best suited for her purposes; so here it might be choosing the use of Advanced Organizers, d = .23. The selection of this one is stronger than that of homework but not as strong as d = .78 cooperative vs. competition. But are such judgments examples of rational policymaking? There is no way of telling a priori, and so the advocates of meta-analysis findings for policymaking must first deal with these issues.

To make matters more difficult, the other aspects of policymaking, implementation and evaluation, also come into play. For instance, if the policymaker chooses to go ahead with the Homework d = .21 finding, what commitments does she have to make for implementing the findings? And, of course, the emphasis is on the plural, findings; the effect size is a summary of previous studies. As pointed out, such studies may have differing designs and ways of conceptualizing the independent variable. The policymaker cannot simply assume that the effect size represents the exact same kinds of studies. But even if it did, the policymaker may not be able to implement the design in her particular situation (e.g., strong political pressures may mitigate against having children placed in experimental and control groups). Then there is the problem of the evaluation process: does she have to make decisions as to what instruments to use for the evaluation, and must these be a selection from those used in the original meta-analysis synthesis, or can she choose some other one? Finally, what standards is she to apply to judge whether or not the policy formulation and implementation, (1) and (2), have resulted in a "successful" evaluation effort (3)? If some of these considerations are plausible, we are confronted with our original paradox: we need empirical findings to make and justify policy, yet they cannot provide the clear cut direction we desire. Paradoxes can be explained, if not completely resolved, and we now turn to some ways of looking more closely at this process in relation to policymaking.

Policy Paradoxes

There are a variety of ways to attempt to address the SSRSP paradox mentioned above, but like all paradoxes, there are usually no completely satisfactory ways to handle them – although much literature exists in this regard (e.g., Rescher 2001). Our purpose is not to explore all these strategies but only mention a few aspects related to the policy issue. One strategy is to raise or lower what we call the "threshold effect." For example, using a broader societal example, social policymakers are often confronted with the equality-equity dilemma: can you provide for both simultaneously, given that you are committed to both? It appears this is not possible since, for instance, if you wish to treat students, for example, equitably (i.e., fairly) by way of some standardized achievement measure, simply doing this will produce a form of inequality. If we apply the threshold effect, we may treat everyone equally by applying such low standards that everyone meets them, thus fulfilling both conditions. This type of "policy" would not be acceptable. Using the same example, another approach would be to create a variety of subgroups and then limit evaluations within but not between the subgroups. Thus, within each subgroup, everyone is being treated "equally" in terms of the classifying variable itself. The acceptable threshold effect to provide for "equity" would be some "acceptable" (again, how determined and by whom?) range of scores (i.e., interval level band). Such arrangements might even be justified by arguing that they constitute some form of "distributive" justice and hence, as policies, are not only acceptable but desirable (Ryan 1993). In passing, it might be noted that the SSRSP model must somewhere along the line incorporate notions of justice, whether distributive, social or some other kind. This consideration, in turn, complicates the social policy issue since it now places another demand on empirical findings as evidence, namely that they be so structured by way of research designs that they lend themselves to a justification of a "justice" definition. However, even the most sophisticated research designs may not settle the equality-equity issue and may indeed exacerbate it.

Now, returning to the meta-analysis induced paradox and this holds for paradoxes in general, one may attempt to resolve it by changing some part of the language associated with it. This would be similar in logic, for instance, of changing some term to make an invalid argument valid (Barker 1989). In our example, we may simply want to argue that "some" types of meta-analyses, or some portion of these meta-analyses, are more useful than others for policy formulation. This strategy keeps the "good" (but, again, determined by whom

and how?) aspects of meta-analysis, while admitting that all cannot be applied simply because they are not relevant for the policy issue at hand. The problem here is that the paradox resolution depends on how "some" is defined and what criteria are applied to differentiate it from "not some." This is similar to the distributive justice situation mentioned above, namely, the criteria chosen are the very ones that are used to "justify" a particular framework, resulting in an undesirable circularity.

Again, a paradox consists in the fact that certain statements that compose it are themselves consistent when taken separately (the premises and the conclusion of an argument) but the conclusion can also be contradicted, and doing so, still produces a *plausible* statement. So, here we have a claim that "good" policymaking is dependent on empirical research; that meta-analysis is one of the best ways to garner the data needed to make "rational" policy decisions; and, therefore, meta-analysis will result in "good" policymaking. However, it may be that "good" policymaking is still possible without employing meta-analysis. But if so, what would have to be substituted? Again, one would have to argue that policymaking can be carried out in the absence of empirical approaches, or that "some" empirical techniques may be appropriate, but not necessarily meta-analysis. Both positions would require further elaboration, of course.

There are other issues that may be related to the paradox of policymaking in the context we have been discussing which are often overlooked. We will only mention a couple. In looking at meta-analysis as one of the potentially useful means of establishing the SSRSP link, we often forget that the subject matter of our investigations is such that techniques such as meta-analysis *must* be used. That is, meta-analysis is a means of circumventing what cannot be done in a democracy: conducting large scale, system wide, experimental studies to determine certain outcomes. This situation may not be true in other fields such as medicine, but it certainly is the case in much social research. The fact remains that, in democratic societies at least, we cannot deliberately manipulate situations and subjects so that specific hypotheses can be tested. In lieu of this, meta-analysis provides us (potentially) with a means of combining like studies to estimate overall effects. However, we have seen, this "aggregation" approach presents as many problems as it solves.

A further issue here is what we will refer to as the "ethics of policy failure." By this we mean the criteria that "ought" to be used to declare a policy has been evaluated and judged to have failed. To illustrate this, let us return to the "homework" example previously used. Let us assume the policy formulation stage is based on a meta-analysis which seems to indicate that assigning home-

work has produced "average" effect sizes. Based on these data, the policymaker decides to implement and evaluate the benefits of homework. Although they are seldom seen in this way, there are actually a range of ethical issues that present themselves, if only implicitly. For example, the policymaker might decide on an a priori hypothesis testing model whereby a level of statistical significance is determined and followed. She is then ethically bound to follow whatever the data dictate, even though she may not be happy with the results. Thus empirical assumptions constrain ethical decisions, although, ironically, we often believe that the two domains are not connected, or somewhat differently, that "significant" findings do not present any ethical concerns since such findings must imply a certain course of action. There are, of course, borderline ethical cases where the null hypotheses are not rejected but the level of significance is "close." If the policymaker is committed to the belief that homework is still a viable policy alternative, then she might still proceed with the policy. Is such a decision "ethical?"

This may be hard to tell: from one point of view it is because she is playing by the rules of the game; but it may also be ethical from the other perspective in the sense that, while not reaching statistical significance, the future implementation of the homework policy may have beneficial results for at least some of the students. These kinds of issues are usually dismissed by social science researchers as not lying in the realm of research but that of "politics." This may be so, but the politics of policymaking is still closely connected to how we do social science research – the two are inextricably linked. And so the policy paradox continues, although in different guises.

Policy Alternatives

Although the politics of educational policymaking cannot be summarily dismissed, a related route to SSRSP is to examine variations on the meta-analysis theme. For empirical studies, whether single or meta-analytic, the major problem remains that of induction as previously mentioned. That is, as a form of the naturalistic fallacy, it is simply very difficult to go from what "is" to what one "ought" to do.

Acknowledging this state of affairs, some researchers have advocated the use of so-called mixed methods to possibly forgo this research-policy link (Tashakkori and Teddlie 2003). A great deal has been written about such methods as

alternative ways to do educational research, and by implication, suggest their relevance for policy formulation (Rallis & Rossman, 2003). Mixed methods basically consist of designing research so that two (usually some version of quantitative and qualitative) methods or designs can be simultaneously utilized, although one method may take precedence over another (i.e., Quan or Qual Dominant) combined with specific sequencing strategies. The overall intent is to make a research case that the investigation of some particular question or phenomenon is better understood if multiple methods or designs are utilized.

For policy purposes, and using a one-study model, the intent would be to see if the policy formulation stage can be more clearly articulated from using such mixed methods. Thus, for instance, using our previous example, the policymaker may believe that the data provided by a mixed methods approach will make it more rational to decide whether the homework policy should be formulated and implemented. In other words, more and (supposedly) relevantly connected information is now available – one informs the other in terms of the research question. A variation on this theme is to make an argument that one particular approach within the panoply of possible approaches would better serve the policymaker. For instance, Miller (2003) argues that a Quan Dominant approach is the best way to proceed since the assumptions of quantitative social science research are better established and can serve as "placeholder" effects for the qualitative dimension of the study. Thus, the researcher may use a two-factor analysis of variance design and apply the qualitative dimension only to findings that are statistically significant. In this way, the quantitative sets the parameters for the qualitative. Of course, an opposite argument (i.e., Qual Dominant) could also be made.

The question of whether a single study mixed method approach, versus either alone, facilitates the policy making process is still an open question. And it must probably remain so, because to "answer" it would require further and perhaps impossible analyses where one would have to compare mixed versus non-mixed strategies for a given policy issue – this is logically possible but not practically feasible. Could mixed methods be used in a meta-analysis scenario? Again, the answer is probably "no" since this would require first a compilation of (let us say) Quan Dominant mixed methods studies and then a related analysis of the qualitative findings across these studies. Again, such a strategy is not impossible but certainly not practicable at this time. To make it so would require some type of large scale effort by the social science research community to invest in the study of some particular phenomenon from a variety of

mixed methods approaches. At best, a single well designed mixed methods study might be useful in shedding additional light on how its findings might contribute to policymaking.

Conclusions

We have attempted to illustrate that the SSRSP link is a complex phenomenon whose dynamics have largely gone overlooked. It is important to clarify these issues since the overarching purpose of social research is to have that research result in something, namely, rational policy. But policy is an illusive concept, one whose parameters are not defined by simply conducting social science research. As we have tried to show, the roots of the problem lie in the fact that the language of research methods cannot be readily translated into what ought to be done. There may be approximations, of course, but these approximations themselves are bound by certain restrictions that militate against a smooth transition from research to policy.

While now rather obvious, let us reiterate the point with a further example or two. For instance, the idea of "confidence intervals" is sometimes touted as a good way to proceed for translating research findings into policy formulation. The logic seems acceptable: if we can determine that an "outcome" lies within a certain "band", and can say this with some degree of statistical certainty, then surely such a finding out to be able to be converted into policy. The problem is that we do not exactly know what confidence interval optimizes the decision to formulate the policy. Too broad of a band incorporates more of the data that may be important for the policy, but again it may be too broad. Conversely, a very narrow confidence interval may give us more precision but be too narrow for the policymaker to take a chance on a policy that she sees as having broader consequences. More generally, by way of analogy, it is the old problem of trying to decide on what is a 'meaningful" level of statistical significance – is the .01 level "better" than a .05 level for instance? The point is that we can estimate the "risk" statistically, but a policy "risk" is different than a statistical one. Moreover, the issue cannot be necessarily solved by the "content" of the problem itself, for who is to say (but the irony of policymaking is that we must say it) that a policy implementing homework is more or less desirable than one which attempts to implement a parental involvement of policy? (Should statistical demands be greater for one rather than the other?)

By way of analogy again, determining rational policy choices on the basis of empirical evidence is a little like understanding the concept of a "limit" in calculus – it may be approached in some (theoretical) way but how do we determine when "enough" is enough? There is no completely acceptable or rational way of doing so. Well, where does all of this leave us then? We have to do something, but we really don't know what will ultimately "work", or perhaps more precisely, what policy "works" is left up to us. Whether our empirical judgments are decisive or not is a matter of how we conceptualize and use them. There is no independent way of telling a priori in which direction one should go.

However, policies need to be formulated, implemented, and evaluated. So, where does that leave us? At best, we must consider policy development as a "manageable-risk" process. That is, in a somewhat similar way that Bayesian theorists assign "subjective probabilities" (Hacking 2001), policy formulation that depends on empirical findings must be assigned a subjective probability that is based on the equally subjective criteria of what constitutes a sufficiently rational reason to say these findings indicate that we should proceed. Do such criteria also constitute a necessary condition (logically) for formulating the policy? Probably not since alternative, and perhaps contradictory, empirical evidence may exist for this particular policy issue. The problem is akin to formulating the "correct" form of a Bayesian statement: should it be, "given this evidence, formulate this policy," or should it be, "given this desired type of policy, use the following evidence?" The first is "predictive" while the second is "ex post facto," but the point is there is really no logic-of-choice where one is definitely better than the other. But, once again, what then should we say: something like, "Given the results of the meta-analysis done on the homework issue, there is a 50-50 chance this policy can be rationally formulated, implemented, and evaluated?" Probably not. Perhaps the best that can be done is that in the evaluation phase, a priori or a posteriori, we set the "bar" either relatively high or low, and then in either case, defend our decision when the results are reported. A high expectation, success outcome, will make us gloat that we did get it right after all, while a failure here will be explained by the need for further research. A moderately low but successful outcome will again permit us to gloat in a more constrained way, while a failure here will also give us the rationale for further study. In some sense the policymaker can never completely lose, since, as with the case of underdetermination of theory by evidence (Laudan 1990), whatever does or doesn't result can be equally well explained.

This is not meant to be a cynical conclusion to the SSRSP issue, but rather a skeptical one. The problem lies, dear Iago, in the paradox and not in the evidence.

References

Achinstein, P. 2001. *The Book of Evidence*. Oxford, MA: Oxford University Press.

Audi, R., ed. 1999. *The Cambridge Dictionary of Philosophy* (2nd ed.). Cambridge, MA: Cambridge University Press.

Barker, S.F. (1989). *The elements of logic* (5th ed.). New York: McGraw-Hill.

Bushman, B.J., and H. Cooper. 1990. Effects of Alcohol on Human Aggression: An Integrative Research Review. *Psychological Bulletin 107:* 341-354.

Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cooper, H. 1998. Synthesizing Research: A Guide for Literature Reviews (3rd ed.). *Applied Social Research Methods Series*, Vol. 2. Thousand Oaks, CA: Sage.

Glass, G.V., McGaw, B., and M.L. Smith. 1981. *Meta-analysis in Social Research*. Beverly Hills, CA: Sage.

Gravetter, F.J., and L.B. Wallnau. 2004. *Statistics for the Behavioral Sciences* (5th ed.). Belmont, CA: Wadsworth.

Hacking, I. 2001. *An Introduction to Probability and Inductive Logic*. Cambridge, MA: Cambridge University Press.

Laudan, L. 1990. *Science and Relativism*. Chicago, IL: University of Chicago Press.

Light, R.J., and D.B. Pillemer. 1984. *Summing Up: The Science of Reviewing Research*. Cambridge, MA: Harvard University Press.

Manicas, P.T., and A.N. Kruger. 1976. *Logic: The Essentials*. New York: McGraw-Hill.

Miller, S.I. 1973. Educational Policy as a Field of Study. *Educational Studies 4*(2): 57-60.

Miller, S.I. 2003. Impact of Mixed Methods and Design on Inference Quality. In *Handbook of Mixed Methods in Social and Behavioral Research*, edited by A. Tashakkori and C. Teddlie, 423-455. Thousand Oaks, CA: Sage.

Miller, S.I., and M. Fredericks. 2000. Social Science Research Findings and Educational Policy Dilemmas: Some Additional Distinctions. *Education Policy Analysis Archives 8*(3): 1-13.

Miller, S.I., and L.A. Safer. 1993. Evidence, Ethics, and Social Policy Dilemmas. *Education Policy Analysis Archives 1*(9).

Rescher, N. 2001. *Paradoxes: Their Roots, Range, and Resolution*. Chicago, IL: Open Court.

Rollis, S.F., and G.B, Rossman. 2003. Mixed Methods in Evaluation Contexts: A Pragmatic Framework. In *Handbook of Mixed Methods in Social and Behavioral Research*, edited by A. Tashakkori and C. Teddlie, 491-512. Thousand Oaks, CA: Sage.

Rosenthal, R. 1991. *Meta-analytic Procedures for Social Research* (rev. ed.). Newbury Park, CA: Sage.

Ryan, A., ed. 1996. *Justice*. Oxford, MA: Oxford University Press.

Salmon, W.C. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton, NJ: Princeton University Press.

Tashakkori, A., and C. Teddlie, eds. 2003. *Handbook of Mixed Methods in Social and Behavioral Research*. Thousand Oaks, CA: Sage.

Steven I. Miller, Ph.D.
School of Education/Department of Philosophy
Loyola University Chicago, USA

Frank J. Perino, Ph.D.
Associate Professor
College of Education
Northeastern Illinois University, USA

Marcel Fredericks, Ph.D.
Professor
Department of Sociology
Loyola University Chicago, USA