Jordan, Gretchen B (2013). A Logical Framework for Evaluating the Outcomes of Team Science, prepared for the Workshop on Institutional and Organizational Supports for Team Science, National Research Council, Board on Behavioral, Cognitive and Sensory Sciences, Committee on the Science of Team Science, October. http://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_085491.pdf

A Logical Framework for Evaluating the Outcomes of Team Science

Gretchen B. Jordan

1. Introduction

This paper has been written to provide ideas on evaluation of outcomes to a panel charged with developing a research plan for study of the science of team science. The National Research Council and funders of research including the National Science Foundation and the National Institutes of Health are interested in knowing how to evaluate the effects of team science beyond scientific impact in order to know better how to enhance the effectiveness of collaborative research in science teams, research centers, and institutes. Collaboration has come to be the norm in publicly funded research and pressure for demonstrating outcomes and value to taxpayers is high, so the question posed resonates with research funders everywhere.

Several "science of science" or "science and innovation" initiatives have formed in the U.S., Europe and Asia since 2005 to address the lack of data to answer this question. One of these is the Science of Team Science, a new interdisciplinary field that is concerned with understanding and managing circumstances that facilitate or hinder the effectiveness of collaborative research, whether these are large or small scientific teams within an organization, or entire research centers or institutes. There is a general belief, supported by some empirical evidence, that scientific breakthroughs occur for teams that would not be attainable by either individual or simply additive efforts.

The evaluation challenge implicit in the question being asked is a large one with four related requirements. First, the outcomes of the science need to be determined. Second the characteristics of the nature of the team, its work, the organization, the sector that is to absorb the research findings, and other influencing factors need to be determined. Only then can the third task, analyzing the "effectiveness" of an individual team or center, be completed. Then to begin to construct theories of what works and doesn't work and why, it is necessary to collect and analyze data similarly across a variety of situations in order to see patterns and begin to generalize. Most would agree that the current outcomes measured in Science of Team Science, primarily bibliometric publication counts, journal impact factors, and levels of interdisciplinarity, are not sufficient for answering the question.

In this paper I propose a logical framework for evaluating a sequence of outcomes for team science considering the characteristics of the teams and the context for the science and its translation into outcomes, building on related research in the science of team science and related literature and practice in research evaluation (not all of which is referenced). A logical framework describes the goals, strategies and activities of the program or organization being evaluated and based on that suggests the set of evaluation questions to be answered and indicators to be collected and analyzed in order to answer those questions. The proposal should be viewed as a suggestion for further discussion and work because evaluation frameworks for something as complex as the effectiveness of team science would normally be done with a team of experts over several months.

This framework builds on my extensive earlier work in the area of logical frameworks for evaluation of science, technology development and technology deployment programs primarily for the U.S. Department of Energy. The Framework pays particular attention to the outputs and attributes of outcomes that depend on the radicalness of the objectives and scope of the science. This builds on and extends the work I did with colleagues related to assessing the research environment. We developed a typology of innovations called “research profiles” (Hage et.al. 2007, Jordan et. al, 2008 and 2012) that has the dimensions of *incremental vs. radical advances*, and *narrow vs. broad scope of focus*. The research profiles are operationalized in this paper as criteria for measuring near term outcomes and in part the importance or magnitude of longer term innovations or potential innovations emanating from a particular science team or research center/institute. The paper will also consider the larger contextual factors that influence the near-term, mid-term, and long-term outcomes of team science. Such factors include the nature of scientific problem, how quickly a solution to the problem is needed, and the readiness of potential users to adopt the translational outcomes of the research (e.g., new technologies or medical treatments).

The paper will propose output and outcome indicators which are based on the logical framework, and options for the timing of evaluations and the questions that would be answered. Categories of indicators to answer those questions are proposed and shown in a diagram, and a demonstration of how indicators might differ by research profile is provided. Methods that could be used to assess outcomes are briefly discussed. The paper concludes with recommendations for further research to direct research to build a science of team science.

2. Evaluating Outcomes of Science Generally - Use a Big Picture

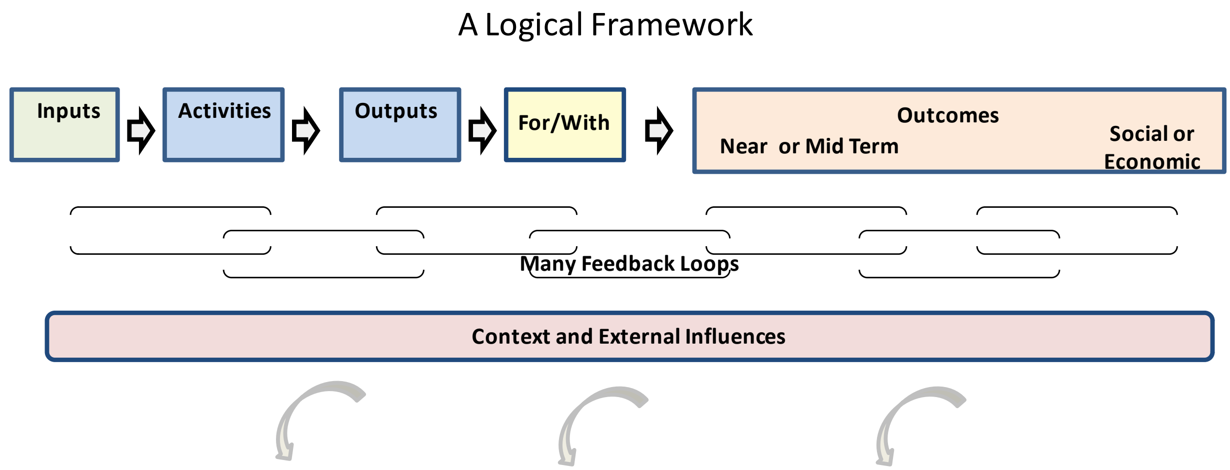
Given the complexity involved in discovery and translating new ideas into application and practice, it is not surprising that current science and innovation policy and program rationale and evaluation use a fragmented picture of the innovation process. The analogy was used at one meeting of evaluators on the subject of people who are blindfolded describing an elephant differently depending on what part of the elephant they are touching. Legs may seem like tree trunks, and ears like large fans. Without looking at a complete picture of the elephant, it isn’t possible to appropriately evaluate how the elephant functions or how various parts contribute to that functioning. Looking at only part of the elephant gives incomplete or incorrect answers. Greg Tassey, in his work on technology policy (2007), argues that the imperative is to switch to a dynamic version of a full life-cycle model of R&D Policy Analysis. Not only is the “black box” model of research and technology development (R&D) not sufficient, but the manner in which a technology diffuses must be much more clearly understood and taken into account.

Analysis and evaluation using an organized picture of the complex life cycle of innovation can identify blockages to innovation and improvements that are needed in existing policy and programs. Evaluation using an agreed upon big picture of the innovation process that includes knowledge production as well as product, process and organizational innovation, could help assess programs fairly within the broader context. Furthermore, similar studies would allow evaluation synthesis of study results and thus better tests of existing theories and new understanding of the underlying reasons programs successfully achieve desired outcomes or not (Jordan 2010). The challenge to develop a logical framework for assessing the outcomes of team science that goes beyond science outcomes requires developing a big picture of the innovation process, in my opinion. What I suggest here is a highly simplified picture of the non-linear model for research and theory on diffusion of innovations. This builds on existing theory-based frameworks and earlier generic logic models developed by the author and colleagues (Reed and Jordan 2007, Jordan 2010).

First, here is a brief introduction to logical frameworks for those not familiar with them. The central element of a logical framework is the logic model, a popular tool for developing and describing the rationale for a policy or program and its context. A logic model is a plausible and sensible model of how the program will work under certain environmental conditions to solve identified problems (Jordan 2013). The logic modeling process makes explicit what is often implicit. Also, if done carefully the process lays out a “theory of change”, the plausible pathways through which resources translate into outcomes, and mediating factors that can help or hinder success at key points. Much has been written about the logic model forming the basis for good evaluation and performance monitoring, as well as its use in program design and building a shared understanding of what an effort plans to achieve and how that will be achieved.

Figure 1 shows a generic logic model, with inputs used in activities that produce outputs for some potential users who then apply the outputs to change behaviors or attitudes or actions putting to play a sequence of near, mid, and longer term outcomes which are the end goal of that set of activities. The logic elements in this diagram are arranged left to right as to an extent they do take place across time, with iteration and feedback loops shown only notionally. Usually a program or group of activities with a goal and a budget, will have multiple activities solving the problem addressed through more than one pathway so a program logic diagram would be multiple lines of interacting logical elements. Winter and Berente (2012) argue that team effectiveness may hinge on participants' understanding how goals of different members of the team affect team goals because this is essential to overall success. I argue that logic modeling and other road mapping exercises can build that shared understanding.

Since programs are designed for a specific contextual situation it is important that that be described in the framework as well. All the major external influences on the program need to be assessed during outcome evaluation if the program wants to explain its success, or lack of it, taking these other influences into account. The indicator data, both qualitative and quantitative, to be collected and analyzed during an outcome evaluation is determined by the logic model. As shown in Figure 1, a good set of indicators will be balanced across all the various elements of the logic model.



Indicators for						
Inputs	Activities	Outputs	Interactions	Near term Outcomes	Mid term Outcomes	Social or Economic Outcomes
Characteristics of likely differentiating factors; External influences on achievement						

Figure 1. A Basic Logical Framework with Logic Model and Accompanying Indicator Categories

3. A High Level Logic Model for Science Teams and Their Outcomes

Let us start examining the outcomes of team science at a high level of abstraction. The end goal of publicly funded team science is to provide economic or social value to the public. That could be a myriad of things, from contributing to the U.S. being a leader in a particular area of science, to improvements in consumer products or the attitudes toward behaviors that increase the risk of disease, or improvements in the healthcare system or health generally. As these examples and Figure 2 show, we suggest that there are three general categories of outcomes of science: effects on the science community, effects on the economy or society, and effects in between these two where the knowledge is being considered or applied by policy makers or businesses or public groups but has not yet translated to social or economic outcomes.

The logic model also calls out the precursors of the outcomes, the building and managing of the teams, their activities and outputs and their interactions with the users of their outputs. Notice that these interactions could be either with users within the scientific community or in other cases, including policy research and translational research, with an application community. These interactions might occur during planning, during the research or after the research report or publication or combinations of these. There are many possible feedback loops between any science project or program and the outcomes. The science can push development of new products or solutions never dreamt of by technology developers, health organizations, and others. Alternatively, organizations developing a technology or seeking a solution to a specific problem will engage scientists in their specific focused area. Early versions of a technology or solution are tested in the actual operating environment and sent back for further research.

This logic model has the same general categories of outcomes in the same sequence as the outcomes model developed through concept mapping for the Transdisciplinary Tobacco Use Research Centers (TTURC) (Stokols et.al. 2003). That model has a group of outcomes labeled collaboration which is called “For/With” in the model here. The TTURC model also groups science outcomes, and then has an area for transition to application and policy use, ending with the sector-specific health outcomes. A nuance in the proposed simple logic model is to show that some team science may have application in the nearer term. Translational science teams working on more incremental problem may affect behavioral risk factors to health quite quickly, for example, if a solution is simple and compelling. Research teams at an automobile company may be called away from designing the brakes of the future to solve a technical problem noticed during testing this year’s model.

Two other elements to notice in Figure 2 are the Inputs and the Contextual characteristics. Because building teams takes time and how they are managed influences their success (Stokols et.al. 2013), this is called out as an input. All science success occurs in a context and to analyze why something may or not happen or did or did not happen, contextual information is needed. Winter and Berente (2012) argue that reflection on contextual issues should be at the foreground of the science of team science.

As part of planning an evaluation, or during project or program design or redesign, key parties would develop a model (or roadmap) with specifics within this generic template, or the more detailed one provided later. Considering the context, one can ask “what non-program factors might prevent this outcome from happening?” Then there can be redesign where possible to ensure that doesn’t happen.

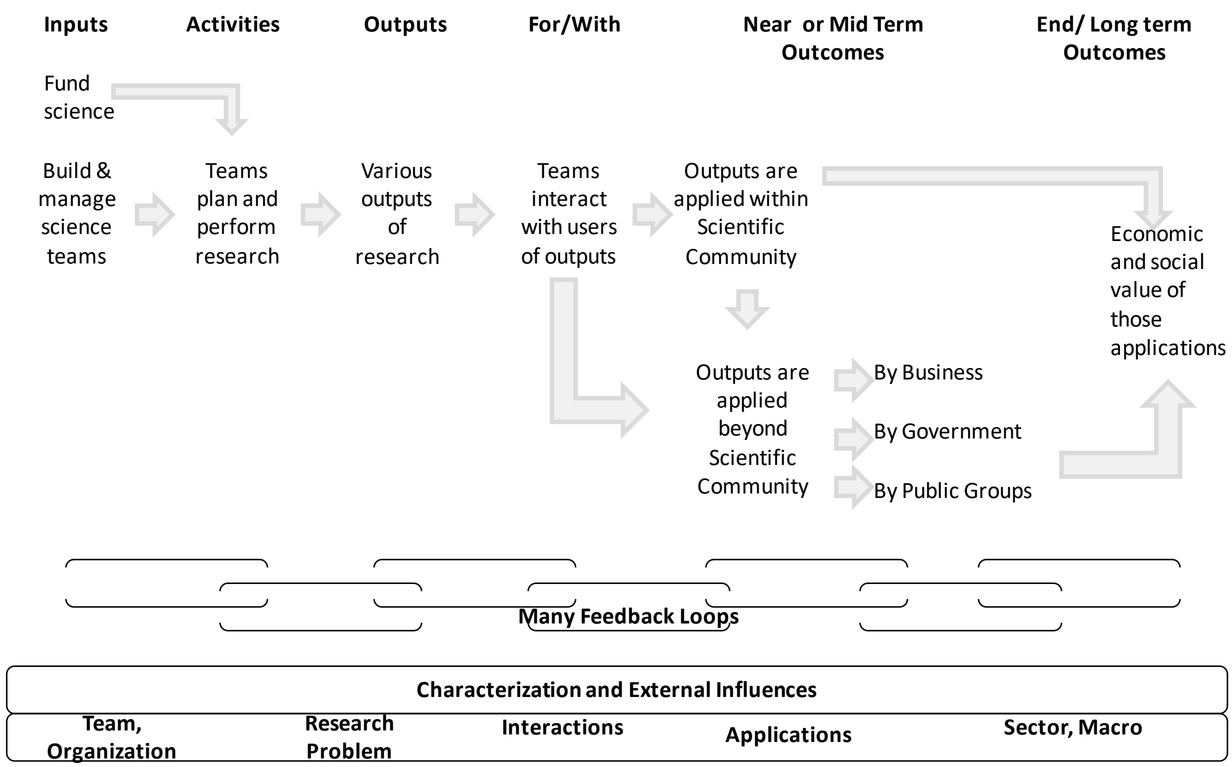


Figure 2. A High Level Logic Model for the Outcomes of Team Science

Thus major characteristics of context are called out in this logic model because if a framework for evaluating outcomes is to help determine how to improve the effectiveness of science teams, the conditions leading to those outcomes also must be known. Teams and their management differ. The timing and breath and radicalness of outcomes will differ in part depending on the arena of research problem (basic, applied, or development research, manufacturing research, or marketing research (Hage et. al. 2007, Jordan et.al. 2008) and on the scope and radicalness of the objectives of the science being done. The team science literature has characterized interactions of teams and team and users (Stokols, 2006), and it is possible to categorize potential or actual applications in terms of the sector and actors involved and their readiness to absorb the near term outcomes of the science team. Finally, there are the general external influences at both the sector and national levels that can help or hinder the success of the science team. This would include general economic conditions, specific government policies, cultural norms and the like. All of this will be discussed in more detail.

4. Outcomes Differ for Different Research Profiles

The knowledge production system consists of a range of research organizations that produce various kinds of research results. Research objectives differ. Thus expectations for performance should differ. In work with Hage and Mote and others we defined Research Profiles (Jordan et.al 2003, Hage et.al. 2007; Jordan et. al. 2012). Figure 3 shows the four types of research profiles associated with two primary strategic choices: the relative degree of risk or desired discontinuity and the relative scope of the research problem or its systemic character. These profiles suggest different management strategies, a contingency theory which is beyond the scope of this paper, but likely to be of interest to the science of team science community. Winter and Berente (2006) argue for considering contingency theory from the organizational sciences. The research of Shenhar (2001) on engineering projects is suggestive of how these two dimensions of research objectives can be operationalized in scientific research, where both the idea of scientific or technological uncertainty and systemic scope are in effect.

We define the research objectives dimension as *Degree of Radicalness* in the scientific or technological advance, on a continuum from incremental (or normal or straightforward) to radical. This can be operationalized looking at the degree of change in the state of the art, centrality of the problem, and the discovery of a pattern that upsets existing theory or a technology that creates a new market niche. For scientific research, the task environment is the knowledge world or "the state of the art," that is, how much is known, and what is considered to be an important scientific concern or requirement. Radical advances in science sometimes occur when a central problem is solved, such as the identification of the structure of DNA. Sometimes this also happens when a major discovery is made or when a research finding challenges an existing theory.

This dimension has been discussed in the science of team science. For example, Stokols, Hall and Vogel (2013) have distinguished between interdisciplinary and transdisciplinary team science in terms of the degree of novelty and innovation reflected in the conceptual frameworks, methodological approaches, and translational advances produced by particular research teams and programs. They discuss the *distance* or tension between the "existing landscape" of research within a particular field, and the imagined "futurescape" of unexplored terrain revealed through a transdisciplinary research program that "pushes the boundaries" beyond existing and widely accepted theories and methods. They also discuss how transdisciplinary research often exerts *transformational* impacts within the realms of theory development, research, and community practice.

The second dimension of research objectives is the *Scope of Focus*, a continuum from narrow to broad, further defined by the number of variables or processes or components or the number of levels or of systems involved, or the extremeness of the environments of the work. The question of the amount of the scientific advance can involve multiple outcomes, that is, the number of variables or processes that are being researched at the same time. Some fields have a systemic quality, that is, a large number of variables have to be considered at the same time. An example would be research to support cessation of smoking cigarettes, a health risk that required research to support and integrate multiple levels of government policies, change clinical practice, and inform and persuade the media and the public. Further, not all scientific or technical problems can be approached with small research teams. The size of teams is often discussed in the science of team science. Many research problems require a large scale focus and large teams. It took a large-scale inter-organizational program to coordinate a range of efforts so that the time to complete the entire genome was accelerated.

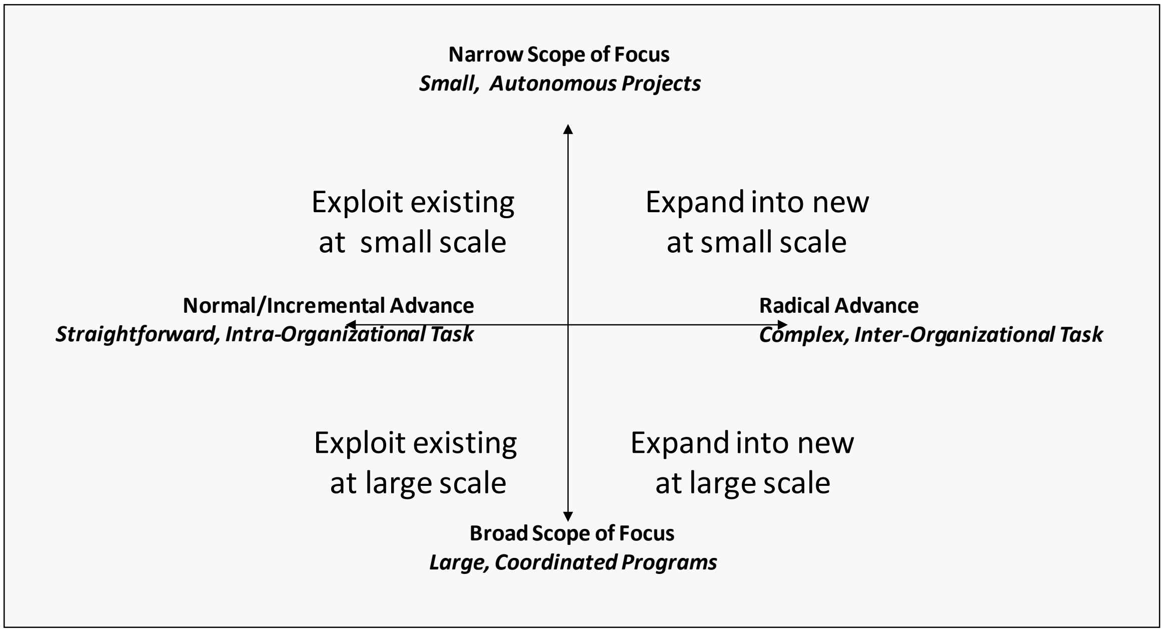


Figure 3. Research Profiles: Different objectives and scope of focus means different tasks and outcomes.

The two dimensions define four research profiles or strategies:

1. Expand into new at a small scale - narrow scope of focus aiming at a radical advance

2. Expand into new at a large scale - broad scope of focus aiming at a radical advance

3. Exploit existing at a large scale - broad scope of focus aiming at an incremental advance

4. Exploit existing at a small scale - narrow scope of focus aiming at an incremental advance

The profile most of interest to the science of team science is the "Expand into new at a large scale," perhaps because of the general public interest in solving big societal problems through research advances as soon as possible. The examples of each profile that follow will attempt to convince readers who are not already convinced that all profiles are present and valuable, depending on the circumstances. Objectives depend on the problem to be solved and timeframes for solutions, and could be either to do good, normal science or to break new ground. Of course, there is the problem that few claim they are doing incremental or normal science. Research managers have more difficulty convincing people to invest in incremental research and researchers are rewarded more for dramatic single discoveries than for a lifetime of smaller steady contributions.

Expand into new at a small scale

This would be research done by an individual scientist or a small team that is very exploratory, or in the stage of preliminary investigation. It may be an area that is uncharted or use a methodology that is unproven, but the impact on science or technology if it works would be dramatic. Moving beyond a bench scale prototype would usually be beyond this narrow scope of focus. Small startup firms are bought out by large firms that can handle the scope required by scale up, for example.

Expand into new at a large scale

One example of this profile is research on a new energy source for transportation such as hydrogen-fueled cars. Research is needed on hydrogen production and hydrogen storage as well as the vehicles that could use hydrogen. There is also research needed on the infrastructure, distribution and delivery systems for use of these vehicles and the safety of such a system

Exploit existing at a large scale

The National Oceanographic and Atmospheric Administration (NOAA) has a mandate that requires a broad scope and expensive and specialized equipment and teams to collect weather, climate change and other relevant data, including satellites, ships, buoys, and planes. Other examples could be found in the area of national security. Shortly after September 11, 2001 the White House Science Advisor announced at a public meeting that emphasis had to be placed on advances that brought forth solutions in the near term, not the longer term. Particularly in industrial research, there is a progression from a radical advance in a product being followed by a period in which incremental advances and modifications are made.

Exploit existing at a small scale

NOAA has a portion of its research that is needed to maintain an unchanging stream of climate data even when upgraded satellites collect that data. Writing the necessary algorithms is just an evolutionary change, yet this is essential to the researchers depending on constant data in order to measure trends. There is also the generalization and codification of a body of research that sets the stage for new exploratory research.

5. Timing of Outcome and Effectiveness Evaluation and Questions Asked

This paper describes just two possibilities for timing of outcome evaluation in order to keep this simple. These is shorter term outcome and effectiveness evaluation three to five years after start of a program or between evaluations, and retrospective outcome evaluation ten years or more after the effort started. Both of these will be more cost effective if there is some routine data collected annually. This would include data on .inputs, activities, outputs, and interactions. Given the often long time lapse between the research and application, and the changes in the teams and their characteristics and interactions, it is best to collect some data routinely. Then this data will be available for the more in depth evaluation, say after five years. Also the initial data collection can include a base line on the science and problems it potentially will inform so that a comparison can be made.

Possible levels of data collection and analysis are evaluation of outcomes of an individual team (which would change some over time) or outcomes in a center or institute with multiple teams preferably all aimed at an area of science or a problem area. Any less focused evaluation would find it impossible to collect necessary data and to separate signal from noise when it came time to look at factors influencing effectiveness.

Shorter term outcome and effectiveness evaluation

Evaluations three to five years after the start of a research effort can examine outputs, connectedness and science outcomes. For some research such as translational research and policy research, the evaluation can also examine progress toward applications. These evaluations may benefit from the body of evidence available from related evaluations. Where studies have been done to gather evidence that links early outcomes such as industry involvement and co-investment to new product development by involved companies, demonstrating co-investment by industry is argument for the potential for longer terms outcomes.

Here are the kinds of questions that could be asked and answered in the relatively short time frame for outcomes to occur.

- What have been the quantity, quality, timeliness and cost of outputs?
- What is the connectedness of team members with the potential users of their research?
- What outcomes on science have been observed?
- Where applicable, what application or adoption infrastructure outcomes have been observed?

Questions about contextual factor could include:

- What are the objectives of the research?
- What have been the inputs? What were the activities?
- What are key characteristics of the team/center? Of the profile of the research?
- What are the characteristics of the interaction?
- What are the characteristics of application(s) and sector of the application?

Questions on effectiveness include:

- Is there correlation between contextual/team characteristics and outputs and outcomes?
- What worked and what worked less well?
- How do outputs and outcomes compare to similar efforts done by individuals?

Retrospective outcome evaluation

Outcome evaluation that looks back on past research (retrospective) is useful for demonstrating team, program, or center effects on long term goals such as improved health status. However, it is not likely enough data will be available to connect outcomes to a particular team in order to analyze effectiveness of particular aspects of the management of that particular team. There are too many other influences over the ten plus year time frame. That said, case studies using similar protocols could draw conclusions using meta-analysis across studies. .

The outcome questions that might be asked and answered in these studies include:

- What outcomes have been observed?
- How do these observed outcomes compare to the stated objectives for the team/Center research?
- How do these outcomes and the value they provide to science and society compare to some standard (such as peer opinion)?
- Can an observed outcome be traced back to team activities (backward tracing)?
- What long term outcomes can be observed where the team/center research contributed (forward tracing)?
- Is there a plausible story of how team/Center activities contributed to short term outcomes which then contributed to midterm outcomes which then contributed to longer term observed outcomes?
- What are the benefits/value of the observed outcomes compared to the expenditures on the research?

(Where possible) questions about contextual factor may include:

- What have been the inputs? What were the activities? What were the objectives of the research?
- What were key characteristics of the team/center? Of the profile of the research? Of the characteristics of the interaction? Of the application and sector of the application?
- What were the macro institutional conditions and rules, that is, general economic conditions, capabilities over time for this particular area of research and application, and influences of modes of coordination in the science field, application area, and between the two?

6. Describing in More Detail a Sequence of Outputs and Outcomes for Team Science

Figure 2 introduced a high level generic logic model for team science outcomes. This section will add more detail to each element of the logic model and then the following section will show the more detailed logical framework with these categories of supporting indicators.

Inputs

The inputs of team science include the funds for the science, of course. Inputs also include the science team and its capabilities, and the capabilities of staff managing or supporting the team. Existing networks of each team members can also be an input since these people may provide ideas and constructive criticism on an informal basis. Inputs also include the available facilities and equipment and what those enable the team to accomplish. All science builds on previous science, so it is important to know about the existing knowledge base, both scientific and technical, in the area that the team draws from and will add to. Finally various aspects of the organizational environment are an input into the research. The science of team science is investigating aspects of this, in particular organizational structure and management. The current Research Evaluation Framework for University Research in the U.K. includes "Research Environment Vitality" as one of its evaluation criteria. In a multiyear study my colleagues and I did for the DOE Office of Basic Energy Sciences, we categorized the factors influencing the research environment in the following groups: (1) Autonomy, Integration of ideas, and Ability to explore; (2) Agile but long term investment, External collaboration, and Focus with defined goals; (3) Rewards and Value-added research managers; and (4) Organizational support, quantity and quality of resources, and coordination by management (Jordan et. al. 2003).

Activities

The activities of research have been variously categorized. The five areas of Kline and Rosenberg's non-linear model are often used and are recommended in this Framework. Similar categorizations are part of the Stage Gate process and Technology Readiness Levels so these could also be utilized. The point is to know for an evaluation of outcomes the type of research done because there are considerable differences across categories in typical time lapses between the research and its application. It is understood that basic and applied research activities are different from development research activities, manufacturing research and market research activities. Some teams integrate across one or more of these areas, and that integrated approach could affect the timing of research outputs and applications and the radicalness of those outcomes.

Outputs

There are four categories of research outputs suggested. The first three of these are categories established by NSF in the 1990s: Ideas, Tools, and People. The fourth is mentioned in the science of team science literature: Planning for transition to application. Ideas include many of the outputs regularly measured, such as knowledge advances and the excellence of the work. Indicators of these include quantity and quality of publications, patents and other intellectual property, and awards and peer recognition. New research tools, techniques and facilities are another valuable output. The people category can include students and post docs educated and trained whether or not they remain in science. Some also include the existence of a critical mass of people or communities of practice in a field as indicators. Preparation for application is likely a useful output indicator because it can be measured in terms of involvement of potential users in planning or teams and production of written plans. Since measuring something puts emphasis on it, this could push behavior in that direction. It also is an indicator of more likely future utilization of research outputs.

Interactions

Because partnerships and hand offs to the target audience for outputs is so essential to achieving program outcomes, many logic models put “For/With” or “Reach” explicitly in the model. For team science, both collaborations and interactions with next stage users are key, and an indicator of potential or actual influence. The term “Connectedness” may capture both the interaction and the level of integration. Team science literature defines different kinds of teams: multidisciplinary, interdisciplinary, and transdisciplinary research and research teams. The latter has teams with multiple functions and downstream users of the research represented. Some have also noted the challenges of inter-sectoral teams such as multiple levels of government. Another way of assessing connectedness is the extent to which there are individuals or organizations who act as intermediaries or boundary spanners between fields or functions.

Possible measures of the level of connectedness could be co-authorship, the density of a network or the integration implied by the mechanism for collaboration and coordination. Mechanisms include joint planning, co-funding, co-location of researchers, establishing common language or curriculum, and assignment of responsibility for bridging various interests and knowledge sets.

Science Outcomes

Science outcomes are “near term” outcomes compared to transitions to application or outcomes on society beyond the area of knowledge aspects. This framework proposes five categories of science outcomes. The first two are not commonly used and are explained more in the section on differences by research profile. The other three are more self-explanatory and commonly used. Effects on the organization or structure of knowledge captures changes in the way science is done, which is of particular interest to team science. Working in transdisciplinary teams is a structural change. Using modeling and simulation and experiments to build theory is another. These changes could be either radical or incremental in aims and effect. The impact on science can be measured by the extent to which the advance in knowledge changes the state of the art in that area of science. Effects on science infrastructure are a near term outcome because the effects provides the resources for future research as well as being inputs and outputs of current research.

Application and Adoption Outcomes of Science

There is a tendency for those interested in evaluation of science or innovation policy to ignore the complex nature of the diffusion of a product, process, or policy, and the arenas of manufacturing and marketing research which are more market focused. Exceptions are cases where there is a clear public good associated with diffusion and adoption of a technology or practice, such as technologies or processes that are more energy efficient or public health practices. One problem in ignoring technology diffusion in research evaluation and policy decisions is the possible disconnect between science advances and what private sector R&D and the market want and are willing or prepared to absorb. Another is holding unrealistic expectations about the infrastructure and compatibility necessary and time frame required for a technology advance to be absorbed.

As part a DOE Office of Energy Efficiency and Renewable Energy (EERE) framework for evaluating the impact of technology deployment programs Reed and Jordan (2007) developed a logic model for diffusion based on the seminal work of Everett Rogers (1995) that defined the diffusion process and what influenced it. This model examines those influences in four domains of the market: the end user of a technology and three infrastructure domains, business infrastructure, government policy and program infrastructure, and information infrastructure. A logical framework for assessing the impacts of health sciences research developed by the Canadian Academy of Health Sciences used similar categories for how the research translates into application. Those are the health industry, other industry, government and public groups (Frank and Nason, 2009]

Societal Outcomes

These outcomes include changes specific to a sector such as health where improvements in morbidity and mortality rates are outcomes, or energy, the environment, national defense and security and the like. There are a number of possible economic outcomes including growth in income, sales, jobs, or cost savings. Often multiple outcomes are demonstrated, with sector and economic outcomes assessed most frequently.

Context

Characterizing, measuring and understanding the context in which outcomes of team science are determined is very important. If the evaluation is to inform improving the selection, management, and implementation of team science, the variety of circumstances requires looking at context to be able to link outcomes to characteristics of teams and policy instruments. For evaluation to demonstrate outcomes only, looking at context is needed to have a plausible explanation of how the program or teams contributed to outcomes because external influences also drive or restrain success

The literature on system level evaluation suggests looking at three levels, micro (individual or team or organization), the meso or sector that is the area for application (such as health, cancer prevention, energy, energy efficiency), and macro level which includes broadly influential institutions and rules (such as property rights, the banking system, free market vs. central control). The meso/sector level is particularly useful because it can serve as a link between the micro and macro levels. Effects on the healthcare system are linked to health research and availability of healthy labor force, for example. Also there are major differences in R&D investment rates and time to market in sectors. For example, time to market for information technologies is short, and for energy technologies it is longer. See Figure 4 for more detail on what might be in each of these categories.

Putting it all together

Figure 4 summarizes the recommended categories of indicators for assessing the outcomes and effectiveness of team science. Often it helps to see the big picture by showing all the outcomes and accompanying categories of indicators that help define those on a single page.

Figure 4. Logical Framework of Indicator Categories for Assessing Outcomes and Effectiveness of Team Science

Inputs	Activities/ Outputs	Interactions	Outcomes: Near Term / Mid Term	Outcomes: Long term
-Funds -Staff/Team quality -Instruments -Knowledge base -Technical base -Research environment	A. ACTIVITIES -plan -investigate -prove concept - prototype B. OUTPUTS 1. Ideas/ Knowledge advances (Excellence, Publications, tech reports, IP, awards) 2. New research tools, techniques 3. People trained 4. Preparation for transition to application [Productivity]	C. CONNECTEDNESS -With other scientists (pre-development) -Across functions with developers, manufacturers, marketing -Inter-sectoral -With intermediaries - With potential application users D. Level of integration (co-located, boundary spanners, etc.) [Indicates influence]	E. SCIENCE OUTCOMES 1. Research activity "performance" 2. Research Agility 3. Organization, integration of knowledge 4. Impact on science -Change state of the art, emerging fields, … 5. Change in science infrastructure 5a. Knowledge Base 5b. Tools, Facilities 5c. People, talent F. APPLICATION OUTCOMES (potential and actual): 1. Industry: new product, process, service 2. Government: policy, program 3. Tech. Infrastructure: standards, generic technology G. ADOPTION INFRASTRUCTRE (potential and actual): 1. Business: distribution channel, logistics, training, etc. 2. Government procurement 3. Public: new media campaign, Advocacy group [Application, Absorptive capacity]	H. VALUE OF THOSE APPLICATIONS: Economic -general -business -energy Social -health -environment -security -other

Context

Micro		*Meso/Sector*		*Macro*
Characteristics of the team (size, diversity, organizational/ management, readiness, etc.)	Nature of the research problem a. research type b. radicalness c. scope	Characteristics of Interactions: a. diversity b. continuity c. mechanism used	Nature of the application of research: a. Breadth b. Timing c. Radicalness of change for application d. Sector speed for technical change e. Sector absorptive capacity, resources	Availability of: -Capital -Capabilities (people, instruments) -Ease of coordination

7. How Outputs and Outcomes Might Differ by Profile

Research profiles were defined based on objectives and scope, so outcomes must differ by research profile. Those differences need to be defined and assessed in the near term in order to learn what and where to make changes that can improve future outcomes, to the extent these are reflected in characteristics of the team and the team's environment.

What is offered here as a starting point for further discussion are differences by profile in six output and outcome areas. The six categories of outcomes were suggested by work Feller and Gamota (2003) during the time when the Program Assessment Rating Tool (PART) was a requirement. They called this set of metrics a "Powerball". The last four of the six are commonly used categories in research evaluation.

1. **Research Performance –** what is observable at the time the research is being performed.
2. **Research Agility –** the flexibility to change course as new learning occurs, balanced with the need for sustained support over a period of time.
3. **Structure/organization of Knowledge –** the structure of the base of knowledge (such as fields) and the way that research is carried out, talked about and planned.
4. **Science Impacts –** advance in knowledge, change in the state of the art, and effects on other research.
5. **Science Infrastructure –** effects on the knowledge pool, the technical base (such as research tools and facilities), and scientifically trained people.
6. **Societal Impact –** effects outside of the science community, on policy or industry or public groups as reflected in behaviors and actions, new products or services.

Table 1 shows how each of these might manifest itself in each of the four [profiles. This is the result of a brainstorming session with faculty with expertise in organizational development, but this has not been published, reviewed or tested. These have some face validity, however, so may be useful to begin a discussion and further work. Here are two of the points of the several points about differences.

- When analyzing performance of a team in the current moment looking neither backward nor forward, the work of large teams would range from high risk with potential for leaps forward with high reward to organized projects making steady progress. For narrow scope teams it could range from a portfolio of high risk projects to a portfolio of mixed projects notable for the high quality of researchers involved.

- The ability to adjust research plans to recent findings in research could range between following an emerging field and consolidating knowledge in an existing field.

Table 1 also begins to examine differences in the timing and breadth of applications from research depending on profile. What seems to be obvious is that incremental research has a shorter time for application than research aiming at radical change. Research with a broad systemic scope of focus aims to affect a larger system, so is likely to have broad system wide outcomes.

Table 1. Success Indicators by Research Profile

Research Profile	Indicators Differing By Profile			Likely Timing, Breadth of Applications
	Outputs	Connected-ness	Near- or Mid Term and Societal/End Outcomes	
Large scope, radical aims	1- Projects have high yield; expected high yield 2- Identified applications for knowledge	Many Diverse (fields, function, geography, organizations)	3- Rapidly deploying activities; strategic coalitions 4- Radically new product or process 5- Converge on theory/project aimed at technical need 6- Influenced public/private sector R&D/outputs	Long time lapse Broader
Large scope, normal aims	1- Organized projects making steady progress 2- Standardized knowledge or language	Many Specialized, Targeted	3- Correct diagnosis of the challenge 4- Incrementally improved product/process 5- Access to, utilization of facilities 6- New standards for quality, reduced harms	Near term Broad or not depending on case
Small scope, radical aims	1- Portfolio of highly unusual projects 2- Change the way people think and ask	Few Diverse (fields, function, geography, organizations)	3- Uncoordinated activities/Emerging fields 4- Radically new idea or prototype 5- An expanding portfolio, risk 6- International thought leadership	Long time lapse Broader

Small scope, normal aims	1- Great contributors participating 2- Develop common language/ teachable points	Few Specialized Targeted	3- Coordinated activities/Revised textbooks 4- Incrementally new idea or prototype 5- Facilitated workshops, colloquia 6- Ideas seeded, awareness fostered	Near term Broad or not depending on case
Key: 1. Research Activity Performance 2. Research Agility 3. 3. Structure/Organization of Knowledge		4. Science Impacts 5. Science Infrastructure 6. Societal Impacts		

8. Mixed Methods Required

The challenge in demonstrating research outcomes is to have reasonable expectations about the performance of the research program for the timeframe within which the program activities are being assessed and to use multiple methods of assessment. The complexity of the questions asked and the variety of indicators need to answer those convincingly suggests mixed methods. Bibliometric analysis and expert judgment need to be supplemented with other quantitative and qualitative methods. Here are a few possible methods, but for more detail see the U.S. DOE guide on R&D Evaluation Methods (Ruegg and Jordan 2006) and the recent handbook on evaluation that uses R&D examples throughout (Link and Vonortas, 2013).

- Collection of routine administrative data including historical expenditures, changes in team members, research objectives, collaborators and their inputs to the research.
- Annual or end of project reports from principal investigators with requirements around reporting progress and which follow a format that makes contextual data mining possible.
- Tracking of Technology Readiness Levels, where appropriate, over time or use of the Stage Gate process for decisions on moving to the next stage of development. Similar readiness scales could possibly be developed for science and for applications such as policy development or service sector take up.
- Social network analysis is potentially very useful for seeing changes in collaboration patterns over time and for identifying boundary spanners and bridging organizations or needs for those. Defining networks can require researcher time, and additional data is needed to know just what is being shared across links or produced by collaboration.

- Mixed method case studies are a primary method and could include use of administrative data, secondary statistics such as number of Ph.Ds. graduated in a field, interviews with various experts, surveys if the population is large enough, and econometric modeling.
- Studies like historical tracing could be done as the transition to application is proceeding. These could document and time stamp movement along pathways such as the movement of research into clinical practice and the effects of that change on technology and techniques used, training required, patient reactions and the like.

9. Areas for Further Discussion and Research

As mentioned in the first paragraph of the introduction, everything about this proposed logical framework for evaluating the outcomes and effectiveness of team science is a candidate for further discussion and work. It is possible that the proposed logical framework, or an improved version of it, could provide the architecture for a roadmap for research on the science of team science. One of the first things done in research is characterization of the object of the research. Another is to develop ways of measuring what is of interest. Yet another is to increase understanding by studying how things evolve over time within their environment and what factors influence evolution or radical change.

The science of team science has concentrated to date mostly on study of the inputs, outputs, and interactions, working on characterization of these and understanding. The proposed framework would broaden that to include different research profiles, additional science outcomes, more attention to pathways of applications that logically lead to societal outcomes, and characterization of the application and adoption area context. As those studying the science of team science have noted, another area of research is development of measurement tools including data collection systems. Finally, theory building requires development of research designs to apply the new tools in ways that find patterns amidst the tremendous variation that is in the system.

References

Feller, Irwin, George Gamota, and William Valdez. "Developing science indicators for basic science offices within mission agencies." *Research Evaluation* 12.1 (2003): 71-79.

Frank, Cyril, and Edward Nason. "Health research: measuring the social, health and economic benefits." *Canadian Medical Association Journal* 180.5 (2009): 528-534.

Hage, Jerry, G.B. Jordan and J. Mote. 2007. A Theories-Based Innovation Systems Framework for Evaluating Diverse Portfolios of Research: Part Two - Macro Indicators and Policy Interventions. Science and Public Policy, 34(10): 731-741.

Jordan, Gretchen B. 2013. Logic Modeling: A Tool for Designing Program Evaluations, in *Handbook on the Theory and Practice of Program Evaluation*, Albert N. Link and Nicholas S. Vonortas, Editors, Edward Elgar Publishing, April.

Jordan, G.B., J. Mote, J. Hage. 2012. Research Profiles: Prolegomena to a New Perspective on Innovation Management, in "Technological, Managerial and Organizational Core Competencies: Dynamic Innovation and Sustainable Development," Eds. F.S. Nobre, D.S. Walker, and R.J. Harris, Business Science Reference (IGI Global), USA.

Jordan, G. 2010. A Theory-Based Logic Model for Innovation Policy and Evaluation, Research Evaluation, 19(4), October 2010, 263-274.

Jordan, G.B. and L.D. Streit. 2003. Recognizing the Competing Values in Science and Technology Organizations: Implications for Evaluation. In *Learning From Science and Technology Policy Evaluation*, P. Shapira, Philip and S. Kuhlman (eds.). Edward Elgar: Cheltenham, UK.

Jordan, G. B., Hage, J., & Mote, J. 2008. A theories-based systemic framework for evaluating diverse portfolios of scientific work, part 1: Micro and meso indicators. In C.L.S. Coryn & Michael Scriven (Eds.), *Reforming the evaluation of research. New Directions for Evaluation, 118*, 7–24.

Jordan, G.B. 2006. Factors Influencing Advances in Basic and Applied Research: Variation Due to Diversity in Research Profiles. In *Innovation, Science, and Institutional Change: A Handbook of Research*, J. Hage and M. Meeus (eds). Oxford University Press: Oxford, 173-195.

Mote, J., G. Jordan and J. Hage. 2007. Measuring Radical Innovation in Real Time. International Journal of Technology, Policy and Management, 7 (4): 355-377.

Reed, John H, G. Jordan. 2007. Using Systems Theory and Logic Models to Define Integrated Outcomes and Performance Measures in Multi-program Settings, in *Research Evaluation*, Volume 16 Number 3 September.

Ruegg, Rosalie, and Gretchen Jordan, 2007. Overview of Evaluation Methods for R&D Programs, U.S. DOE, March.

Shenhar, Aaron J. "One size does not fit all projects: exploring classical contingency domains." *Management Science* 47.3 (2001): 394-414.

Stokols D. Toward a science of transdisciplinary action research . American Journal of Community Psychology. 2006 ; 38 (1): 63 – 77 .

Stokols D , Hall KL , Taylor B , Moser RP. The science of team science: overview of the fi eld and introduction to the supplement . American Journal of Preventive Medicine. 2008 ; 35 (2 suppl): S77 – S89 .

Stokols D , Fuqua J , Gress J , Harvey R , Phillips K , Baezconde-Garbanati L , et al. Evaluating transdisciplinary science . Nicotine & Tobacco Research. 2003 ; 5 (suppl 1): S21 – S39 .

Stokols D , Hall K , Moser R , Feng A , Misra S , Taylor B. Evaluating crossdisciplinary team science initiatives: conceptual, methodological, and translational perspectives . In: Frodeman R , Klein J , Mitcham C , eds. Oxford Handbook on Interdisciplinarity . New York : Oxford University Press;2010 : 471 – 493 .

Stokols D , Misra S , Hall K , Taylor B , Moser R. The ecology of team science: understanding contextual infl uences on transdisciplinary collaboration . American Journal of Preventive Medicine. 2008 ; 35 (2 suppl): 96 – 115 .

Tassey, Gregory. *The technology imperative*. Edward Elgar Publishing, 2007

Winter, Susan J., and Nicholas Berente. "A commentary on the pluralistic goals, logics of action, and institutional contexts of translational team science." *Translational Behavioral Medicine* 2 (2012): 441-445.