Learning Science in Informal Environments
Commissioned Paper

Evaluation of learning in informal learning environments

Institute for Learning Innovation

July 21, 2007

I. Overview
In 2006, the National Research Council initiated a study on Learning Science in Informal Environments. The purpose of the study is to synthesize a range of relevant literatures and recommend strategic directions for future research in the area. In the course of working on this study the Committee has found one of its challenges to be the identification and assessment of evaluation studies of informal science programs, in particular those which have probed science learning outcomes. To that end they commissioned the Institute for Learning Innovation to produce a paper that would help them discern the state of evaluation practice, the range of methods used to assess learning in these settings, and the quality and strength of the evidence for such learning. This paper will be used by the Committee as they address the issue of evaluation in their report and make specific recommendations to their sponsor, the National Science Foundation, and the field of science learning in informal environments at large about how to best support and shape quality evaluation practice (questions, methods, etc.) in the future. In addition, the paper addresses the contribution that evaluation has made to the field's understanding of the impacts of informal science learning experiences and suggests how evaluation practice can be enhanced in the future.

The Institute was asked to address the following specific questions:

§ Definitions: How is learning defined in the evaluation of science learning in and from informal environments? How is evaluation defined, and in particular is this definition the same or distinct from a definition of research?

§ Methods: What methods (design, techniques, units of analysis), measures (standardized and evaluation specific), and research questions are "typical" in the current evaluation of Learning Science in Informal Environments (LSIE) outcomes? What unique challenges and opportunities do informal learning environments present to evaluation and what are the characteristics of appropriate methods and measures for learning in LSIE evaluations? How has research informed what methods are appropriate?

§ Findings: What do the findings across a range of evaluation studies indicate about what and how much is learned in informal science learning environments vis-à-vis a range of learned proficiencies/outcomes? Is there convergence in the findings that would support a theoretical model of evaluation? Do any innovative evaluation strategies stand out as particularly useful in the LSIE context?

- § Outcomes: How can ~~program~~ evaluation findings contribute to the understanding of how people benefit from experiences in informal science learning environments?
- § Challenges and opportunities: What are some of the major changes/advances in the field of informal science learning environments evaluation and where is the field heading?
- § Recommendations: Based on this analysis, what is the future outlook for research/evaluation practice in the Learning Science in Informal Environments field? What would we recommend as some promising directions?

II. Defining Informal Science Learning (ISL)

> *If you don't know where you're going, any road will get you there.*
> Flying Karamazov Brothers

In the late 1980s Mary Ellen Munley, a leader in museum education, wrote a paper in which she argued persuasively about the importance of asking the right question(s), suggesting that evaluators could benefit greatly from the insights of the Flying Karamazov Brothers (Munley, 1986, 1987). Although tongue in cheek, Munley's point was well taken—since evaluation is a process and a tool with which to understand one's intentions and accomplishments it is critical to set out having some idea of what one is actually attempting to accomplish and ultimately evaluate (appreciating that it is also important to be open to unintended consequences which we will explore at a later point in this paper). Since this Committee's charge is to explore *learning science in informal environments* it seems only appropriate to try to define what is meant by that term in as clear and consistent a manner as possible for the purposes of this paper.

However, as the Committee has come to appreciate defining this term is no small feat, something we have also appreciated and thought about a great deal. First, defining learning generally is challenging under any circumstances since it is simultaneously a process and a product, a verb and a noun (Falk & Dierking, 1995). Even social scientists investigating learning, including learning researchers, psychologists, and sociologists, have difficulty agreeing on a single definition (e.g., Bransford, 1979 - Bransford, J., Brown, A., & Cocking, R. (2001); Churchland, 1986) and for the most part they have investigated learning under the narrowly defined compulsory circumstances of formal education and vocational training. The task of defining learning, in terms of the what, where, how, why and with whom becomes even more difficult in informal learning environments, for example, when individuals visit a museum, watch documentaries on television, play an interactive web-based game, participate in an after school science program, read articles and books, etc.

However, there is another issue complicating the task. As the National Association of Research in Science Teaching (NARST)'s Ad Hoc Committee on Informal Science Education noted in its 2003 policy statement (Dierking, Falk, Rennie, Anderson, & Ellenbogen, 2003), although *informal learning* is the most commonly applied term for the science learning that occurs outside of the traditional, formal schooling realm (pre-

college, university and advanced degrees), the term has significant limitations because it artificially delimits efforts to describe the type of real-world learning that humans engage in daily; learning that occurs across a broad spatial and temporal context, both inside and outside of schooling. The term focuses on the attribute of where learning occurs ("informal"), rather than describing the nature of the learning. If one focuses on the *nature or processes* of learning, one appreciates it as a biological or ontological process in which the process of learning is independent of the context in which it occurs, or as some have stated, including us: "Learning is learning, no matter where and how it occurs."

Complicating matters even more are the many terms used to describe "informal learning" in different communities of practice or contexts of learning (i.e., environmental education or international development). For instance, The Encyclopedia of Informal Education (2006) defines and differentiates between formal education/learning, informal education/learning and non-formal education/learning. Interestingly, although museums have traditionally referred to themselves as "informal education" institutions, by the criteria of this framework they are actually "non-formal" education institutions (the museum sets a learning agenda and determines the learning outcomes, even if those are not met). Complicating matters further, the term "informal learning" also carries different meanings in different academic and professional contexts. Ultimately it can be argued that the term "informal learning", and thus the term "informal science learning," while widely used, is confusing and ill-defined.

In an effort to acknowledge these definitional inconsistencies and try to arrive at some common language, representatives from over two dozen federal agencies, nonprofit professional organizations, and not-for-profit organizations met throughout 2006 to discuss and come to agreement on some common definitions of terminology typically used by interpreters, environmental educators, historians, and others in informal settings such as parks, aquariums, zoos, nature centers, historic sites, and museums. Called the Definitions Project, this effort was initially funded by the U.S. Environmental Protection Agency and also organized by the National Association for Interpretation (NAI), in cooperation with the U.S. Fish and Wildlife Service and the Institute for Learning Innovation. The ultimate goal of the project was to create a sense of community amongst a range of organizations that, while sharing many goals and approaches to learning and education, traditionally have worked in isolation from one another. There is now a prototype website which this Committee and the field may find useful (http://www.naimembers.com/definitions/index.cfm). The website is not considered "public" as of yet but project staff are welcoming visitors and further input from the field. The long-term plan for the website and glossary, a major product of the effort, is that it will continue to grow each year as new terms and examples of existing terms are added by a review panel staffed by agencies and organizations related to the field of informal education and interpretation.

Based on this quick analysis of terms, our recommendation would be to follow those of the NARST Ad Hoc Committee and the Definitions project. Both groups felt that it was critical to separate the setting from the definition of learning occurring in the setting,

recognizing that many classrooms can be considered informal environments or contain informal elements and that not all experiences in informal science institutions are self-directed or free choice. The NARST Ad Hoc Committee also felt that there is a need for a more appropriate moniker for this type of learning and this area of study. Many possibilities were suggested, (e.g., self-directed, free-choice, lifelong, public understanding of science, etc.). Although the group ultimately did not choose a term, there was unanimous agreement among members of the Ad Hoc Committee that the term for *learning* in these settings should not be "informal science learning."[1] Instead they argued for a learning term that would reflect the nature of such learning including characteristics such as self-motivated/self-directed, voluntary, personal, on-going, learner-centered, guided by the learner's needs and interests, contextually relevant, collaborative, non-linear, open-ended, and with a high degree of choice as to what, when, how, where and with whom to learn (Griffin, 1998; Falk & Dierking, 2000). We feel that the definition should be even broader and include other disciplinary approaches such as anthropology and sociology so that discourse, networks, communities of practice, apprenticeships, gaming, serious leisure, the study of hobbies and so on would also be included in a perspective on learning outside of schools and formal education and training.

Thus we would recommend framing this study as being about *learning science in informal environments* (LSIE) rather than *informal science learning* (ISL). Although this may seem like a subtle distinction, we feel it is an important one, and we refer to LSIE instead of ISL throughout the chapter. . The term LSIE also aligns well with the approach we believe the Committee has taken which includes learning that occurs in various venues or configurations such as museums & designed spaces, family and everyday learning, and after/out of school and adult programs.
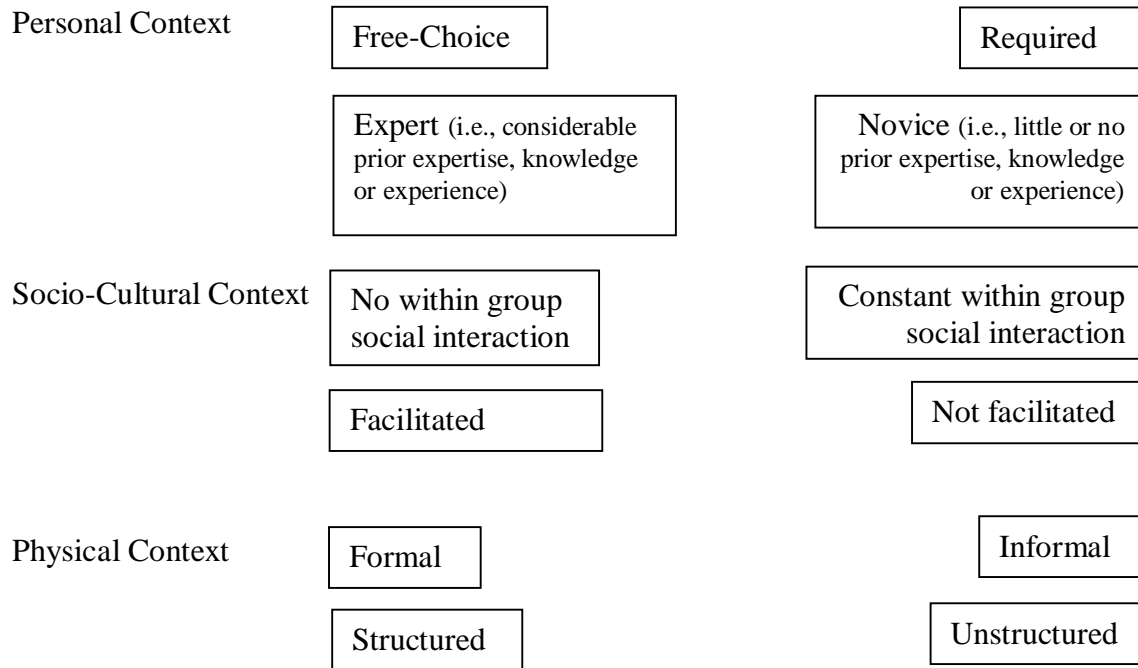
The Institute for Learning Innovation has also been trying to expand our framework for conceptualizing the complexity of learning in informal environments. In a study paper for the Consensus study, Falk, Dierking & Storksdieck 2005) suggested an epistemology for the nature of learning (and thereby factors that influence the success of a learning process) which includes a minimum of six dichotomies, two from each of the contexts of the Contextual Model of Learning discussed more fully later in the chapter (See Figure 1).

This epistemology represents a small selection of the Contextual Model of Learning dimensions, but these were ones that in recent research emerged empirically as the most important (Falk & Storksdieck, 2005). The goal is to provide a simplified version of a potentially more comprehensive framework for researching such learning, one that appreciates the complexity and multi-dimensional nature of LSIE. Each of the six dichotomies represents a complex continuum of variables. For instance, "free-choice vs. required" encompasses not only the conditions under which the learning is occurring, but also important motivational aspects such as the degree of intrinsic versus extrinsic motivation or the degree to which the learner is intentionally choosing to learn about a

---

[1] They did concur that "informal science education" was an appropriate term to describe the informal science enterprise and field.

specific aspect of science. Interestingly, Kit Klein and her former colleagues at the Center for Informal Science Teaching and Learning (CISTL) have explored a similar model though it contains many more variables (Klein, Astor-Jack, Jordan, Addelson, Rowe & Kassing, 2005).

Personal Context	Free-Choice	Required
	Expert (i.e., considerable prior expertise, knowledge or experience)	Novice (i.e., little or no prior expertise, knowledge or experience)
Socio-Cultural Context	No within group social interaction	Constant within group social interaction
	Facilitated	Not facilitated
Physical Context	Formal	Informal
	Structured	Unstructured

Figure 1. Potential Epistemology for Lifelong Science Learning

Understanding the purposes and agency for LSIE may help in teasing out its essential elements. For example, a learners learning objectives in informal science learning settings can range from seeking out personally relevant information (for instance, researching specific information on the Internet) to identity formation and personal transformation. Thus agency it is an important distinguishing characteristic when defining formal, non-formal or informal/free-choice contexts of learning: Does the venue have a teaching agenda? Does the learner have a learning agenda? Who determines what is learned? And how structured is the learning/teaching process?

There are some other issues to consider also. As a growing number of researchers choose to try and understand how people come to learn about science, increasingly they appreciate that learning rarely, if ever, occurs and develops from a single experience. Rather, learning in general, and science learning in particular, is cumulative, emerging over time through myriad human experiences, including, but not limited to experiences in museums, schools, while watching television, reading newspapers and books, conversing with friends and family, and increasingly frequently, through interactions with the Internet. The experiences children and adults have in these various situations dynamically interact to influence the ways individuals construct their scientific knowledge, attitudes, behaviors and understanding. In this view, learning is an organic,

dynamic, never-ending, and quite holistic phenomenon of constructing personal meaning. This broad view of learning recognizes that much of what people come to know about the world, including the world of science content and process, derive from real world experiences within a diversity of appropriate physical and social contexts, motivated by an intrinsic desire to learn.

Historically much of the research on science learning in informal environments has occurred within museum-like settings. Learning from museums, or other comparable educational institutions (zoos, aquariums, nature centers, etc.), actually represents a nominal number of the situations in which this type of learning occurs. However, the growing body of research on learning in and from museum-like settings provides an important baseline of understanding about learning in such free-choice situations. Still lacking, though, are comparable studies of learning from film, radio, community-based organizations like scouts, and summer camps, home and friends, the work place, the Internet and a whole range of other real-world situations.

III. Defining Evaluation

As suggested earlier, evaluation is a process and a tool with which to understand one's intentions and efforts to accomplish those intentions. Ideally evaluation is conducted to ensure that programs are on track and successful and is a key to effective practice, in other words evaluation is the flipside of good planning. During crucial stages of program development, effective evaluation practice documents or measures achievements or outcomes against intended goals and objectives (while also being open to unanticipated outcomes as well). Evaluation is thus not merely a tool to assess program outcomes and document success or failure; it also ideally is a process which contributes to decision-making at key points of project development and implementation and can be used to ensure success throughout the process of project development. In order to serve such a broad purpose, evaluation research ideally is integrated fully into project development defining such important project milestones as:

- § Identifying target audiences (e.g., general museum visitor, urban youth, schoolchildren, decision-makers in the community);
- § Defining realistic goals and objectives for each specific target audience;
- § Identifying all important stakeholders in the project and involving them in the process by asking about their needs and perspectives;
- § Creating programs and program elements that address each objective for each of the target audiences (if there are more than one);
- § Avoiding programs and program elements that cannot be mapped to an objective or goal;
- § Testing all intermediary steps and products to ensure their efficacy;
- § Documenting the degree to which the program has achieved its goals and objectives for each of the previously defined target audiences.

In the ISE field it is customary to identify four stages of evaluation. The four stages are:

(a) Front-end evaluation. These provide input to decisions about how to develop a program in advance of the planning stage. Generally it provides background information for future project planning. It typically is designed to determine an audience's general knowledge, questions, expectations, experiences, learning styles and concerns regarding a topic or theme.

(b) Formative evaluation. These provide information to improve the program during the design and development stage. Formative evaluation studies typically provide information about how an interpretive media or program can be improved and occur while a project is under development. It is a process of systematically checking assumptions and products in order to make changes that improve design or implementation.

(c) Remedial evaluation. This form of evaluation provides information to improve a project during implementation and allows for corrections once projects are underway. Remedial Evaluation is the assessment of how all the individual parts of an interpretive media or program work together as a whole; like formative evaluation the goal of remedial evaluation is to improve educational effectiveness and insure achievement of goals and objectives.

(d) Summative evaluations. These assess outcomes or impacts of a "settled" project. Summative evaluation is conducted after an interpretative media or program is completed and provides information about the impact of that project; what is assessed should be tied to project goals and objectives, however there should be an effort to document unintended outcomes also.

Unfortunately evaluation as often described and associated only with written surveys or interviewers armed with clipboards. However, effective evaluation uses the entire spectrum of educational and social science research methodologies, including surveys, face-to-face and phone interviews, observations, focus groups, videotaping, audio-taping, and so on. Ideally a mixed methodology is utilized so that data can be triangulated to produce robust results.

There is an important role that summative evaluation can play in enabling "reflective practice" and institutional learning but there are some barriers which will be discussed further in the scan of the field. For although summative evaluation can be a tool to inform the field and world of the consequences of efforts, the nature of proprietary information, the lack of publishing avenues (though that is changing), and competing websites (Online Evaluation Research Library and informalscience.org) can actually impede dissemination and progress.

Although summative evaluation and research often use similar approaches and tools, there are distinct differences particularly in terms of their goals and purposes. Summative evaluation studies tend to ask *whether* and *to what degree* specific project-oriented goals have been met, documents unanticipated outcomes and ideally also addresses what contributed to success or failure; ideally summative evaluation is rigorous project- specific research but at times it does use ad-hoc methods to address a range of project objectives. Although not necessarily the case, the evidence bar can be lower in evaluation since the goal is to demonstrate achievement to a smaller, more homogenous

audience (for instance, funder and project staff); research findings need to convince an entire professional community. Different disciplines have developed their own unique rules for determining what is considered acceptable research as well as what are acceptable questions, tools and methods. In recent years there has been an increasing cross-fertilization of methods and approaches across disciplines.

IV. A Scan of Current ISE Evaluation Practice

Based on discussions with the working group, we analyzed the 82 studies that currently are posted on informalscience.org. We also sent out an invitation along with an evaluation practice matrix to several colleagues (Randi Korn, Beverly Serrell, Minda Borun, Jeff Hayward, Joshua Gutwill, Lorrie Beaumont, Barbara Flagg, Carey Tisdal, Steven Yalowitz, Wendy Meluch, George Hein, Kit Klein, Deborah Perry & Saul Rockman) requesting that they send us sample studies and/or use the matrix to "code" their study. The evaluation matrix includes criteria such as the underlying theoretical framework of the study (if any, and why chosen); target audiences, methodologies, methods, and major outcomes, broadly defined (A copy of the matrix is in Appendix A).

We received feedback and sample studies from all these evaluators. In addition we surveyed Institute researchers for their comments on the matrix and evaluation studies they felt were good examples of ISE studies. Based on feedback from the working group, we also scanned the afterschool/family/community involvement and youth development arenas, reviewing approximately 10 studies in those categories. In the case of the afterschool/family/ community involvement and youth development arenas, not all projects evaluated were science-focused, but the studies represent useful windows into the approaches used by related fields to evaluate their efforts so they were included. Also where relevant, particularly in the areas of theoretical frameworks and outcomes measured, we have incorporated data from a recently completed study for the Institute for Museum & Library Services, *Engaging America's Youth* (Koke and Dierking, 2007)*, focused on museum and library involvement in youth development efforts (n=247).

We should mention one caveat before discussing the analysis. One researcher/evaluator cautioned us against referring to the studies we gathered as *representative*, preferring instead a label such as *selected* and we thought the point was well-taken. *Representative* implies that they are in some way typical or an exemplar of some larger construct, but pointed out that s/he would hate to think that the two studies s/he shared were considered representative of anything other than being studies which were chosen because they have some important and interesting characteristics. As you know we did ask these evaluators/researchers to send us a study or two that they felt represented their work but that notion of representative is quite different than its usual use, particularly when discussing research. Some of the questions addressed in this section include:

- What unique challenges and opportunities do informal learning environments present to evaluation?
- What methods (design, techniques, units of analysis), measures (standardized and evaluation specific), and research questions are "typical" in the current evaluation of LSIE learning outcomes? What are the characteristics of appropriate methods

and measures for learning in LSIE evaluations? How has research informed what methods are appropriate?

- What do the findings across a range of evaluation studies indicate about what and how much is learned in informal science learning environments vis-à-vis a range of learned proficiencies/outcomes? Is there convergence in the findings that would support a theoretical model of evaluation? Do any innovative evaluation strategies stand out as particularly useful in the LSIE context?

(a) Evaluation Studies Posted on informalscience.org

More than half of the 82 evaluation studies posted on informalscience.org are summative studies (47) with one of those being a blended research/evaluation study and more than a quarter (28%; n=29) are front-end studies (one of these was an extensive literature review to guide a new initiative and one was described as a front-end study which would guide the renovation of an exhibition so it had a formative component also). The few additional studies include 3 formative studies and 3 studies that though called summative were efforts to evaluate an exhibition or interpretive approach before it was going to be re-worked so they also had a formative aspect. In fact one evaluator we contacted suggested that with all the renovation going on in museums they thought that more of these kinds of studies would be helpful since "I think many institutions will need these types of studies in the future."

Currently almost three-quarters of the studies posted (61) focus on exhibitions, with five of those focusing on exhibitions with strong program or web site components. Sixteen of the studies are media projects with almost all of those (14) being television programs, or evaluations of activity guides or web sites that were products of a television program. One of the media evaluation studies focuses on a large format film and accompanying activities and one was for a training web site for professionals. Interestingly, though exhibitions often include similar components such as activities or web sites, their evaluation studies either are comprehensive and include an evaluation of all aspects of the project in one summative or solely focus on the exhibition. It was more common to see individual evaluation studies for the different components of a project (television show, activity guide, collaboration component and so on) in the media arena. It was also in the media arena where three of the projects were specifically designed for the professional audience, reflecting the new NSF guidelines initiated in 2004.

Important to note though is that out of the 82 posted studies, only 4 are program evaluation studies (one for a one-day teacher workshop focused on inquiry). Although there are four additional studies that focus on exhibitions with integrated programming or in one case an exhibition with an associated intern program (counted above), this is the extent of posted program evaluation studies. Since most of the posting is voluntary and consistently done by only a few evaluators who tend to focus on exhibition evaluation, this is not a surprising observation, but the NSF may want to be more assertive about studies being posted once they are completed and may want to be proactive in terms of the types of studies (front-end, formative, summative) and their focus (exhibitions, programs, media, etc.) that are included.

(b) In-Depth Analysis

Approximately 32 studies were analyzed in depth, 20 of them were ISE, NIH or IMLS-funded and 12 were studies in the afterschool/family/community involvement and youth development fields. Seven (7) of these studies were front-end, and with the exception of one, which was a front-end evaluation of a trail in a national park, all were studies focused on exhibitions. Given that the term "front-end study" was coined by Chandler Screven specifically related to the exhibition development process this is not surprising. Although we did not review any, it is common in the afterschool/family/community involvement and youth development fields to conduct needs assessments which are similar in concept to front-end studies.

The remaining 25 studies were summative evaluations. Although there was more variety among these studies, in terms of what was evaluated summatively, including a citizen science program, 2 activity guides associated with media projects, a television program, a teacher workshop, on-line and face-to-face training for afterschool providers, 4 family support & early/elementary childhood intervention programs providing an array of social, educational, and health services to parents and preschool-aged children and 5 afterschool and summer programs for upper elementary and older youth, the majority of the projects, particularly ISE-funded projects, were exhibitions (our invitation to colleagues in no way specified exhibitions). According to David Ucko, Program Officer, community and youth awards currently make up about 15% of the ISE portfolio of projects, both in terms of numbers and budgets, which is a growing number.

We only had budget sizes for 22 of the evaluation studies we analyzed. Front-end studies were most typically in the range of $10-$25,000 and summative evaluation studies were typically in the $25-$50,000 range. Audiences evaluated in both front-end and summative studies of exhibitions were most often casual visitors including children in families or young children and caretakers, but often specifically stating "not children in school groups." Audiences for many media projects and most community-based projects, both those with an ISE focus and those in the wider afterschool/family/community involvement and youth development fields, included children/leaders from afterschool programs or community-based organizations and most of these evaluation studies focused on families and communities described as underserved.

Only a few of the exhibition projects, whether front-end or summative, include theoretical frameworks. The theoretical frameworks identified in about six of the external studies were constructivism, misconceptions and naturalistic inquiry. The Institute customarily uses the Contextual Model of Learning as an organizing framework for many of its evaluation studies, ensuring that the personal, sociocultural and physical contexts of the experience are considered and when possible included as independent variables. No media evaluations included theoretical frameworks. As we will discuss in more depth in Section V of this chapter, the lack of theoretical frameworks in evaluation studies of exhibitions is likely due to the fact that many of these projects are not conceptualized from the outset in terms of such frameworks but it is important to note. There is some effort to include theoretical frameworks in some of the ISE community-

based efforts, primarily positive youth development concepts, Community of Practice and activity theory models and some specific frameworks depending upon the focus of the project, for instance, researchers at the Institute have used the Ecologies of Parent Engagement model proposed by Calabrese Barton, Drake, Gustavo Perez, St. Louis & George (2004) to guide a program evaluation for a .parent involvement project funded by ISE.

In terms of design and methods, there was very little variation, particularly among the exhibition projects. Front-end evaluation studies tended to include focus groups, observations (when an exhibition was being renovated) and/or structured or in-depth interviews; in one case a front-end for a renovation of an existing exhibition using a longitudinal web survey demonstrated that learning from a visit was retained over a 4-month period. In terms of the summative studies all (with the exception of one study which was a pre- post-test design), were post-test only designs, utilizing timing and tracking and exit interviews for data collection. In addition to these methods, a few exhibition studies also included cued observations and interviews in which visitors were asked to spend time in galleries and were observed and interviewed after (they spent considerably more time in the exhibition when cued). One colleague we contacted begins the summative evaluation process by creating a matrix to organize what the team wants to find out and to think of as many possible ways to gather data, however, similar methods as described for other studies were utilized in the two examples of studies s/he sent. This approach is similar to developing a Logic Model with project teams, something the Institute now does at the outset of all major projects, ideally at the very first planning meeting. We have found it a useful approach enabling teams to focus on their goals and objectives and to operationalize for the evaluation team what success will look like. We have also used card sorts and drawing activities (particularly with children), both in front-end and summative exhibition studies, and are aware of these as methodologies utilized by others, however, only one of the external studies we reviewed included card sorts and a drawing activity as a method. As budgets and timelines allow, the Institute tries to build longitudinal components into as many of its studies as possible. One external study included web surveys which were administered 1-4 months after the visit, otherwise all data was collected as visitors exited the exhibition.

Only a few of the external exhibition evaluations reviewed identified specific independent variables however, those that were identified included self-reported age, education, occupation and in the case of families, parental report of children's age. Location related to tracking and timing data was also used on occasion as an independent variable. In one case, two different prototypes of trail guides were independent variables. The Institute customarily includes age, gender, race/ethnicity, frequency of visitation or use, membership if appropriate, knowledge of/ interest in the content, level of social interaction, length of stay, crowdedness, and staff/volunteer facilitation, when appropriate, as independent variables in exhibition studies.

In terms of the evaluation of media projects, almost all included pre-post-measures and a few included observations of children or adults watching the media piece or using the activity guide. As stated earlier, in all cases research participants were either part of

intact groups such as a scout group or an afterschool group or were constructed focus groups in which individuals were invited to attend at a scheduled time. The only RCT design in any of the 25 summative evaluation studies we reviewed was a media project in which afterschool leaders were randomly assigned to one of two treatments: participation in an on-line training program solely or participation in an on-line training program with a face-to-face training session also.

In terms of the evaluation of programs there was greater variation in designs and methods, with distinct differences between ISE-focused evaluation studies and those in the wider afterschool/family/community involvement and youth development fields. ISE program evaluations tend to utilize longitudinal pre-post test designs that included structured and open-ended interviews, site visits, observations, focus groups, the creation of cases studies and dependent measures developed and piloted by the evaluators themselves. The Institute has also attempted to use time series designs in on several program evaluation studies, in which impacts are analyzed in terms of the number of experiences engaged in or some combination of possible interventions over time but attrition and small frequencies of participants in some aspects of the program have limited our ability to discern clear patterns.

Program evaluation studies in the wider afterschool/family/community involvement and youth development fields also include longitudinal pre-post test designs, but also on occasion time series designs (for instance, a program focused on young children in which there were child-focused and parent education components in which some research participants engaged and others did not). Case studies and interviews with participating families are also common in the wider afterschool / family /community involvement and youth development evaluations.

There is a distinct difference in dependent measures utilized in evaluation studies in the wider afterschool/family/community involvement and youth development arenas. Most measures are standardized tested instruments and scales (e.g., Learning Accomplishment Profile, GED completion, school readiness-gain scores on Preschool Inventory and Peabody Picture Vocabulary Tests), or in the case of public health projects (physical fitness tests, blood samples, and so on) or other measures. In one project the evaluation compared an analysis of videotaped interactions of parents playing with children prior to participation in the project and then after and also compared skills and competencies of children who participated in the program with those who did not.

Interestingly, our research revealed a great deal of debate in the wider afterschool/ family/community involvement and youth development arenas about both RCT designs and standardized and tested outcome measures (this will be discussed in more detail in Section VII for the LSIE community specifically) . This debate was illuminated well in one project, a program designed to promote child and family health after school (Sass & Blumenthal, 2006). Project staff described an ongoing challenge in identifying appropriate external evaluation partners because of these very issues, a challenge we believe is mirrored in the LSIE community. Because of the focus of the project, project staff wanted to include public health researchers at local universities as part of the

evaluation team to provide feedback on community factors, components, and measurable outcomes, however, in the course of collaborating they have also encountered differences in perspective regarding rigor and program appropriateness. Public health researchers emphasize experimental control and outcome measures, while this program values "kids' choice"—that is, allowing children to participate in activities that interest them, rather than requiring mandatory participation. In addition, some of the common measures in childhood obesity research (e.g., children's weight, body composition) are not consistent with their program's philosophy or model. They are continuing to seek common ground with public health researchers while maintaining their program goals. Clearly, this is an unfolding issue in many disciplines but the bottom line seems to be that RCT is an appropriate approach for some projects and some questions.

Another related issue that is being discussed in the wider afterschool/ family/community involvement and youth development arenas is the desire for researchers, practitioners, and policymakers to move beyond the question of whether programs matter for youth, families and communities given that research suggests that they do, to questions about why, how, and for whom these programs matter and matter most. This is a discussion that to some degree is happening within the ISE field, particularly among those engaged in developing, implementing and evaluating youth and community programs, however, we feel it is a worthwhile conversation for other aspects of the LSIE field and has implications for the design and methods used to evaluate these efforts.

Related to the issue of defining outcomes is the debate about the impact of the outcome-based evaluation movement on current evaluation practice (Figure 2 on the next page shows the outcomes-based logic model). Proponents led by the United Way and social service community, and increasingly federal agencies such as the NSF and Institute for Museum & Library Services (IMLS) point to the benefits of being clear about one's intended outcomes (reminiscent of "If you don't know where you're going, any road will get you there"). In particular, they argue that it is important to distinguish between "outputs"—for instance, what activities are part of the project or how many people participate in those activities, and "outcomes"—what you are trying to accomplish through the activities. Interestingly though, we have been struck that in some of our work outputs actually can be outcomes, a case in point is a IMLS-based partnership project we evaluated five years ago. A science-rich institution was trying to build better relationships with neighboring underserved communities, something they had never done before. A pure outcomes-based approach would say that the partnership and relationship-building was an out put, rather than an outcome of the effort but given the history of the institution vis-à-vis the local community we argued that in this case the creation of the partnership was actually an outcome. Fortunately, the IMLS agreed. We were struck when analyzing common outcomes in the wider afterschool /family/community involvement and youth development arenas that outputs were common outcomes there, for example, attendance and participation of families in programs and activities and the number of parents who completed GED or other educational goals. In very at-risk communities where participation in programming is infrequent, what might typically be called an output may in fact be an outcome.

Those who are wary of this approach suggest that it is easy to focus on the lowest common denominator, outcomes that are easy to measure, rather than broad, less tangible ones. Others also worry that by defining intended outcomes so tightly, it is difficult to be open to the unanticipated outcome, impacts that often are the most interesting of all. There is also concern that such evaluations only focus on whether goals were accomplished, rather than probing how, why and for whom impacts were observed. In our opinion there are important perspectives on both sides. As we suggested it is common for us to engage a team in creating a Logic Model early in the life of a project, however, we also endeavor to design an evaluation in such a way that one can observe unintended outcomes and probe the deeper questions of why, how and for whom.

We also feel that first and foremost, any discussion of outcomes needs to be imbedded within a person's overall experience and daily life. As we will discuss more fully in Section V, people are motivated to engage in these experiences for a multitude of reasons: to spend quality time with family/children/friends, to take part in a culturally enriching activity, to reinforce and shape identity
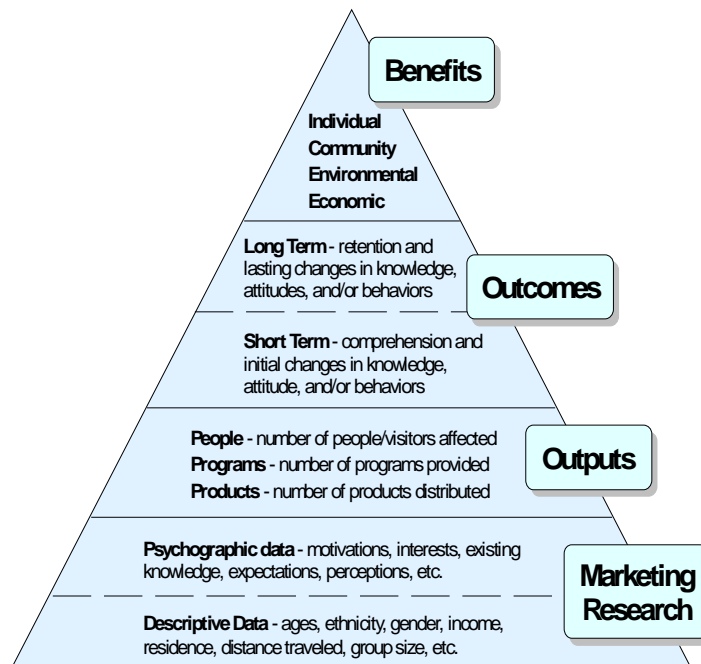
Figure 2. Outcomes/Outputs Pyramid

and to "learn new things." Given these realities, it seems unnecessarily restrictive for the field to define learning outcomes too narrowly. Instead a balance must be struck, since just as a narrow definition of outcomes may fail to document the true impact of museum experiences, a set of outcome variables that is defined too broadly may lead to ambiguous undefined results.

The following represent some criteria and a potential framework for defining outcomes within the social, cultural and historical context of people's lives. Thinking broadly and contextually about the outcomes of these experiences may help ISIs face one of their greatest challenges: how to allocate scarce resources across a potentially limitless set of competing visitor needs and expectations.

Outcomes should be grounded in the role of the institution in the community and situated within the contexts of people's lives. The reasons for what someone experiences and learns, let alone why and how someone engages in such experiences, are inextricably bound to the socio-cultural context in which that experience occurred. As suggested earlier, the development of potential outcomes should as often as possible be guided by the learners themselves: Let categories emerge from ISE experiences as opposed to defining *a priori* what the outcomes of an experience will be. Outcomes can be based on the goals and objectives of a project (and therefore would be closely tied to exhibition and program design), or they can be "unanticipated" and not planned, yet be valuable to the visitor. Unanticipated outcomes can also be defined and measured.

Outcomes need to be aligned with learner expectations. People engage in these activities with prior experiences and hence prior expectations for what the experience will be like (even people who do not traditionally participate in these experiences). These expectations strongly influence the outcome of ISE experiences since motivations and expectations directly affect what people do and what they learn from an experience. When expectations are fulfilled, learning is facilitated. When expectations are unmet or reinforced if they are negative, learning can suffer. Thus it is important for ISE practitioners to understand people's expectations and needs in order to design experiences that facilitate and learning.

Outcomes should be open-ended and flexible. Learning is a highly personal process, and although it is impractical to define outcomes for each and every person, it may be useful to distinguish groups of people and to develop experiences and desired impacts based on such psychographic factors as levels of prior interest and knowledge. Such an approach will enable practitioners to develop multi-layered experiences with resulting multi-layered outcomes but in order to do so they will need to understand their audience and be able to understand the potential multiple outcomes that a range of types of learners might make.

Outcomes should be assessed at different points in time. Short-term outcome measures have long been used to assess the impact of LSIE experiences but have also been criticized for not adequately capturing the true impact of them. We suggest that any discussion of outcomes needs to distinguish whether one has assessed the immediate outcome (when the person has just completed the experience), or whether some longer-term effects are assessed. As a field we also need to better operationalize what long term is—studies with longitudinal

components sometimes assess the learning within days, months or years later but there is no clear understanding of when the best time to do so is. A study teasing out some of these issues is greatly needed.

Outcomes can be expressed at different levels. Learning not only occurs in different domains (i.e., interest, awareness, knowledge, skills, attitudes, etc.), they can also be assessed at different levels. Our research suggests that there are three basic levels at which outcomes can be measured: at the level of the individual, the level of the social group, and the level of the community. Outcomes defined on the level of an individual would answer this guiding question: How is the individual influenced by the experience? However, we can also ask: How is the entire social group with which the individual visited the museum influenced? Did group members learn about one another, reinforce group identity and history? Did they develop new strategies for collaborating together? And finally, we can also define outcomes on the community level: How does the exhibition, media piece or program influence the local community? Realizing the multi-level nature of outcomes is especially relevant today as museums work to document their public value and are more and more frequently setting priorities that include group and community transformation.

Ultimately, defining and prioritizing outcomes requires value judgments. All outcomes are not created equal. In certain contexts, some outcomes will be more appropriate, valuable, or useful than others. In ranking outcomes, ISIs are making value judgments as much as they define their purpose and their contribution to their communities. ISIs are also communicating a strong message: why we are we here! This is their unique contribution and their niche. Thus an outcome should be specific to the ISE experience, and be something that is more optimally accomplished through that experience. This last criterion is particularly important for the continued valuation of ISIs and LSIE outcomes within society. These institutions need to be able to point to their singular contribution to the menu of learning options people can avail themselves of, a notion of complementary learning we will discuss in more detail in Section VI. Ultimately, the discussion about outcomes is a discussion about the contribution of ISIs as cultural institutions in our society. ISIs are not extensions of classrooms, nor are they playgrounds and entertainment centers. Hence, any outcomes used to document their impact should reflect their uniqueness.

V. Typical Science Learning Outcomes

Background Throughout this chapter we have focused fairly exclusively on discussing the processes involved in assessing the impact of LSIE experiences. We have defined evaluation, discussed some of its stages, the influence of outcomes-based evaluation on the field and some of the changes and advances we have been observing. But as we have also suggested, evaluation is merely a tool. Ideally, it is a process that can be used to provide answers to the essential questions of any field. Why are we here? What are we (and even more importantly, those we are trying to reach!) trying to accomplish and take

17

away from LSIE experiences? How do we know whether we have accomplished these goals? Are there unanticipated outcomes that are important to document also?

Before launching into a discussion of specific science learning outcomes though it is important to point out that there is still some debate in the field about the assumption that learning is an outcome, either a possible outcome or interestingly even a desirable one. For instance, back in 2002, Ted Ansbacher argued whether learning was the right general word to describe the impacts of LSIE experiences. In a response to that article, Dierking, Cohen Jones, Wadman, Falk, Storksdieck & Ellenbogen (2002) acknowledged that this was certainly not a new argument--the field has been debating the issue for many years. But what we did suggest is that after a great deal of research by us and others and interactions with literally thousands of visitors, we feel that learning is the correct term with a critical caveat— the learning must be broadly defined, clearly operationalized and emerge from the meaning the person makes of the experience, rather than what the ISE professional thinks the outcome should be (this does not mean that the designer of the experience has no goals but that in the development of the experience, be it an exhibition, program, web site or game, s/he must be aware of the visitor's perspective in order to negotiate and accomplish our desired outcomes). We argued further that when one used this broader frame to conceptualize and generate potential outcomes, they were similar to outcomes recommended by Ansbacher (2002).

Although this debate was a few years back, we are still aware of people that do not entirely buy into the notion of learning as the outcome for LSIE experiences, so it is important for the Consensus Study group to understand the context in which their report, focusing very directly on learning in and from informal science environments, may be received by some. In our opinion the debate is moot. A growing number of studies (Falk, 2001; Falk, Dierking & Storksdieck , 2005; Falk, Storksdieck & Dierking, 2006; Miller, 1998; 2001; National Science Board, 2002; Weiss, et al, 2005) demonstrate that schooling is necessary but not sufficient to support lifelong science literacy. For example, seventy-five percent of Nobel Prize winners in the sciences report that their passion for science was first sparked in non-school environments. We also now appreciate that the public acquires science information continuously across their day and throughout their lives. As an organization, the Institute is in full support of learning, broadly framed, being the outcome of such experiences in fact, the Institute's primary not-for-profit mission is to understand, facilitate and advocate for the importance of such learning. We believe that learning is an outcome but at a pragmatic level also appreciate there is important political currency in the word. This is particularly the case in a changing world with diminishing resources and a need for innovative approaches to education and learning, a niche that free-choice learning organizations and institutions by their very nature are designed to fill. Free-choice learning organizations and institutions have worked to be recognized in their communities as important components of the overall learning infrastructure.

Consequently, rather than rejecting the notion that LSIE experiences foster, support and reinforce learning, we have argued for years that the field should be playing an advocacy

18

role which supports broader views of where, when, why, how and with whom learning takes place (2000, 2002, 2003). Fortunately that message is being communicated by others as well beyond the immediate ISE field. In 2005, the Harvard Family Research Project initiated the notion of complementary learning, advocating for the essential role that schools and a variety of nonschool learning supports, such as families, early childhood programs, out-of-school time activities and programs, higher education, health and social service agencies, businesses, libraries, museums, and other community-based institutions can play *together* in supporting learning. Hopefully this will be an outcome of the Consensus study as well.

In order to make sense of the breadth of possible science learning outcomes in and from informal learning environments, we feel that is it is critical to first understand why people of various ages engage in ISE experiences. Research suggests that the motivation behind most adult's free-choice science learning[2] is a desire to satisfy a personal sense of identity and to fulfill personal intellectual and emotional needs. For instance, adults visit settings like national parks, science centers and botanical gardens to satisfy their intellectual curiosity and stimulation, as well as to fulfill their need for relaxation, enjoyment and even spiritual fulfillment (Ballantyne & Packer, 2005; Doering & Bickford, 1996; Doering & Pekarik, 1996; Storksdieck, Ellenbogen & Heimlich, 2005; Hood, 1983; Kaplan & Kaplan, 1982; McManus, 1992; Moussouri, 1997; Paris & Mercer, 2002; Prentice, Davies & Beeho, 1997; Rosenfeld, 1980).

Adults take their children to these settings because they feel such experiences are worthwhile, educational and fun, and in the process they learn science (Borun, Chambers & Cleghorn, 1996; Borun, Chambers, Dritsas & Johnson, 1997; Dierking, Luke, Foat & Adelman, 2000; Dierking & Falk, 2003; Falk & Dierking, 1992;). Adults also encourage their children to participate in a wide variety of after-school and extra-curricular experiences including 4-H, scouting and summer camp experiences, many of which also support science learning (e.g., Dierking & Falk, 2003; Ponzio & Marzolla, 2002; McCreedy & Zemsky, 2002; St. John, Carroll, Hirabayashi, Huntwork, Ramage, & Shattuck, 2000). Everyday learning outside of institutions, including science learning during travel or when watching science-related specials on television, complement the list (Crowley & Galco, 2001; Falk, 2001; Falk & Dierking, 2002; Storksdieck, Ellenbogen & Heimlich, 2005). Despite the fact that participating in family experiences on the part of children may not be free choice at all, research suggests that when given a choice children indicate they would prefer to participate in these experiences with family rather than as part of school groups, with the reasons often reflecting the free-choice and social nature of such experiences.

Unfortunately, with few exceptions, the families taking their children to informal science settings and encouraging their children to participate in OST science-related experiences are predominantly higher income and Caucasian. However, greatly influenced by the

---

[2] Learning in which individuals engage throughout their lives to find out more about what is useful, compelling or just plain interesting to them in the area of science; when they have the opportunity to choose what, where, when and with whom to learn. Such science learning is motivated by the learner's needs and interests, and for the most part is intrinsically motivated and under the choice and control of the learner.

growing afterschool and out-of-school time (OST) arena and specifically in science, NSF ISE's efforts to support community-based science education efforts, there are increasing opportunities for youth and families from poor and underserved communities to engage in OST science experiences. According to the Harvard Family Research Project's Study of Predictors of Participation in Out-of-School Time Activities (2007), participation rates in before and after school programs have increased at all levels of family income with the greatest increase among the lowest income youth. They attribute this trend to an increasing policy focus on the benefits of OST, along with extensive funding for the 21st Century Community Learning Centers program and suggest that policymakers and the public need to continue to focus on equity to ensure that this trend continues. Thus it is encouraging to know that equity issues are a focus of the Consensus study as well.

Despite the fact that children may prefer to participate in LSIE activities with their families, it is important to acknowledge that there are a number of children who only participate in LSIE experiences through school visits and programs. In addition, a more focused type of science learning occurs when individuals search for specific science information on the Internet, in libraries, through conversations with peers or other knowledgeable persons or when they participate in science-related hobbies and special interest groups (Chadwick, 1998; Crowley and Jacobs, 2002; Eveland & Dunwoody, 19972001; Gross, 1997; Proctor & Dutta 1995; Russell Tytler, Duggan, & Gott, 2001). Science learning in these situations is driven by specific, intrinsically motivated needs, e.g. finding out more about an ailment, or about pest control when aphids infest the garden. While there is anecdotal evidence that this type of focused free-choice science learning is highly effective, little systematic research has been conducted to assess the degree to which such learning occurs and the conditions which influence it.

(a) Breadth of LSIE outcomes

With this background in mind and based on the scan of ISE and related fields, what LSIE outcomes (and in the case of the wider afterschool/family/community involvement and youth development studies what general outcomes) were observed across the evaluation studies we reviewed? Were discernible patterns observed?

As a reminder, approximately 32 studies were analyzed in depth, 20 of them were ISE, NIH or IMLS-funded and 12 were studies in the afterschool/family/community involvement and youth development fields. Seven (7) of the studies were front-end, and with the exception of one, which was a front-end evaluation of a trail in a national park, all front-end studies were focused on exhibitions. The remaining 25 studies were summative evaluations which were used to develop the following lists of observed outcomes within ISE exhibition, youth and community and new media projects and studies in the wider afterschool/family/community involvement and youth development fields.

Observed LSIE Outcomes in Summative Evaluation Studies of Exhibitions We analyzed 15 summative evaluation studies that focused on exhibitions. These included studies conducted at a wide range of ISIs. These included: seven science centers: New York

Hall of Science, NYC; Franklin Institute Science Museum, Philadelphia; the Exploratorium, San Francisco; Museum of Science, Boston; Maryland Science Center, Baltimore; California Science Center, Los Angeles; and the Museum of Science & Industry, Chicago; three zoos and aquariums: San Francisco Zoo; Shedd Aquarium, Chicago, and Monterey Bay Aquarium; Monterey, CA; one children's museum, Bay Area Discovery Museum, Sausalito, CA; one history museum, Chicago Historical Society and one university visitor center: Cornell University's Nanobiotechnology Center. Two summative studies were included from Monterey Bay Aquarium and two from the Exploratorium accounting for the 15 summative studies we analyzed.

The content areas explored in these projects were equally diverse including: what is known and not known about dinosaurs and how scientists understand them; local animals and habitats; African savannas; nano-biotechnology; wild reef-sharks, shark myths and mysteries; jellies; Chicago sports; space topics including the search for life; genetics, evolution; and, science process topics.

The following are common outcomes observed in these summative exhibition evaluation studies:

- Interest and engagement (often measured in terms of long holding times & positive affect, as well as self-report)
- Understanding specific scientific concepts and underlying mechanisms including biological & cultural evolution, individual growth & development, conservation; most often post- only designs
- Misconceptions/confusions about specific scientific concepts and underlying mechanisms; most often post- only designs

Less commonly observed LSIE outcomes in these summative exhibition evaluation studies include:

- Understanding of specific scientific concepts and underlying mechanisms that persists over time (4 months was the longest time period measured)
- Emotional connection
- Extensions to other learning experiences (school projects, checking out a book, family conversations)
- Understanding the scientific enterprise; what we know and don't know about a topic and why
- Engaging visitors in developing their own narratives and pretend play around exhibition topics
- Attitudes about specific content ideas, often controversial (e.g. evolution)

Observed LSIE Outcomes in Media Evaluation Studies We analyzed 6 summative evaluation studies that focused on media. These included studies conducted for two public television stations with strong track records for informal science media development: WGBH, Boston and Twin Cities Public Television, St. Paul, MN; and one for a children's museum, an internet-based design technology learning environment. Most of the projects focus on summative evaluation of the media program developed but

given the nature of these projects they also involve interactions among additional venues including botanical gardens, natural history museums and science/technology center venues/exhibitions, the home, the Internet and other web-based activities. However, one of the WBGH projects was a self-directed, online tutorial program designed to help after-school educators learn how to lead hands-on science activities with children. The content focused on in these efforts also varied but tended to be more general topics than exhibitions and included engineering, general science, science inquiry and technology concepts.

The following are common outcomes observed in these summative media evaluation studies:

Media piece alone:
Understanding specific scientific concepts and underlying mechanisms including engineering, general science, science inquiry and technology concepts, most often pre-post designs

Media piece with accompanying guide/web site:
Understanding specific scientific concepts and underlying mechanisms including engineering, general science, science inquiry and technology concepts, most often pre-post designs
Attitudes/interest/affect about and toward specific content ideas, most often pre-post designs

One less commonly observed outcome in these summative media evaluation studies was:

No difference in on-line and face-to-face training

Observed LSIE Outcomes in Program Evaluation Studies We analyzed 4 summative evaluation studies that focused on programs. These included studies conducted at two museums/science centers: Fresno Metropolitan Museum and Franklin Institute Science Museum one aquarium, New Jersey State Aquarium, Camden, NJ (though this project was a collaboration that also included the Franklin Institute Science Museum, the Philadelphia Academy of Natural Sciences, and the Philadelphia Zoo) and one university: University of Minnesota. Because we received so few summative program evaluation studies we also included findings from the *Engaging America's Youth* research study that Institute researchers recently completed (Koke & Dierking, 2007) for the Institute for Museum & Library Services. This study surveyed 247 projects, all past recipients of IMLS funding for youth development efforts; 52% of the sample were museums (n=128), 40% were libraries (n=99) and the remainder fit into "the other" category (n=20), primarily institutions in the formal education sector such as community colleges, schools or universities, though in these cases they were merely the administrators of the grant. All projects took place in the informal environment of a museum or library. Twenty-two percent of these programs (n=54) served youth indirectly by producing a product such as a curriculum or web site, or focused on adult leader outcomes, or did both.

The following are common outcomes observed in these summative program evaluation studies include:

- Attendance & participation of target audience in evaluated program & activities
- Interest and engagement (often measured in terms of repeated attendance & positive affect, as well as self-report), sometimes pre-post designs
- Understanding specific scientific concepts and underlying mechanisms; most often pre-post designs and often persisting over time
- Self-reported increased interest and feelings of competence in science among children and youth
- Self-reported increased self-confidence and improved social skills
- Extensions to other learning experiences (school projects, checking out a book, family conversations)
- Self-reported improvements in youth's relationships with family members and the community
- Self-reported improvements in academic achievement since participation
- Improved perceptions/connections by youth to cultural institutions and positive adult role models in community

Less commonly observed LSIE outcomes in these summative program evaluation studies include:

Changes in behavior (self-report such as re-landscaping, planting, etc.)

- Impacts on adults participating in programs (e.g. awareness by adults for importance of learning in general and science, specifically or increased interest / engagement /knowledge on part of adults participating
- Extensions to other learning experiences (school projects, checking out a book, family conversations)
- Understanding the scientific enterprise; what we know and don't know about a topic and why
- Attitudes about specific content ideas
- Increased perception among youth that they are important resources in the community

Observed Outcomes in the Wider Afterschool/Family/Community Involvement & Youth Development Programs Of the 37 studies we analyzed in depth, 12 were studies in the wider afterschool/family/community involvement and youth development fields. Four of these were family support & early/elementary childhood intervention programs providing an array of social, educational, and health services to parents and preschool-aged children, five were afterschool and summer programs for upper elementary and older youth and three combined afterschool and family support components. These programs were primarily situated or coordinated through schools though they were led and funded by a diverse set of organizations including city governments, private foundations, cultural institutions and not-for profit organizations focused on youth and community development.

The following are common outcomes observed in these wider afterschool/ family/ community involvement and youth development programs that we reviewed:

Number of parents who completed GED or other educational goal

Attendance & participation of families in center's programs & activities

Children's self-reports about behavior changes and positive conversations with parents

Knowledge, attitude, and behavior changes observed among youth and children related to particular behaviors (e.g., nutrition and physical activity) with pre-post designs

Knowledge gain among parents (e.g., knowledge regarding healthy food preparation and physical activity) with pre-post designs

Increased scores on standardized measures/scales as evidence that the program enhanced some dimension of child development or adult literacy, by a certain percentage over what would be expected without intervention with pre-post designs

Less commonly observed outcomes include:

Statistically significant positive change in the parent-child interactions from pre-to post- videotaping

Staff participating in training gained knowledge and ability to implement the program effectively

Attitude and behavior changes among parents with pre-post designs

Local outcomes such as significant increases in earned income, a reduction in welfare, increase in adult education, decrease in number of pregnancies, and an increase in birth with pre-post designs.

(b) Quality of Current Evaluation Practice

In gathering examples of current evaluation practice, we took three avenues: We utilized studies (1) posted publicly on informalscience.org; (2) from the Institute's project portfolio; and, (3) requested from other evaluators in the field. As noted in the previous section we also received permission from the IMLS to include findings from the recently completed *Engaging America's Youth* research study. When we approached the other evaluators in the field they were willing to share their reports with us but with the understanding that individual studies would not be identified or judged and that they were studies that *the evaluators* deemed were *representative* because of their important and/or interesting characteristics. They were not necessarily exemplars.

We appreciate the reason for your request for us to assess the quality of the available evaluation studies *as a whole* though appreciate that this is still a difficult challenge given the variety of research perspectives and frameworks (mostly implicit) guiding this work and the fact that evaluation studies are quite different depending upon whether they are an exhibition, media or youth/community program. Taking all of this into consideration, our criteria for selecting studies to analyze in depth were fairly basic: (1) at a minimum studies needed to have measured outcomes (interestingly not all did); (2) the outcomes needed to have been measured in a way that we felt had integrity (baseline data of some kind was collected though it should be noted that in many cases, particularly in exhibition evaluations, these data were self-report data, albeit collected in a high quality manner; (3)

samples were collected in an appropriate manner (either random or strategic) and, (4) sample sizes were adequate (at least 50). We had hoped to use the presence of theoretical frameworks, clear identification of independent and dependent variables, pre-post designs and creative methodologies as criteria also, but they were not consistently included in studies so we did not use them as overall criteria, though studies that included these aspects are definitely a part of the analysis.

With these basic criteria in mind, we feel that the studies we included in this analysis provide a reasonable understanding of what is being accomplished in the LSIE field, particularly given the current environment which typically includes inadequate funding for evaluation (this has been changing recently with the NSF's leadership, and for the most part, a lack of clear guidelines for what constitutes high quality evaluation). We think that it is fair to say that for the most part evaluators in the field have done what was expected but not necessarily more. Most studies are quite informal, lack clearly delineated independent and dependent variables and actually lack vision in terms of what *could* be measured rather than what *should* be measured. We hope that that this report, the focus on IRB approval, and the current effort by the NSF to implement some common metrics, will help to raise the quality of evaluation efforts, encourage innovative approaches and methods and ensure that there are some common elements that enable true comparison among and between projects.

(c) Characteristics of a Quality Evaluation Study

In terms of what constitutes a quality evaluation study, Institute for Learning Innovation researchers feel it includes the following:

(1) It is based on a theoretical framework that ideally also drives the project being evaluated;
(2) Research questions/hypotheses are clearly stated;
(3) Independent and dependent variables are clearly delineated and defined in ways that are objectively measurable;
(4) All appropriate IRB procedures and ethical standards of research are followed;
(5) If the goal of the evaluation is to measure change, then baseline data are collected;
(6) All data samples are collected in an appropriate manner (either random or strategic) and sample sizes are predetermined to insure adequate analysis by appropriate statistical tools;
(7) All data analysis is conducted using appropriate statistical tools and procedures (both quantitative and qualitative); and,
(8) All conclusions are reported in a way that clearly identifies not only findings but implications of the evaluation; all stated in clear English in a manner that affords easy comprehension by lay professionals.

Section VI: Contributions of Findings to Understanding of the Benefit of ISE Experiences

(a) Theoretical frameworks that support LSIE evaluation

A variety of theoretical frameworks have been applied recently to understand, define or evaluate science learning in informal settings. Most of these frameworks have attempted to provide a broader perspective on learning outcomes, yet are compatible with the nature of informal/free-choice learning. These models, frameworks or theories include some discussed in earlier sections that are based on or framed within constructivist and socio-linguistic theories:

- § The Contextual Model of Learning (Falk & Dierking 2000), a general framework for understanding free-choice learning (see also Falk & Storksdieck, 2005, for an application and quantitative validation of the model); the model focuses on 10-12 key personal, social and physical factors that alone, and in conjunction with one another, foster free-choice learning. The model stresses visitor agenda, personal motivation, the sociocultural nature of learning, the importance of physical context and long-term outcomes.

- § The Integrated Experience Model, a constructivist model developed to explain specific affective and cognitive environmental learning experiences in media-rich out-of-school learning environments like planetaria, is based in part on the Contextual Model of Learning and focuses on the learner's willingness to learn or receptivity/openness for certain information in the learning process.

- § Situated/Enacted Identity (Falk, 2006; Rounds, 2006) focuses on audience expectation and audience agenda in terms of true, underlying interests that are intimately linked to the audience's enacted identity during a visit or free-choice learning experience. This framework is based on a large body of literature that considers the entry narrative of the visitor as a key factor in understanding motivation and learning from an informal/free-choice learning experience.

- § Family learning though not a theoretical framework per se has been an important way of reframing LSIE experiences that changes the focus from any one individual in a learning group, such as the child, to the entire family group (Dierking, Luke, Foat & Adelman, 2001; Ellenbogen, Luke & Dierking, 2004; Astor-Jack, Whaley, Dierking, Perry & Garibay, 2007; Ellenbogen, Luke & Dierking, 2007; Ash, 2003; Crowley & Galco, 2001; Ellenbogen, 2002; 2003; Borun, Dritsas, Johnson, Wagner, Fadigan, Jangaard, Stroup & Wenger, 1998).. In this context "learning" is defined as "a joint collaborative effort within an intergenerational group of children and significant adults." Outcomes include learning science concepts, attitudes and behaviors, but also learning about one another and the members within that group, as well as shaping and reinforcing individual and group identity. Family learning approaches are grounded in socio-cultural theories and are currently transforming the way some museums and

science centers such as the Children's Museum of Indianapolis (Dierking, Andersen, Ellenbogen, Donnelly, Luke & Cunningham, 2005), Pittsburgh Children's Museum and Chicago Children's Museum are reorienting their missions, educational strategies, and experiences.

Other frameworks have recently emerged and have been used to inform evaluation studies in LSIE, such as Community of Practice (CoP, see Lave & Wenger, 1991), Positive Youth Development (PYD) and Possible Selves.

- § CoP has become a common lens through which many in-depth community-based efforts and professional development projects in the informal science education arena are being developed and assessed. In this framework, the model can be used to understand participants' trajectories from science novices (peripheral members of the science community according to CoP) to more active and core members, engaging in authentic science and sometimes even participating in apprentice-like activities with scientists, engineers and technicians. The framework can also be interpreted more broadly and not used in the strict sense of concentric circles of decreasing professional identity (the notion of peripheral, active and core members) in a group of people who share common goals, resources or ideas. For instance, Sharp (1997) defines a "community of practice" as an informal network of practitioners with the shared need to work more effectively or to understand their work more deeply, who collaborate over a period of time and through extensive communication to develop a common sense of purpose in order to share their knowledge and experience. In this manifestation, CoP provides a particularly strong underpinning for many online resources and discussion groups, particularly those whose goal it is to be sustainable without ongoing financial and personnel resources.

- § PYD and Possible Selves frameworks have also been used recently primarily in assessing youth programs since they provide a framework that recognizes impacts that transcend typical STEM cognitive, attitudinal or behavioral change outcomes (Koke & Dierking, 2007; Luke, Stein, Kessler & Dierking, in press). They are also grounded in sociocultural theory and address the broader developmental needs of youth, in contrast to traditional deficit-based models which focus solely on youth problems, such as substance abuse, conduct disorders, delinquent and antisocial behavior, academic failure, and teenage pregnancy. A more recent PYD framework describes six characteristics of positively developing young people which successful youth programs foster. Proposed by Lerner and referred to as the Five C's: cognitive and behavioral competence, confidence, positive social connections, character, caring (or compassion) and contribution, to self, family, community, and ultimately, civil society, they suggest that when young people manifest these five C's across their development, they can be described as thriving. The framework referred to as 'Possible selves' (Stake and Mares, 2005) proposes that individuals' perceptions of their current and imagined future opportunities serve as a motivator and organizer for the individual's current task-related thoughts, attitudes and behaviors, thus "linking current specific plans and

actions to future desired goals". The basic concept is that if programs wish to effect behavior change, they must first change self-concept (Markus, 1986) . In short, a program may be measured as successful if it positively impacts self-concept. In that sense, PS is close to the concepts of 'self-efficacy' that have been identified as crucial to behavioral change (Kollmus & Agyeman, 2002). Since PYD and Possible Selves frameworks focus on the true developmental needs of youth, they encourage program development that supports youth more deeply, or even defines personal growth outcomes as the major objectives for LSIE programs, thus allowing for positive program impact where previously only small or seemingly negligible change in traditional learning occurred. The Institute for Learning Innovation has utilized these frameworks in the evaluation of roughly a dozen such programs in the last five years or so, and the frameworks not only allow for more meaningful program development and evaluation, but also provide program staff with a powerful language for communicating about their work and its benefits.

(b) Science Learning Outcomes in Light of Lifelong Learning

Over the course of a lifespan, people learn most of what they know outside of school and formal learning arenas. As suggested, researchers now appreciate that the public acquires science information continuously across their day and throughout their lives – even school-aged children utilize a wide range of non-school sources for constructing their science understanding (e.g., Anderson, 1999; Bransford, 1997). Many researchers have suggested learning is a natural human process - learning does not begin and end at a specified time, or when someone external to the individual determines learning should occur. Although the free-choice/informal learning that occurs throughout a person's lifespan at times utilizes skills such as functional literacy that are acquired to a large degree through formal schooling, all learning is cumulative. Estimates within the U.S. suggest that, across the lifespan, approximately 3% of an individual's life is spent in school, university/community college/technical school training, or professional development. As discussed in Section II of this chapter here is a tremendous body of literature discussing, and often arguing, meanings of informal, nonformal, and incidental education. While, for the purpose of instruction and planning, these distinctions are important and should not be minimized, for the *learner*, be it child, adult, professional, tourist, hobbyist, activist, parent or one of countless other roles, life and all the non-schooled learning that occurs within it is seamless. The idea of free-choice learning shifts the perspective from that of the institution to that of the learner (Falk & Dierking, 2002).

So although some of what the public knows about science is shaped by compulsory schooling, people construct their understanding in this area over the course of their lives, in many places and contexts, and for a variety of reasons. People learn about science on the job, while engaged in personal investigations, through civic organizations and during their leisure time (Anderson, 1999; Anderson, Lucas, Ginns & Dierking, 2000; Falk, 2003). People report that they have high levels of interest in science-related topics, and

their interests and motivations in specific science topics tend to vary depending on both their prior knowledge and the connections they are able to make between various science topics and the reality of their daily lives.

The 1998 NSF report on the public's attitude and knowledge of science and technology (National Science Board, 1998) reinforces these findings, stating that "Americans get most of their information about public policy issues from television news and newspapers." Sixty-eight percent of the US adult population reports that it watches TV news for at least one hour every day, while about 46% read a newspaper on a daily basis (Table 1). Another 28% listen to radio news for at least one hour a day. Newspaper and other print media are a more important source of information for people with higher levels of formal education, while the reverse is true for television. Listening to the radio does not seem to vary as function of level of education.

About 15% of all Americans read at least one science magazine per month, while 53% reported watching one science TV program during the same time. Almost twice as many men than women seem to read science-related magazines (20% versus 11%), and more than a quarter of those with college degrees and higher reported reading at least one science magazine per month, while just 14% of those who did not hold college degrees did so, and just 9% of those without a high school degree did. In other words, reading about science in magazines is strongly influenced by gender and levels of formal education. Not so watching science on TV, at least when this is measured as a percentage of the population watching science on TV rather than total time spent watching TV. There is hardly any difference between the genders, or the level of formal schooling.

Table 1. Public use of various sources of information, by selected characteristics: 1997 and 2001 (Percentages)

Source	All Adults	Male	Female	Less than HS	HS	BS/BA	Graduate
Print							
• Read newspaper every day (1997).	**46**	49	43	41	44	53	59
• Read newspaper every day (2001).	**42**	45	39	23	44	48	60
• Read at least one newsmagazine every day (1997).	**14**	15	14	6	14	22	27
• New magazine read regularly (2001).	**16**	17	14	7	13	25	31
• Read at least one science magazine per month (1997).	**15**	20	11	9	14	25	29
• Science fiction books or magazines read regularly (2001).	**16**	16	17	7	19	13	14
Radio, TV							
• Watch TV news for at least one hour every day (1997).	**68**	63	72	80	67	59	54
• Watch TV news every day (2001).	**63**	60	66	61	66	57	63
• Listen to radio news for at least one hour every day.	**28**	29	27	31	27	27	27
• Watch at least one science TV program per month.	**53**	56	50	45	55	57	52
Public library, museums							
• One visit per year to a public library (1997).	**70**	68	72	52	71	87	84
• One or more visits per year to a public library (2001).	**75** (31*)	71	78	60	74	85	85
• Five visits per year to a public library (1997).	**45**	41	48	31	43	68	63
• Five or more visits per year to a public library (2001).	**48**	42	53	27	48	62	67
• Visit science museums, zoos, etc. at least once a year (1997).	**60**	63	59	34	64	78	75
• Visit science museums, zoos, etc. at least once a year (2001).	**66** (37*)	64	68	54	64	81	83
Book purchase							
• Purchase at least one book per year.	**61**	55	56	33	62	85	87
• Purchase at least one *science* book per year.	**31**	32	29	9	31	51	56
Sample size (1997)	**2000**	930	1070	420	1188	257	135
Sample size (2001)	**1571**	751	823	116	834	393	221

* Answers by 16,028 Europeans as part of the Eurobarometer 55.2 (December 2001). Source: European Commission, 2001.
Adapted from Science and Engineering Indicators, 1998 and 2002.

Similar results were found in a study of Los Angeles residents (Falk, Brooks & Amin, 2001). Individuals were asked to indicate which sources they used to keep current in science (Table 2). For this population, books (not for school) emerged as a more important source than television, but the popular media was not far behind along with life experiences such as personal health issues and the acquisition of information needed for day to day living. Museums (including science centers, zoos, aquariums, etc.) emerged as statistically comparable to schooling as a source of continuing public education.

30

Table 2: Ranking of Sources Relied Upon "Some or A Lot" or "Not At All" for Learning About Science and Technology (N = 1007)

Rank Order	Relied Upon "Some or A Lot"		Relied Upon "Not At All"	
	%	Category/ Source	%	Category/ Source
1st	76%	Books, magazines, not for school	37%	Radio, audiotapes
2nd	74%	Life experiences	23%	On the job
3rd	74%	TV, cable	18%	Family/ friends
4th	68%	School, courses	15%	School, courses
5th	65%	Museums, zoos	13%	Museums, zoos
6th	57%	On the job	10%	TV, cable
7th	55%	Family/ friends	8%	Books, magazines, not for school
8th	31%	Radio, audiotapes	8%	Life experiences

From: Falk, Brooks & Amin, 2001

(c) Are some learning outcomes unique to informal environments?

As we discussed in Section II, learning is learning. However, by this point in the chapter it should also be clear that different environments, be it a classroom, a museum visit, a family playing a computer game at home or an adult participating in a dance session, afford different kinds of outcomes. The impacts of the range of LSIE experiences are broad, depending upon the personal, sociocultural and physical contexts of the experience. What is important in defining and operationalizing these outcomes is to understand the basic characteristics of learning science in free-choice/informal environments:

§ Learners' agendas and motivations drive the outcomes
§ Outcome goals developed by project staff may or may be shared by the learners
§ Learning is constructed meaning applied by the learner
§ Learning is continual
§ Learning is cumulative
§ Learning is horizontal (in other words, synthesized across a variety of learning experiences)

Lifelong free-choice learning is different from classroom learning and the field is well served to consider some of the following ideas when evaluating the learning that results from these experiences:
§ Understand the often complex agenda and the mixed motivations of free-choice learners. Free-choice learners who visit zoos, watch a documentary or visit a national park often seek to combine enjoyment and learning experience: for them, learning occurs through fun, and is mostly highly personal in nature. Jan Packer (2006) has termed this type of learning experience as "learning for fun" and
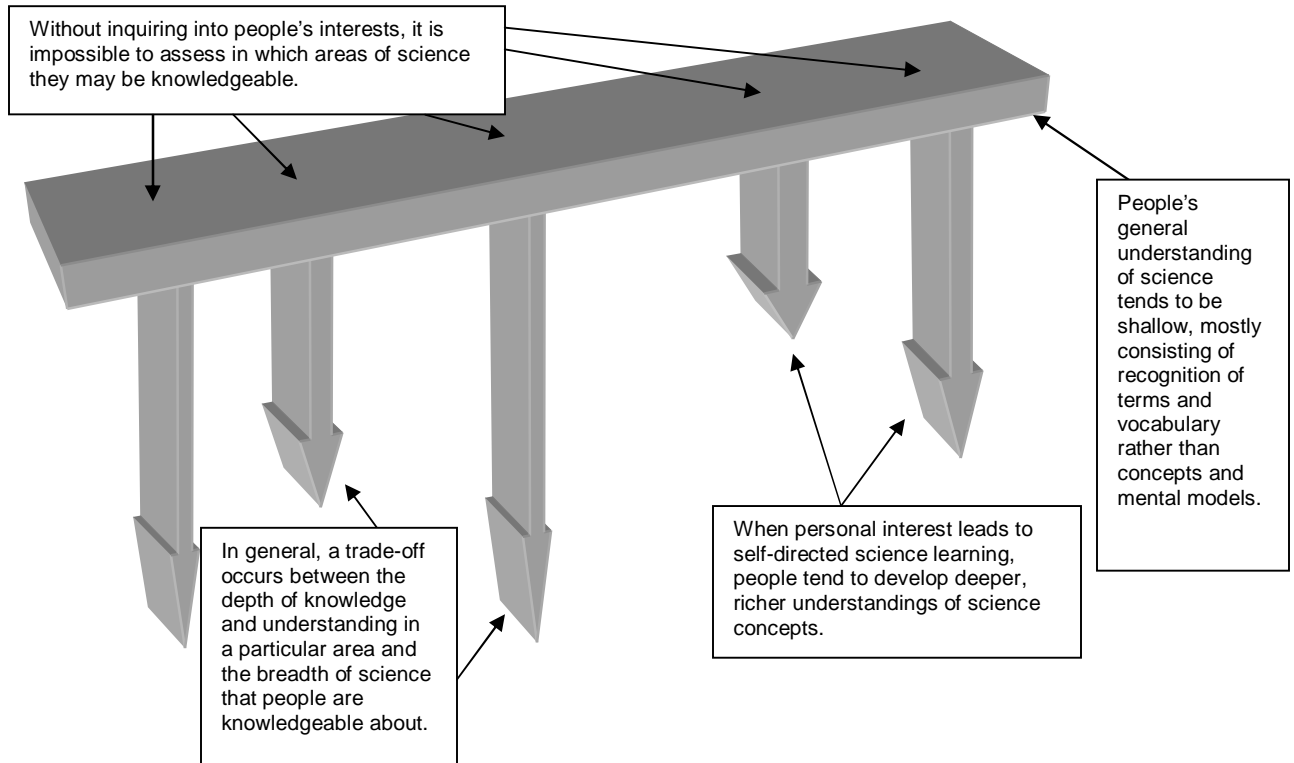
Marilyn Solvay, Director of the Maine Historical Society calls it "laughing and learning."

§ Understand the target audience and define it well. A LSIE activity that seeks to address everybody is most likely satisfying no one as it will likely fail to serve individual needs, wants or expectations sufficiently. Solid front-end evaluation or needs assessment is crucial and is best embedded in a development model that links audiences to expected outcomes through clearly defined experiences.

§ Lifelong, free-choice learners are not empty vessels: they come with prior knowledge and understanding—sometimes with alternative conceptions and other times knowledge exceeding those designing the experiences. Learners bring with them awareness, attitudes, interest and intentions and a lifetime of experiences leading to this moment. Again, solid front-end evaluation is needed to uncover these aspects of one's audience, particularly since we know that learners are not necessarily open to new ideas if those are not embedded into the learners pre-existing cognitive and emotive background.

§ In contrast to a school environment, where educators can make some reasonable assumptions about the background of the students, free-choice/informal audiences tend to be heterogeneous, even when they have self selected to be present in a program. It is only possible to understand the audience when you involve them or when you talk to them or otherwise engage with them. This does not necessarily need to involve costly structured evaluation studies. You can provide many low-cost and low effort opportunities to hear from participants, including brief conversations, short feedback forms and debriefing conversations, as long as you keep an open mind for the need to hear from the audience.

§ There is no such thing as an average person: assuming that learners in a non-school setting are seeking the same outcomes from a single experience is to set oneself up for failure. Rather than understanding the "audience", consider the many audiences—smaller groups with similar backgrounds can be addressed in similar ways, but it is important to remember that one size does not fit all.

§ Impact, or audience outcomes, should be in part defined by the target audience and carefully calibrated to what is possible to achieve for that particular audience: do not overestimate (or underestimate) the depth of impact you can achieve. Some audience members may be ready to change their behaviors, others might be skeptical and need more convincing, but both could be in the audience.

(d) What is known about the cumulative effects of science learning across time and contexts?

As stated above, the cumulative effect of science learning through assimilation and accumulation (to use Piagetian language) does not only form the basis for curriculum development or entire school systems (as in Germany or Switzerland), it also applies to the concept of free-choice science learning, just that the learning path is less determined by external forces to the learner and more by a constant interplay between motivation, curiosity, learning experiences and follow-up across a variety of settings and situations. Multiple experiences reinforce learning. Motivation and cognitive effects intertwine: some locations motivate for active follow-up or ready for subsequent reinforcing

experiences; others provide content or process or cognitive knowledge (Falk & Storksdieck, 2005).

Knowledge that has been gained from learning science in informal environments, however, tends to be concentrated on focused areas of interest (Falk & Dierking, 2002). Figure 3 provides a visual representation of this concept.

Without inquiring into people's interests, it is impossible to assess in which areas of science they may be knowledgeable.

People's general understanding of science tends to be shallow, mostly consisting of recognition of terms and vocabulary rather than concepts and mental models.

In general, a trade-off occurs between the depth of knowledge and understanding in a particular area and the breadth of science that people are knowledgeable about.

When personal interest leads to self-directed science learning, people tend to develop deeper, richer understandings of science concepts.

Figure 3: People's science knowledge in a simple schematic. The general knowledge is represented by the shallow bar, specific knowledge that is deeper and richer is represented by downward pointing arrows. The length of the arrow (or depth of the bar) represents the depth of knowledge and understanding, while the width represents the scope.

While knowledge of free-choice learners is often highly focused on the role of personal interest, it is important to note that most "learners" learn very little within the context of a brief "treatment." However, this ought not to be considered an entirely negative finding. School learning is rarely assessed based on a one-or two-hour class, yet free-choice learning is often assessed specifically after exposures that rarely exceed one to two hours. Falk & Storksdieck (2005) have demonstrated that learning can occur, but that it is likely restricted to those who bring with them interest and a learning agenda, and low to moderate prior knowledge. However, even those visitors to a science center exhibition who did not learn when learning was measured in a pre/post design took away something important: the potential to learn later what Bransford, et. al. refer to as preparation for future learning (PFL). Those visitors whose interest was sparked learned in the months that followed the science center visit, by engaging in other free-choice experiences.

(e) Public Learning Arenas and the Development of Science Knowledge and Interests

Science and technology education has been a priority in the United States for close to 50 years. In consort with these national priorities, researchers have been increasingly interested in investigating the public's attitude towards and knowledge of science (National Science Board, 2002).

These studies can be categorized into four main types:

1) Studies reporting how and where people learn science.
2) Studies reporting the public's attitudes and interest in science.
3) Studies reporting public scores on a variety of tests of science knowledge.
4) Studies reporting public attentiveness toward science and technology.

A few emerging patterns from these studies

The free-choice learning sector, in particular the popular media, is a source of science learning that influences public understanding of science.
As detailed earlier, the fact that the public acquires its science information outside of school appears to be a robust finding that holds up for all science subjects, despite the frequency with which science is taught in schools. For example, the most mentioned source of information about the Solar UV was the television weather forecast (35% of adults); followed by other television programs (30%) and newspapers (31%). Research in other Western countries corroborates these findings. Lehmann (1999) asked 1670 German middle and high school students for their source of detailed information about two specific science topics, the ozone hole and global climate change. Even for school-aged individuals, television and print media ranked higher than school as sources of information. It may be that beyond simply having the opportunity to learn about science outside of school, free-choice learning sources may present science and technology topics in ways that are more appealing to the general public, and hence are more memorable.

Interest and attitude toward science and technology
"Interest" is an important filter for selecting and focusing on relevant information within a complex environment (Falk & Dierking, 2000). In that sense, the psychological state of mind referred to as "interest" is an evolutionary adaptation to select what is perceived as important, thus relevant from the environment. People pay attention to those things that interest them, and hence, interest becomes a strong filter for what is attended to and learned. There are a range of additional factors that act as filters:
- Personal relevance, a precursor to interest (Pope & Gilbert, 1983).
- Attitudes and awareness of issues and
- Values and beliefs (Guba & Lincoln, 1989; Lucas & Roth, 1996).

In other words, a person uses his/her beliefs, values, or attitudes as filters with which to make sense of new information. For example, if information does not "pass the test" of our pre-conceived notions, we may disregard it. If we believe that there is no such thing

as global warming, we may discount information to the contrary. The same is true for "personal relevance." A person may pay attention to news about hypertension if that person, or someone close to that person, suffers from high blood pressure. The reverse may be true as well. Without a personal connection, or "hook" in the terminology of the news business, we will likely not pay attention to a story. These findings are corroborated by recent research by Falk and other researchers at the California Science Center (Falk, Brooks & Amin, 2001). When the public was asked to talk about science topics they were personally interested in, as opposed to science in general, the where, how and why of their science understanding was far more clear.

Level of interest in science and technology: Differences between topics

The public finds science and technology to be a highly interesting topic (Falk, et al., 2001). This high interest is true of all sectors of society; there were no significant differences across gender, race/ethnicity, education level or income. However, the specific areas of science in which the public is interested, do vary. Many research studies confirm the notion that making a personal connection to a science subject, or finding relevance to one's daily life, tends to increase people's stated, self-reported interest (see Table 3 for an overview). The general adult population is mildly interested in space exploration and nuclear energy, somewhat more than mildly interested in new scientific discoveries new technologies, and environmental issues, and fairly interested in medical discoveries. The US ranking is similar to that of Europeans, where medicine and the environment were the two areas of greatest scientific interest (European Commission, 2001, National Science Board, 2002).

Table 3: Level of public interest in selected policy issues: 1992–2001, selected years (Percentages and Mean Index Scores)

Year	1992		1995		1997		1999		2001	
Issue	VI	MIS	VI	MIS	VI	MIS	VI	MIS	VI	MIS
New medical discoveries	66	82	69	83	70	83	68	82	65	80
Environmental pollution	59	77	53	74	52	72	51	71	48	70
Issues about new scientific discoveries	36	61	44	67	49	70	45	67	47	69
Use of new inventions and technologies	37	64	43	66	47	69	41	65	43	66
Space exploration	22	47	25	50	32	55	28	51	26	50
Sample size	2,001		2,006		2,000		1882		1,574	

VI = Percentage of respondents who answered "very interested"; MIS = Mean Index Score. The original responses were converted to a 0-100 index: "very interested" = 100; "moderately interested" = 50; "not at all interested" = 0. SOURCES: Science & Engineering Indicators, 2002 (and earlier years).

Interest levels in these subjects vary further depending upon the specific science theme. For example, people's interest in space exploration is, on average, moderate. In 2001, almost as many people stated that they were "very interested" in space exploration (26%) than were "not interested" (27%). In comparison, only a sixth of Europeans (17%) declare an interest in astronomy and space. In contrast, 65-70% of people surveyed consistently reported very high levels of interest in medical discoveries, a subject that appears to have a closer relevance to people's lives. About half of the US general public indicates that they are "very interested" in environmental issues (mean index score of 70). However, these values have been gradually declining in the last ten years, from a high in 1992 of 59% who were "very interested" in environmental issues (Mean Index Score =

77). Interest in general scientific discoveries has been on the rise, from a low in 1992 of just 36% who were very interested to the current 47%.

Clearly, the general public's interest in specific science subjects shifts over time. Not surprisingly, then, interest levels in various science-related issues seem to fluctuate depending on when a survey was conducted, and what scientific issue was currently captivating the public's mind. It is not clear exactly what factors are responsible for these changes in public interest. However, various studies suggest that changes in people's interests and attitudes largely reflect the information that is filtered to them through the media (Gelbspan, 1998).

Knowledge & Understanding
The traditional method for measuring learning, or "science literacy," has typically been to ask textbook questions and to judge the nearness of an individual's answer to the expert's version of the scientific story. Based on what learning researchers now know about the nature of learning, this may not be the best approach to truly documenting what it is people understand about the world around them. Questions that are asked with an understanding of the ways in which people are likely to have incorporated salient aspects of a scientific idea into their own lives appear to be more appropriately measuring people's general level of science knowledge and understanding. Very few studies have been designed with this learning model in mind however. With that caveat, the studies that have been conducted still represent elements of what people know about pre-defined areas of science. Repeatedly, studies suggest that most people possess a very basic understanding of science and scientific principles and a wide range of studies have shown that the adult public and students alike often have alternative conceptions regarding the true nature of scientific cause and effect (see for instance Anderson, 1999; Roschelle, 1995). Consequently, most researchers have concluded that science knowledge in the general public tends to be shallow, and often does not incorporate a level of conceptual understanding that would be necessary to foster behavioral change, or apply that knowledge to everyday life (. You may recall Figure 2, "People's science knowledge in a simple schematic" which depicts this notion of learning well.

What most people seem to know about science tends to be fairly limited to understanding concepts and terms, but does not generally extend to general principles, explanatory models, and causes and effects. Because people tend to learn in free-choice settings and become interested in science-related topics through personal connections, it is not surprising that a meta-analysis of the literature suggests that people tend to have a somewhat accurate level of understanding in areas that touch their immediate life circumstances. In other words, though people seem to generally possess low levels of science knowledge, they are more likely to understand in-depth, situation-specific knowledge about science topics that have relevance to their daily lives than they are to have generalized knowledge within myriad science areas. However, some of the basic ideas that people have about scientific processes and causes and effects are alternative conceptions that lead to the development of mental models which do not conform with those of experts. It is also clear that prior knowledge, interest and sociocultural and

personal context have an impact on what it is people actually learn. What follows is a summary of the state of knowledge related to public understanding of science.

Perception of knowledge: Feeling informed
It is widely assumed that self-reported levels of knowledge are somewhat unreliable measures for the true status of science literacy. However, these self-report measures are important indicators in terms of assessing the public's interest in and need for information about specific topics. Table 4 gives an overview of the US public's sense of being informed about selected science issues over the course of the last 10 years.

Table 4: How well informed Americans think they are about selected policy issues: 1992–2001, selected years (Percentages and Mean Index Scores)

Year	1992		1995		1997		1999		2001	
Issue	VI	MIS	VI	MIS	VI	MIS	VI	MIS	VI	MIS
New medical discoveries	22	51	23	52	28	56	25	53	21	51
Environmental pollution	29	57	24	52	23	51	21	48	18	47
Issues about new scientific discoveries	12	39	13	42	19	48	17	44	14	42
Use of new inventions and technologies	10	38	12	40	16	44	17	43	12	38
Space exploration	9	33	9	33	16	41	13	37	10	32
Sample size	2,001		2,006		2,000		1,882		1,574	

VI = Percentage of respondents who answered "very well informed"; MIS = Mean Index Score. The original responses were converted to a 0-100 index: "very well informed" = 100; "moderately well informed" = 50; "poorly informed" = 0.
SOURCES: Science & Engineering Indicators, 2002 (and earlier years).

Interest in scientific issues seems to be linked to self-perceived levels of understanding about science topics. On the one hand, this may suggest that people who are interested seek out information and therefore subjectively feel informed. On the other hand, because there is information that is easily available and understandable, people may actually pay attention and develop an interest in the topic. A third hypothesis would be that the correlation between self-assessed knowledge and interest is based on a reinforcing mechanism, where initial interest and available information feeds an ongoing cycle of new and relevant information, subsequent increased knowledge, interest and awareness to pay attention to yet additional information. Interestingly, the public's self-perceived understanding of science jumped between 1995 and 1997, and suggests that such patterns can change rather dramatically within a relatively brief period of time. Self-perceived knowledge can also increase slowly in the general public, as happened in the area of new inventions and technologies; or it can wane. In 1992, 29% of those responding felt "very well informed" about environmental issues, a number that has declined gradually to 18% in 2001.

Overall, people tend to self-report that they have at most moderate levels of knowledge in science. With the exception of space exploration, almost half of all those responding stated that they felt "moderately well informed" about various science issues across the board (a third for space exploration). The difference in overall, or average perceived levels of knowledge between the various science topics stems from the ratio between those who report to be "very informed" and those who believe they are "poorly

informed." While these two categories are balanced for the areas of medical discoveries and environmental issues, the gap increases considerably for the other areas reported, and is most dramatic for space exploration; more than half of those responding felt that they were poorly informed. When the responses for interest and perceived levels of being informed are compared, it seems clear that the public is interested in learning about science, particularly in the areas of medical and environmental science.

How is self-reported knowledge related to "actual knowledge?"

The question remains whether self-reported levels of being informed about a subject translates into actual knowledge about the subject. The Eurobarometer 55.2 survey (European Commission, 2001) explored this issue in some depth. In a dual set of questions, Europeans were first asked to state their perceived level of comprehension about a series of issues, ranging from air pollution to mad cow disease. Afterwards, they were asked to rate a series of statements as 'true', 'false', or 'don't know' (Tables 5 and 6).

Table 5: Perceived comprehension of selected science issues (% EU 15)
"Could you tell me whether you have the impression that you understand each of these topics of not?"

Subject	I think I understand	I don't think I understand	Don't know
Air pollution	85.3	12.1	2.6
Mad cow disease	76.6	18.8	4.6
The greenhouse effect	72.9	22.4	4.8
Holes in the ozone layer	72.6	23.1	4.2
Global warming	72.3	23.4	4.3

Source: Eurobarometer 55.2 (European Commission, 2001)

Table 6: Knowledge and perception of topical scientific subjects (% EU 15)
"In your opinion, are the following statements true or false?"

Subject	True	False	Don't know
Holes in the ozone layer will cause more storms and tornadoes	55.7	**22.7**	21.6
The greenhouse effect can make the sea level rise	**74.7**	8.9	16.4
Mad cow disease (bovine spongiform encephalopathy) is due to the addition of hormones in cattle feed	49.2	**32.1**	18.7
Mad cow disease presents no danger to man	14.6	**78.3**	7.1

Source: Eurobarometer 55.2 (European Commission, 2001)

About 72 to 73% of Europeans state that they believe they understand the greenhouse effect, global warming, or the ozone hole. Yet, 56% of those responding believed -- incorrectly -- that holes in the ozone layer will cause more storms and tornadoes. This belief is held by a majority of those with high levels of formal education (53%), and is even prevalent (47%) among Europeans who on previous tests across various scientific disciplines exhibited a high level of scientific knowledge. Even those responding who believed they understood "holes in the ozone layer" did not give a correct answer much more often than those who did not state that they understood the issue of "holes in the ozone layer" (59% of "true" answers for the self-reported knowledgeable population).

However, self-reported comprehension can also be indicative of real, albeit shallow, knowledge. For instance, three-quarters (75%) of the Europeans asked during the

Eurobarometer survey in late 2001, believed that the sea level could rise as one of the physical effects of the greenhouse effect, a correct assessment. This proportion rises to 84% among those who replied to the previous question that they understood the "greenhouse effect."

The free-choice sector, in particular the popular media, sets the agenda for public understanding of science.
A large percentage of the specific knowledge individuals acquire during formal schooling wanes over time unless it is periodically practiced or renewed, although some of the more general knowledge may be retained over a long period of time. The specifics are continuously renewed by sources of information primarily of a free-choice nature: television, newspapers, books and museums. This suggests that the free-choice learning sector is a critical actor in developing, reinforcing and sustaining the general public's interest and understanding in science and technology.
If this is the case, who sets the agenda and who decides what topics of science are presented to the public? What are the rules and guidelines for the type of information made available to the public? Although many people typically discuss learning as if it occurred exclusively in schools, a great deal of research suggests that, in fact, learning occurs in many different arenas throughout a person's lifetime. What is unclear is to what extent people are learning in these areas related to specific subjects. Although there are numerous sources for acquiring science knowledge, we have found it valuable to sort them into three broad educational sectors (after Falk & Dierking, 2002): school, work, and free-choice learning. The free-choice learning sector affords people the opportunity for lifelong, voluntary, intrinsically motivated and largely self-chosen and self-controlled learning.

The fact that most Americans receive their scientific information from popular media as described earlier, in particular the news media, has profound impacts on what it is the public learns in terms of science, and how science is understood by the public. Journalists use the news values of timeliness, proximity, prominence, consequence (importance), and human interest to judge whether a science-related story is worth reporting. These are, however, not necessarily criteria scientists would use to judge the importance or significance of "science news". News media also tend to cover dramatic events rather than chronic issues (which is what most environmental and medical issues tend to be). News stories need a "hook" or "news peg" to hang the story, therefore, the media has an intrinsic bias towards catastrophe. There is also a profound difference between news on TV and news in print, as one of the important selection criteria for TV is that a news story be visually appealing on screen; a criterion that leads to a strong dependence on film material that has entertainment value.

For instance, visitors to the American Museum of Natural History, when asked about their interest in specific infectious diseases, indicated that they were most interested in learning about Ebola, AIDS, and hepatitis (Giusti, 1996). That Ebola was among the top three infectious diseases mentioned by the general public indicates that highly publicized events, such as the outbreak of Ebola in some African villages, strongly influenced public perception.

While the media has been widely criticized for its choice and presentation of science news, some of the elements that "sell" science in the mass media might also be relevant for museums and other free-choice learning environments since the logic of drawing attention and keeping the public attuned applies in similar ways. Harz and Chappel (1997, p.93; cited in National Science Board, 2000) note that "...Two things...are vital and...found in nearly all good stories about science: relevance and context. Since so much of science is incremental, the reporter and the public need special help in placing research in the context of the big picture." An often-forgotten aspect of the way the mass media reports science stories is the fact that few editors or reporters have any formal training in science [in stark contrast to many ISE media developers]. Although half the journalists who participated in a First Amendment Center survey had covered science, only six percent reported having science degrees. Hence, journalists are not much different from the general public when it comes to attitudes and knowledge about science.

News decision makers may decide not to cover science stories. These "gatekeepers" may believe that their readers or listeners are uninterested in science stories and/or will not be able to understand them; in that they may allow the bad experiences they may have had with high school or college science courses to influence their decision-making about what science news to print or air. Many journalists may also think that, because their publications or programs are devoting what seems to them sufficient space or time to stories about medicine and health that this translates into an adequate job of covering science in general. The bottom line is that this is an area of influence for the field and a niche for future research and advocacy.

VII. Major Changes/Advances in ISE Evaluation Practice
As we have discussed throughout this chapter, evaluation methods include an entire spectrum of educational and social science research methods: questionnaires, surveys, close-ended and open-ended interviews (face-to-face, phone and increasingly email), observations, focus groups, videotaping, audio-taping, among other approaches. However, in addition to an increasingly rich set of methods are other advances, many resulting from an increasing cross-fertilization of methods and approaches from other social research disciplines, as well as ISE's emphasis on strategic impact ("raising the bar" and demonstrating impact on the ISE *field,* beyond those reached directly by project deliverables) and innovation, requiring that projects "push the envelope" and build upon prior work in ISE specifically and educational research generally, a model similar to the NSF or NIH scientific research model.

(a) Embedding evaluation studies within theoretical frameworks

Although one would hope that a quality evaluation is grounded in a theoretical framework of some kind, ideally the one guiding the conceptualization and framing of the effort being evaluated, this has not always been the case, even for summative evaluation studies as the previous section demonstrated. The reasons for this stem in part, from the fact that the ISE field has tended to approach its work in a fairly a-theoretical fashion. Many of the field's practitioners entered the field from a STEM discipline and are not

familiar with the large body of social science research that could guide the conceptualization and implementation of their ISE efforts. Thus, because an evaluation study generally is guided and framed by the project being evaluated, ISE evaluators have focused on the very specific task of assessing the degree to which individual projects have achieved their goals, often not incorporating larger theoretical frameworks into the design of the evaluation study. In fairness to those working in the ISE field, some of this a-theoretical culture also relates to funding priorities which for the most part until recently have not required a theoretical perspective nor emphasized strategic impact and innovation, which often necessitate broader, more theoretically grounded approaches. It is also important to say, despite the rhetoric of many social scientists that have "discovered" the ISE arena in the last 5-10 years that many theories have not seemed particularly relevant to the work at hand, and the practical applications of the research in the area are far from obvious. Time, resources and capacity are also tremendous obstacles to meaningfully integrating theoretical frameworks into day-to-day practice, be it the overall project work or the evaluation specifically. The challenge for ISE practitioners is in identifying relevant frameworks (or helping to create/modify ISE-specific approaches) and then being able to translate findings from this research into practical applications for their ISE activities that can then also ground subsequent evaluation and improvement. This is a tall order actually but it has been a critical step in improving overall practice and certainly evaluation practice specifically. For even formative evaluation and usability studies can be conducted within the context of theories of human behavior and well-being.

Fortunately this is changing, in part because as the field is maturing, with more social scientists from other disciplines engaged in collaborative work within the field and also an increasing number of ISE-specific researchers, evaluators and practitioners who are grounded in solid social science approaches. Also as noted above, the ISE Division of the NSF has raised the bar itself by emphasizing strategic impact and innovation, requiring broader perspectives grounded more fully in theory. As a consequence, an increasing number of projects and evaluation studies are being grounded in theoretical frameworks borrowed and/or modified from other social science disciplines or in ISE-specific frameworks being created to incorporate the unique aspects of learning in this arena.

Frameworks borrowed and/or modified from other social science disciplines can be seen to "operate" at two different levels: a macro level and a micro level. Macro level frameworks are overarching and support the overall approach to the research methodology utilized within an evaluation study. Such overarching frameworks include Grounded Theory, Action Research, Positivist, or Interpretive theoretical approaches. Micro level frameworks are more specifically focused on the actual activities/interactions observed within a system though because of their epistemological origins they often are grounded in more macro level theories (e.g. socioculturalists tend to approach their work from Grounded Theory, Action Research, or Interpretive stances).

Some promising micro frameworks from other social science disciplines include sociocultural theories such as Community of Practice, Activity Theory and Possible

Selves frameworks; cognitive theories such as situated cognition, everyday cognition and social cognition; and, and identity-focused theories such as Situated Identity. There are also frameworks being borrowed and/or modified that are specific to particular types of ISE efforts. For example, there are a number of ISE projects that focus on conservation and/or the environment in which the desired outcomes often are changes in behavior. Many of the evaluation studies in these projects employ models and theoretical frameworks focused on behavior change. Some of these are borrowed and/or modified from the environmental education field such as the *Theory of Reasoned Action* and the related *Theory of Planned Behavior* (Fishbein & Ajzen, 1975 and Ajzen & Fishbein, 1980); Hines, Hungerford and Tomera *Model of Responsible Environmental Behavior,* based on Ajzen's theory of planned behavior (Hines et al.1986-87, Hungerford et al.1990, Sia et al. 1985); Fietkau and Kessel use sociological as well as psychological factors to explain pro-environmental behavior in their Model of Ecological Behavior (1981); Blake's *Value-Action Gap* model that focuses on real constraints; or recent amalgams that take into account personal, societal, economic and value aspects of behavioral change (Fliegenschnee & Schelakovsky, 1998; Kollmus & Agyeman, 2002).

Other models have been modified from the public health arena (Prochaska Stage Model of Behavioral Change, Prochaska & DiClemente, 1986; Prochaska, DiClemente & Norcross, 1992; Prochaska, Redding, Harlow, Rossi & Velicer, 1994).). In addition as suggested in the previous section many youth development projects are employing specific theoretical frameworks derived from child development and sociology such as Positive Youth Development (PYD). In particular, PYD is being explored as a theoretical framework for youth development efforts in this arena (Dierking & Falk, 2003; Koke and Dierking, 2007; Luke, Stein, Kessler & Dierking, in review), providing a broader frame in which to observe the impact of informal science learning experiences on youth that include learning science concepts and processes, but also include attitudinal and behavioral outcomes such as youth's increased commitment to learning and the development of social competencies and positive identity. Adopting the PYD framework has provided a larger perspective in which to discuss how youth participating in these efforts are being impacted in meaningful and research-based ways. .

There are also ISE-specific frameworks being created that incorporate the unique aspects of learning in this arena. One such framework is the Contextual Model of Learning discussed as a framework for many Institute exhibition evaluations, first proposed by Falk and Dierking in *The Museum Experience* and refined and expanded upon in *Learning from Museums.* The framework describes three overlapping contexts which serve as an organizing and operationalizing frame for relevant theory in this arena: (a) the *personal* context is an umbrella for theories relevant to the individual including among others, motivation, interest, prior knowledge and identity; (b) the *sociocultural* context is an umbrella for relevant sociocultural theories such as Community of Practice, Activity Theory and Possible Selves frameworks among others; and (c) the *physical* context is an umbrella for relevant environmental psychology theories such as behavior settings and place-based learning, among others.

In another example of an ISE-specific framework, the Center for Informal Learning and Schools (CILS) was recently named the Research & Evaluation Center for the NSF Academies for Young Scientists, a new initiative designed to increase student interest in STEM education and careers through out-of-school-time programming. The CILS research team is applying a theoretical framework they have constructed that draws on cultural historical activity theory, the new institutionalism in organizational theory, and theories of inquiry-based science to determine and document the ways in which the OST programs support and expand student participation in science education activities.

(b) Linking Evaluation Studies to Broader Mission and Vision Impacts

In an arena of increased accountability, many funders are requiring that projects demonstrate broad strategic impact. Thus, evaluators are increasingly being asked that the findings of their studies either contribute to the field (even some private funders are requiring that projects they fund demonstrate wider impacts and serve as proof of concept), or in some cases show whether and to what degree entire communities are affected or how projects contribute to the overall mission and goals of an organization. Thus the purpose of the evaluation becomes one of not merely improving an individual project as it is developed or assessing the degree to which an effort achieves its goals and objectives; rather, the evaluation assessed the overall usefulness of the project in reaching institutional or community goals. In this way, individual evaluation studies can become part of larger strategic assessments or ideally, the design for the evaluation is whole cloth with the strategic effort itself. In the U.S. this approach is referred to as strategic or asset evaluation or institution-wide assessment and there is a recent *Curator* issue devoted to the topic (Koster & Falk, 2007). Such an approach is also sometimes referred to as "validation" in Europe though the way validation is implemented in most European projects is that there is a summative evaluation team *and* a separate validation team.

The pressure to take such an approach is coming from a variety of directions, much of it external, including OMB requirements for federal agencies to demonstrate the impact of *program* areas rather than *individual* projects, private funders, and Directors and Boards at institutions seeking to know how their efforts address broader strategic needs in their local communities (rather than the traditional needs of PI's, project directors, or education directors, who have focused on whether and to what degree the project has accomplished its specific goals). Interestingly another driver for this kind of evaluation is increasingly the actual communities involved in the efforts. In particular, organizations supporting populations referred to as underserved are increasingly demanding that any efforts involving them be able to demonstrate real impact on the individuals and communities on which the effort is focused in order to continue participation (Reardon, Sorenson & Clump, 2003).

(c) Embedding Evaluation (and Research) into the Design and Implementation of Projects

We have observed this trend in two ways. First, the whole approach to project design is becoming far more developmental and design-based. In part this is because of ISE's new

guidelines which necessitate a backwards" design approach in which project staff ideally first identify the intended impacts they want to accomplish and then move forward by targeting an appropriate audience, developing the project design, assembling the project team and any partners involved and delineate project deliverables. This approach means that both front-end and formative evaluation studies are seen as critical steps in the design process. Although for many years there been efforts to meaningfully integrate evaluation into the project development process, this trend seems more frequent, natural and organic than earlier "team" efforts. Findings are being used to improve projects and there is increasing openness and acceptance that evaluation is helpful, that audience perspectives matter and that quality evaluation requires expertise and experience. This is resulting in new models of cooperation; for instance some projects are embedding rapid evaluation into the design and development process. The emphasis of rapid evaluation is on stressing capacity building and project staff involvement (even sometimes in data collection). Reports are delivered as bullets or PowerPoints, and followed up quickly in meetings or via phone by the evaluator and project team in order to discuss the results and brainstorm solutions.

Research-based projects are also using iterative design-based approaches in an on-going manner. The *Active Prolonged Engagement (APE)* project, a four-year NSF-funded effort at the Exploratorium, is an excellent example. Both an exhibition development project and an audience research study, the primary aim of *APE* was to explore approaches that might shift the role of visitors from passive recipients of information to active participants in the learning experience. In projects like this, the documentation of the process, often including interviews and videotapes from focal participants, becomes an important finding. In the case of the *APE* project 30 *APE* exhibits were developed and evaluated, a workshop was hosted by the Exploratorium, and a publication, conference presentations, and journal articles documenting the findings of the *APE* Team resulted.

(d) Longitudinal and Life-Wide Designs

As the Contextual Model of Learning acknowledges the experiences and knowledge learners bring to ISE experiences, as well as the subsequent reinforcing experiences in which they engage, are important factors to take into consideration when evaluating ISE efforts. The sum total of the experience is not just what happens at the time of the experience, be it inside the walls of a museum or in a after-school program at a community center - so much is determined by events that happen in the time before an individual arrives at the museum, and equally importantly, in the time subsequent to the visit. Without this longer perspective one cannot hope to entirely understand the learner's experience, thus studies are beginning to focus on science learning as part of lifelong learning; the perspective of "horizontal" learning.

Taking a more longitudinal approach to data collection allows researchers to get a more holistic picture of the role of LSIE experience within peoples' lives. Most commonly, longitudinal designs involve conducting follow-up interviews with participants/users weeks, months, and even years after the experience and results suggest that long-term learning is a result. Researchers have repeatedly shown that many of the conversations

that begin in the museum continue once families are back at home (c.f., Astor-Jack, Whaley, Dierking, Perry & Garibay, 2007). In addition, people are able to describe specific exhibitions and program elements without prompting, indicating the general durability of LSIE experiences. Other studies of long-past museum experiences (Falk & Dierking, 1997) have indicated that they are part of lasting memories because they often are not only a topic of discussion at the moment, but long after the experience. This is consistent with the notion that memories are one product of social discourse.

Ethnographic case studies that involved a long-term relationship between the researcher and a set of families who visited museums frequently, allowing for repeated observations and interviews before, during, and after museum visits (Ellenbogen, 2002, 2003), suggested that conversational connections between museum experiences and real-world contexts are frequent and yet must be examined carefully since the connections are not always obvious to those outside the family. This research was important because it also explored the life-wide notion of observing and talking to people in the various settings they experience in their lives. As a field, there is much to be learned by situating outcomes within larger frames of time, culture and space. The work of the Learning in Informal and Formal Environments (LIFE) Center at University of Washington should also broaden our understanding in this area.

e) Increased acceptance of qualitative/mixed methods and triangulation

With the need to document not only whether, but how and why informal learning experiences are effective the need for methodological explanatory power has increased in recent years, resulting in evaluators employing mixed methods and triangulation designs. There has also been a need to develop more rigorous quantitative scales specific to ISE, introducing item development and analysis to the field of ISE evaluation (this rigorous approach to instrument development has long been a part of more formal evaluation practice).

In addition, as qualitative approaches have been adopted, strategies for validating such data have also been integrated within designs. A variety of strategies can be used to ensure the validity of such approaches, including triangulation by method and analysis (content, cross-case, and recursive analysis), identification of patterns of contradiction, as well as coherence (Sewell, 1999), and periodic opportunities for study participants to review researchers' analysis and interpretive results to ensure validity from the perspective of the participants. Although a few evaluators still express preference for either quantitative or qualitative research/evaluation designs, mixed- or pseudo-mixed methods are far more common. Increasingly, at least the rhetoric of the field suggests that one must tailor the approach to the project and the types of questions being asked.

f) New methodologies in evaluation studies

As the ISE field is maturing and becoming more diverse, many new methodologies are being developed or adapted from other arenas. In fact as is the nature of most disciplines, methods are being constantly updated, old methods rediscovered and newly adapted to new theoretical models that are more akin and responsive to the free-choice nature of

much of this learning. For example, John Falk along with other Institute colleagues (Falk, Moussouri & Coulson, 1998) adapted concept mapping, developing a unique more constructivist way of using it and analyzing resulting data than used in the formal education arena. Personal Meaning Mapping (PMM) as this approach is called, is grounded in a relativist-constructivist paradigm, which recognizes that individuals visiting and/or participating in programs in informal, free-choice settings bring varied backgrounds and knowledge to the experience. This varied background and knowledge, as well as the social and physical context of the experience itself, shapes how a person perceives and processes the experience. Consequently, PMM is designed so that there is no specific "right" answer to demonstrate impact. Instead, this approach measures the unique conceptual, attitudinal and emotional impacts of a specified learning experience on an individual, focusing both on *the degree of the change* but equally on *the nature of the change.* It quantifies learning along four dimensions which are for the most part independent (Extent, Breadth, Depth and Mastery). Thus, it enables investigators to probe meanings that can not easily be gleaned from methods such as written questionnaires or surveys, structured or semi-structured interviews, or observational methods. Although it was initially designed as a summative pre-post tool, it has also been used in front-end studies and post-only summative designs. Other methodologies adapted from anthropology and cognitive science research include the use of journaling, think-aloud techniques, visual documentation, discourse analysis, portfolio assessment, and embedded assessment.

In some cases, methods that have become standard in the field have also been adapted. For instance, Institute colleagues have modified the traditional "timing and tracking" approach, creating an unobtrusive structured observation based on holistic measures, that recognizes that though the amount of time spent in an exhibition is a good quantitative indicator of visitors' use of a gallery space or exhibit element, it often poorly reflects the quality of the visitors' experience engaged with an exhibition. Therefore to complement quantitative measures, we developed a quality ranking scale with which researchers can assess the quality of interactions that visitors have in various sections of an exhibition or at specific exhibit components. It involves time to some degree but not solely.

In addition we have also explored some new approaches to analyzing data. For example, one focus for us has been efforts to tackle what we refer to as the variability problem. One of the implications of constructivist notions of learning is that the impact of an experience will vary widely across learners because of the tremendous range of experiences, knowledge, attitudes, interests, and motivations with which they have. These differences directly affect how each person perceives the experience, the sense s/he makes of it and ultimately, the degree to which it influences their learning. Theory, and increasingly ISE learning research, suggests that these outcomes will be highly personal and quite variable across learners. We have had success aggregating data from groups of people with similar entering knowledge, behavior and attitudes, demonstrating that changes are more discernible when people are grouped in meaningful ways. For example, in a study at the National Aquarium in Baltimore, significant changes in knowledge, understanding and attitudes emerged for similar groups of people that were not seen when the data was analyzed as a whole. We replicated this approach with

equally successful results in a subsequent study at Disney's Animal Kingdom (Dierking, Adelman, Ogden, Lehnhardt, Miller & Mellen, 2004)

g) Increased use of technology in evaluation and its associated affordances and problems (online surveys, recordings, video, hand-held devices, etc.)

Technology is influencing all aspects of our lives, so it should be no surprise that it is also having a tremendous impact on evaluation, both in terms of expanding data collection techniques, but also in terms of providing additional ways to represent and share research findings. The use of technology is also expanding the ability for researchers across the world to collaborate in meaningful ways, sharing and discussing data throughout the research process.

The use of these technologies can be categorized into some different groups. There are technologies being utilized that facilitate tracking, particularly in physical settings such as museums and science centers where visitors have a great deal of choice and control over where they go and one-on-one tracking can be time-consuming and expensive. These include cell phones and other hand-held devices, exhibitions with built-in tracking software that can capture real-time visitor response patterns/behavior, the use of GPS to track visitor use of space and RFID tags. In addition hand-held devices are being used to gather specific information from visitors, for example, gathering triggered opinions. Other devices being used to collect large scale data within museums and other settings include interactive theaters. Tracking of a sort also is increasingly imbedded into web sites and within computer interactive programs and these logging analysis systems are improving.

Technology is also revolutionizing data collection approaches such as the development and administration of questionnaires and observations. For example, the use of online surveys is increasing with excellent off-the-shelf software programs available (Websurveyor and Survey Monkey are two used frequently). In terms of other data collecting devices, small but highly effective microphones and file-producing software that enable effective use of audiotape data and sophisticated videotape software such as inqscribe, that allows researchers to work more effectively and efficiently with digital video and audio data, transcribing and subtitling more easily, are transforming ethnographic and qualitative data collection and analysis. There even are some efforts to use techniques developed early in the field's history, such as galvanic skin response as a tool for measuring visitor affect. Although the use of technology is for the most part having a positive influence on evaluation, there are challenges in terms of sampling with on-line surveys and web logs, and privacy issues.

h) The Randomized Control Trials (RCTs) Debate

A debate has been underway over the last few years in the ISE field specifically, and social science research community as a whole (as suggested in the Section IV) , regarding efforts by the federal government to have federal agencies (who support much of the research) to emphasize the benefits of RCTs to the exclusion of other research paradigms

(in fact there was an entire issue of *Educational Researcher* devoted to the topic; will find reference) In response, the NSF ISE Division, now part of the Division on Research in Learning in Informal and Formal Settings, hosted a NSF workshop on informal STEM evaluation last month to discuss the metrics developed for the Academic Competitiveness Council working group on Informal Education and Outreach. These metrics are intended to provide a common framework for assessing project impacts. PIs will be asked to identify one or more of these metrics appropriate to their projects, whether targeting public or professional audiences. Our purpose for being there was to discuss the metrics and the creation of a brief publication, designed to assist prospective PIs in planning, and current PIs in entering, project impacts into NSF's new online monitoring system and database. At this meeting (and earlier meetings last summer) there was an effort to discuss the limitations of RCTs in general and specifically within the context of LSIE. Some of the issues raised included the fact that learning is complex and multi-faceted, individual yet socially mediated within a rich physical setting. In addition, audiences are heterogeneous in terms of age, gender, background and knowledge. All of these present tremendous challenges to creating valid studies that utilize RCT design.

There was also a great deal of discussion about what makes this field unique, particularly the effort of ISE of late to emphasize strategic impact, build on prior work and educational research, while also supporting projects at the frontiers which may be too exploratory to easily fit into a RCT design. There was concern that adherence to common metrics, whether intended so or not, might ultimately limit the innovative and cutting edge nature of efforts in this arena. One other issue was raised regarded Levin's contention that it is impossible to balance precision, generalizability and reality at the same time, a faulty assumption not always explicitly acknowledged in RCT designed experiments.

This is an unfolding issue but ISE program officers have argued that as with any other discipline, RCT is an appropriate approach for some projects and some questions. They have asserted that it is best suited to informal *learning research* projects in their portfolio, investigating project-level "interventions" via exhibitions, media, web, and programs where generalizable, publishable data is critical, going beyond the more typical project that develops very specific deliverables and evaluates their specific impact. In these cases the added effort, time and resources required for a RCT study are justified. An ongoing challenge has been developing a relationship with an external evaluation partner. Although conversations with public health researchers at local universities have provided BEST Fit with feedback on community factors, components, and credible outcomes, we have also encountered differences in perspective regarding rigor and program appropriateness. These researchers emphasize experimental control and outcome measures, while LA's BEST values "kids' choice"—that is, allowing children to participate in activities that interest them, rather than requiring mandatory participation. Moreover, some common measures in childhood obesity research (e.g., children's weight, body composition, physical fitness tests, blood samples) are not consistent with our philosophy or program model. Still, BEST Fit continues to seek common ground with public health researchers while maintaining our program goals.

Another interesting point is that most people who say they are utilizing a RCT design are actually using a quasi-experimental design (Ferraro, 2005). Very few studies are able to truly use random selection because of the difficulties inherent in the process (Chamberlain, 2007). However, setting that observation aside, Chamberlain's recommendations for utilizing RCT design include:

1. It is important that one understands the power and limitations of RCT's. Although one can feel confident in the findings, they are often disappointing in terms of the information they provide, even in a study that demonstrates the significant strong of an intervention. This is because the purpose of RCT's is to answer was there an impact – and how much of one? RCT designs do not answer how the impact was made and why it happened. In order to provide that level of information mixed methods must be included.
2. Recruiting a large enough sample to make an RCT design valid can take far more time than one expects – Chamberlain recommends a year to recruit the sample.
3. The best incentives were a delayed incentive – if you are in the control group for this study, we will give you the intervention for free afterwards. This beat out cash and many other efforts.
4. Know what the IRB permission standards are before you begin recruiting a sample – the permissions you will need will strongly influence who will want to participate.
5. Recruit very widely – try and recruit more than you actually need for the study because people will drop out or be eliminated.
6. Plan a significant number of site visits to ensure 1) intervention sites are implementing the program as you desire and 2) that control groups are not being contaminated by the study.

One other point is worth considering. Lee Cronbach, a respected educational psychologist and psychometrician, has recently written several papers demonstrating that, when well done, matched control groups are equally reliable and valid as RCT's, removing selection bias to the same degree.

(i) The Influence of the Institutional Review Board (IRB) on evaluation practice

Another extremely important influence on evaluation practice currently is the requirement that projects receive IRB approval for their research/evaluation designs. Although in concept the notion of conducting research in ethical ways that protect and safeguard the privacy and autonomy of research participants seems paramount and it requires the team early on to be organized and systematic, there are significant challenges. These include: 1) the IRB process was developed in the context of large scale RCT medical experiments and thus many of the assumptions are not valid for the types of projects and environments in which ISE projects are focused; 2) the process is for the most part coordinated through universities and thus ISIs (and even more important, evaluators in the field) are not fully equipped to support the process; and 3) because by its very nature require the IRB process requires a level of consent and development of instruments prior to beginning the effort, there is concern that it limits naturalistic inquiry that often best fits within the free-choice learning context, affording no spontaneity or

ability to adapt; in fact there is grave concern that the process actually interferes with the free-choice nature of most of these experiences.

VIII. Conclusions & Recommendations

 (1) We learned from the research and evaluation of LSIE that we reviewed some of the following:
- § Combined, interwoven and interlocked learning outcomes ought to be expected (for instance, facilitating interest in science, creating awareness for science careers and strengthening a sense of Possible Selves, while also learning specific science processes and concepts as part of a summer camp program at a marine science center).
- § Brief experiences in LSIE are common, particularly with family visitors to museums and museum-type settings. These brief encounters tend to lead to more shallow "science learning" for most visitors, while deep, personally relevant learning only occurs for some visitors whose prior knowledge and experience facilitated such impacts. Some LSIE experiences (in-depth youth programs for instance) are deep, transformative and lasting.
- § All levels of learning ought to be expected, depending on the learning opportunities provided and the learning agenda of the free-choice learner.
- § It is often the combined effect of a multitude of LSIE experiences (and more formal experiences) that has an effect on the learner. Isolated events tend to have, on average, small effects that are difficult to demonstrate; yet most evaluation studies focus on the impact of an individual experience in isolation of other life experience that the learner may have.
- § Indicators that individual learning opportunities CONTRIBUTE to learning need to be developed in order to more accurately assess the effect and quality of individual experiences. LSIE experiences ought to be assessed and appreciated as part of a larger context of lifelong free-choice learning and a culture of lifelong learning in society.
- § People will learn what they want to and/or need to learn; creating awareness, interest and appreciation as precursors and sustainers for learning is essential in LSIE.
- § Learning as reinforcement/reconnection/reshaping: much LSIE is not about acquiring new knowledge, but is a reshaping of what was already known.
- § Learning as expanding what one knows: LSIE seems to provide opportunities, if on average in limited ways, to expand the body of knowledge and to create new knowledge AND understanding, primarily when the learner had a personal NEED to learn.
- § LSIE experiences tend to be highly intense and often exhausting, without the learners' awareness (see museum fatigue).

    LSIE creators need to have realistic learning goals that can be accomplished for the particular learners they engage.

(2) LSIE evaluation in the future

- § Should remain highly specific and assess the value of specific experiences (programs, exhibits, etc.) within the context of the learner's life.More proof of concept and design research evaluation studies are needed to support cost-effective, practical s evaluation.

True appreciation of specific learning opportunities is gained probably only when "science learning" is considered within or in combination with other personal but mostly implicit goals and agendas of the learner.

(3) Other thoughts
- § Few evaluation studies actually use theoretical frameworks (though this may be changing).
- § Outcome-based evaluation (OBE) has its benefits (focus, accountability, clarity, scope), but critics question its overall ability to capture all aspects, particularly unintended and incidental outcomes. OBE creates a need for a strong theoretical background, which in turn narrows projects.
- § IRB and RCT issues, as well as discussions about appropriate methods, can complicate the practice of evaluation and holds the potential to push out smaller evaluators, leading to a consolidation of the industry and added cost to projects.
- § Need for more sharing.
- § Need to diversity methods.
- § Need for regular meta-analyses
- § Finding, communicating and facilitating this balance will be a major challenge for the NSF ISE Program's new common metrics and online data monitoring system
- § Move beyond questions of whether programs matter for youth, families and communities given that research suggests that they do, to questions about why, how, and for whom these programs matter and matter most.
- § Still needed: a strong case for the importance of LSIE based on realistic notions of outcomes that can be accomplished in the range of types of experiences (1-2 hour visits, intense in-depth programs) and that are afforded by the variety of informal environments (museums & designed spaces, family and everyday learning, and after/out of school and adult programs). Need to see the benefits in light of all such experiences (infrastructure of LSIE)
- § Need to consider how these experiences contribute to what someone knows, understands, can do or feels. How do these experiences foster and reinforce interest and understanding and how do they connect to the other science learning experiences people have, both in other informal environment and in and from formal environments
- § No dominate theoretical model, but a variety of prominent ones (1) CoP, PYD and Possible Selves for youth and community-based programs and some professional development efforts, also social networking theories for some professional development and youth efforts; (2) behavior change models for environmental/public health/health projects; & (3) Contextual Model of Learning as a framework for theories related to the personal, physical and socio-cultural dimensions of learning and theories of identity and situated/enacted identity in and from exhibitions

- § Some innovative approaches to measuring outcomes include PMM, pooling data from people at similar stages, portfolios, and youth-created products for youth programs
- § New federal effort to focus on some common metrics (in the case of the NSF: changes in awareness, knowledge, or understanding; changes in engagement or interest; changes in attitude; changes in behavior; development/reinforcement of new skills and others and establish an on-line data monitoring system) could help to contribute to the understanding of how people benefit from experiences in informal science learning environments, however, it is important that these common metrics are interpreted broadly and that measures to assess them are creative, valid and reliable. It is hoped that they do not become standards that free-choice learning professionals feel they need to design/teach to.
- § Embedding evaluation studies within theoretical frameworks – leadership role for this study to recommend some directions and encourage projects to use such models in the conceptualization and framing of the effort being evaluated
- § Embedding evaluation (and research) into the design and implementation of Projects
- § Linking evaluation studies to broader mission and vision impacts – although currently this is most often being manifested in efforts to increase attendance there is another way of viewing this which is to make the goal linking science learning in these environments to public value and difficult issues: the empowerment of youth, families and communities through science; accomplishing complementary learning goals
- § "Backwards" design approaches in which project staff ideally first identify the intended impacts they want to accomplish and then move backward to target an appropriate audience, develop the project design, assemble the project team and any partners involved and delineate project deliverables. Front-end and formative evaluation studies are seen as critical steps in this design process.
- § Research-based projects are also using iterative design-based approaches as in *Active Prolonged Engagement (APE)* project, a four-year NSF-funded effort at the Exploratorium
- § Taking a more longitudinal and life-wide approach to data collection allows researchers to get a more holistic view of the role of LSIE experience within peoples' lives and the connections between the various science learning experiences they have.
- § Rigorous, multi-method approaches are critical
- § Increased use of technology in evaluation, along with the associated affordances and problems should be explored (online surveys, recordings, video, hand-held devices, etc.)
- § An important statement could be made about the role of and judicious use of RCTs in the practice of ISE evaluation.

IX. References

Adams, M., Falk, J.H. & Dierking, L.D. (2003). Things change: Museums, learning, & research in Xanthoudaki, M., L. Tickle & V. Sekules (Eds.) Researching visual

arts education in museums and galleries: An international reader, Amsterdam, Kluwer Academic Publishers.

Adelman, L., Dierking, L.D., and Adams, M. (2000). Summative Evaluation Year 4: Findings for Girls at the Center. Technical report. Annapolis, Maryland: Institute for Learning Innovation.

Anderson, D. (1999). *Understanding the impact of post-visit activities on students' knowledge construction of electricity and magnetism as a result of a visit to an interactive science centre.* Unpublished doctoral dissertation, Queensland University of Technology. Brisbane, Australia.

Anderson, D., Lucas, K. B., Ginns, I.,S. & Dierking, L.D. (2000). Development of knowledge about electricity and magnetism during a visit to a science museum and related post-visit activities. *Science Education.* 84(5) 658-679.

Ansbacher, T. (2002). "What are We Learning? Outcomes of the Museum Experience. *Informal Learning Review.* (ILR No. 53, March-April).

Ash, D. (2003). Dialogic inquiry in life science conversations of family groups in a museum. *Journal of Research in Science Teaching, 40*(2), 138-162.

Astor-Jack, T., Whaley, K.K., Dierking, L.D., Perry, D. & Garibay, C. (2007). Understanding the complexities of socially-mediated learning. In Falk, J.H., Dierking, L.D. & Foutz, S. (Eds.). *In Principle, In Practice: Museums as Learning Institutions.* Lanham, MD: AltaMira Press.

Ballantyne, R. & Packer, J. (2005).Promoting environmentally sustainable attitudes and behaviour through free-choice learning. *Journal of Environmental Education Research*

Blud, L.M. (1990). Social interaction and learning among family groups visiting museums. Museum Management and Curatorship, 9, 43-52.

Borun, M., Chambers, M., & Cleghorn, A. (1996). Families are learning in science museums. Curator, 39(2): 123-138.

Borun, M., Dritsas, J., Johnson, J.I., Peter, N.E., Wagner, K.F., Fadigan, K., Jangaard, A., Stroup, E., & Wenger, A. (1998). *Family learning in museums: The PISEC perspective.* Philadelphia, PA: The Franklin Institute.

Bransford, J. (1979). *Human cognition: Learning, understanding and remembering.* Belmont, CA: Wadsworth Publishing Company.

Bransford, J., Brown, A., & Cocking, R. (2000). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

Calabrese Barton, A., Drake, C., Gustavo Perez, J., St. Louis, K. & George, M. (2004). Ecologies of parental engagement in urban education. *Educational Researcher, 33*(4), 3-12.

Chadwick, J. (1998). Public Utilization of Museum-based WorldWideWeb Sites. Unpublished doctoral dissertation. University of New Mexico.

Chamberlain, A. (2007). Randomized Control Design Issues. Presentation by Anne Chamberlain, Research Scientist, Success for All Foundation. Baltimore, MD.

Churchland, P. S. (1986). *Neurophilosophy: Toward a unified science of mind-brain*. Cambridge, MA: The MIT Press.

Crowley, K. & Galco, J. (2001). Everyday activity and the development of scientific thinking. In K. Crowley, C. D. Schunn, & T. Okada (Eds.), Designing for science: Implications from everyday, classroom, and professional settings. Mahwah, NJ: Erlbaum

Crowley, K. & Jacobs, M. (2002). Islands of expertise and the development of family scientific literacy. In G. Leinhardt, K. Crowley, & K. Knutson (Eds.) *Learning conversations in museums*. Mahwah, NJ: Lawrence Erlbaum Associates.

Dierking, L.D., Falk, J.H., Rennie, L., Anderson, D., & Ellenbogen, K. (2003). Policy statement of the "Informal Science Education" Ad Hoc Committee. Journal of Research in Science Teaching, 40(2), 108-111.

Dierking, L. D., Andersen, N., Ellenbogen, K.M., Donnelly, C., Luke, J.J. & Cunningham, K. (2005). The Family Learning Initiative at The Children's Museum of Indianapolis: Integrating Research, Practice & Assessment. *Hand-to-Hand*. Washington, D.C. Association of Children's Museums.

Dierking, L.D., Luke, J. J., Foat, K.A. & L. Adelman. (Nov/Dec, 2001). The family & free-choice learning. Museum News.

Dierking, L.D. (2002). The role of context in children's learning from objects and experiences. Invited chapter for (Ed.) Paris, S. G. Multiple Perspectives on Children's Object-Centered Learning. New York, NY: Erlbaum.

Dierking, L.D.; Cohen Jones, M.; Wadman, M.; Falk, J.H.; Storksdieck, M. & Ellenbogen, K. (2002). Broadening Our Notions of the Impact of Free-Choice Learning Experiences. *Informal Learning Review* 55: 1, 4-7.

Dierking, L.D. & Falk, J.H. (2003) Optimizing out-of-school time: The role of free-choice learning. New Directions for Youth Development. 97 (Spring), 75-88.

Dierking, L.D., Adelman, L.M., Ogden, J., Lehnhardt, K., Miller, L. and Mellen, J.D. (2004). Using a behavior change model to document the impact of visits to Disney's Animal Kingdom: A study investigating intended conservation action, *Curator,* 47(3), 322-343.

Doering, Z.D. & Bickford, A. (1996). Visits and visitors to the Smithsonian. In: M. Wells & R. Loomis (Eds.) <u>Visitor studies: Theory, research, and practice, Vol. 9</u>, Jacksonville, AL: Center for Social Design.

Doering, Z.D. & Pekakirk, A.J. (1996). Questioning the entrance narrative. <u>Journal of Museum Education</u>, <u>21</u>(3), 20-22.

Ellenbogen, K.M. (2002). Museums in family life: An ethnographic case study. In Leinhardt, G., Crowley, K. & Knutson, K. (Eds.). *Learning conversations in museums*<u>.</u> Mahwah, NJ: Lawrence Erlbaum Associates.

Ellenbogen, K.M. (2003). From dioramas to the dinner table: An ethnographic case study of the role of science museums in family life. <u>Dissertation Abstracts International,</u> <u>64</u>(03), 846A. (University Microfilms No. AAT30-85758)

Ellenbogen, K.M., Luke, J.J. & Dierking, L.D. (2004). Family learning research in museums: An emerging disciplinary matrix? In Dierking, L.D., Ellenbogen, K.M & Falk, J.H. (Eds). *In Principle, In Practice: Perspectives on a Decade of Museum Learning Research (1994-2004)*, Supplemental Issue. <u>Science Education</u>. *88*: 48-58.

Ellenbogen, K.M., Luke, J.J. & Dierking, L.D. (2007). Family learning research in museums: Perspectives on a Decade of Research. In Falk, J.H., Dierking, L.D. & Foutz, S. (Eds.). *In Principle, In Practice: Museums as Learning Institutions.* Lanham, MD: AltaMira Press.

European Commission. Eurobarometer 55.2. 2001.

The Encyclopedia of Informal Education (2006) [footnote website, http://www.infed.org/encyclopaedia.htm]

Eveland, W. & Dunwoody, S. (1997). Communicating science to the public via "The Why Files" World Wide Web. Paper presented at 1997 International Conference on the Public Understanding of Science and Technology, Chicago, IL.

Eveland, W.P. & Dunwoody, S. (2001). User control and structural isomorphisms or disorientation and cognitive load. <u>Communication Researcher</u>, 28(1), 48-78.

Falk, J.H. (1984) Public institutions for personal learning. *The Museologist. 46*(168), 24-27.

Falk, J.H. (Ed.).(2001).Free-choice science education: How we learn science outside of school. New York: Teachers College Press.

Falk, J.H. (2006). The impact of visit motivation on learning: Using identity as a construct to understand the visitor experience. Curator, 49(2), 151-166.

Falk, J.H., Brooks, P. & Amin, R. (2001). The role of free-choice learning on public understanding of science: California Science Center LASER Project. In: J.H. Falk (Ed.) Free-Choice learning: Building the informal science education infrastructure. New York: Teachers College Press.

Falk, J.H. & L.D. Dierking, Eds. (1995). Public institutions for personal learning: Establishing a research agenda. Washington, D.C.: American Association of Museums.

Falk, J.H. & Dierking, L.D. (2000). *Learning from museums: The visitor experience andf the making of meaning*. Walnut Creek, CA: Altamira Press.

Falk, J.H. & Dierking, L.D. (2000). *Lessons without Limit: How Free-Choice Learning is transforming education.* Walnut Creek, CA: Altamira Press.

Falk, J.H. & Storksdieck, M. (2005). Using the *Contextual Model of Learning* to understand visitor learning from a science center exhibition. *Science Education*, 89, 744-778.

Falk, J.H., Scott, C., Dierking, L.D., Rennie, L.J. & Cohen Jones, M. (2004). Interactives and visitor learning. Curator, 47(2), 171-198.

Falk, J.H., Dierking, L.D. & Storksdieck, M. (2005). Lifelong science learning research. In J. Moon, (ed.). Informal Science Research. Washington, DC: Board on Science Education, National Academy of Science.

Falk, J.H., Dierking, L.D. & Storksdieck, M. (2007). Investigating public science interest and understanding: Evidence for the importance of free-choice learning. *Public Understanding of Science, 15.*

Falk, J.H., Moussouri, T. & Coulson, D. (1998). The effect of visitors' agendas on museum learning. Curator. 41(2), 106-120.

Ferraro, P.J. (2005). *Are We Getting What We Paid For? The Need for Randomized Environmental Policy Experiments in Georgia.* Atlanta, GA: H$_2$0 Policy Center, Georgia State University.

Gelbspan, R. (1998). *The heat is on: The climate crisis, the cover-up, the prescription.* New York, NY: Perseus Books.

Griffin, J. (1998). *School-museum integrated learning experiences in science: A learning journey.* Unpublished PhD, University of Technology, Sydney, Sydney.

Gross, L. (1997). The impact of television on modern life and attitudes. Paper presented at 1997 International Conference on the Public Understanding of Science and Technology, Chicago, IL.

Guba, E., & Lincoln, Y. (1989). *Fourth Generation Evaluation.* Beverly Hills, CA:Sage

Harvard Family Research Project. (2007). *Findings from HFRP's study of predictors of participation in out-of-school time activities: Fact sheet.* Cambridge, MA: Author.

Hood, M. (1983). Staying away: Why people choose not to visit museums. Museum News, 61(4), 50-57.

Humphrey, T. & Gutwill, J.P. (2005).*Fostering Active Prolonged Engagement: The Art of Creating APE Exhibits.* San Francisco, CA: The Exploratorium.

Kaplan, S. & Kaplan, R. (1982). Cognition and environment. New York: Praeger.

Klein, K., Astor-Jack, T. . Jordan, J. Addelson, B., Rowe, S. & Kassing, S. (2005). Defining Informal and Formal Science Education: In J. Moon, (ed.). Informal Science Research. Washington, DC: Board on Science Education, National Academy of Sciences.

Kollmus, A. & Agyeman, J. (2002). Mind the gap: Why do people act environmentally and what are the barriers to pro-environmental behaviors? *Environmental Education Research. 8(*3):239-260.

Koke, J. & Dierking, L.D. (2007). *Engaging America's Youth*: The Long-Term Impact of Institute for Museum and Library Services' youth-focused programs. Unpublished technical report. Annapolis, MD: Institute for Learning Innovation.

Koster, E. H. & Falk, J. H.. (2007). Forum: Museums in the Value Chain of Society. *Curator.*

Lave, J. & Wenger, E. (1991). Situated learning: Legitimate peripheral participation. Cambridge: Cambridge University Press.

Leinhardt, G., Crowley, K. & Knutson, K. (Eds.), Learning conversations in museums. Mahwah, NJ: Lawrence Erlbaum Associates.

Roth, M. & Lucas, K.B.(1998). From "truth" to "invented reality": A discourse analysis of high school physics students' talk about scientific knowledge. Journal of Research in Science Teaching. 34(2)**:**145 - 179

McCreedy, D. & Zemsky, T. (2002). *Girls at the Center: Girls and adults learning science together*. Philadelphia: The Franklin Institute/Girl Scouts of the USA.

McManus, P. (1992). Topics in museums and science education. Studies in Science Education, 20, 157-182.

Miller, J. D. (1998). The measurement of civic scientific literacy. Public Understanding of Science, 7, 1-21.

Miller, J.D. 2001. The acquisition and retention of scientific information by American adults. In: J.H. Falk (Ed.). Free-choice science education: How we learn science outside of school. pp. 93-114, New York: Teachers College Press.

Moussouri, T., (1997). Family agendas and family learning in hands-on museums. Unpublished doctoral dissertation, University of Leiscester, Leicester, England.

Munley, M.E. (1986). Asking the right questions. *Museum News* 64.3: 18—23.

Munley, M.E. (1987). Intentions and accomplishments: Principles of museum evaluation. in Jo Blatti (Ed). *Past Meets Present: Essays About Historic Interpretation and Public Audiences*, Washington, DC: Smithsonian Institution Press.

National Science Board (2002), *Science &Engineering Indicators – 2002.* Arlington, VA: National Science Foundation (NSB-02-1)

Paris, S.G. & Mercer, M. (2000). Connecting museum objects with personal experiences. Paper presented as part of a paper set, Museum Learning Collaborative: Studies of learning from museums, at the annual meeting of the American Educational Research Association. New Orleans.

Ponzio, R. & Marzolla, A.M. (2002). Snail trails and science tales: Inventing scientific knowledge. *Canadian Journal of Environmental Education*. 7(2).

Pope, M. & Gilbert, J. (1983). Personal experience and the construction of knowledge in science. *Science Education*, 67(2): 193-203.

Prentice, R., Davies, A., & Beeho, A. (1997). Seeking generic motivations for visiting and not visiting museums and like cultural attractions. Museum Management and Curatorship, 16(1), 45 - 70.

Proctor, R. W., & Dutta, A. (1995). Skill acquisition and human performance. Thousand Oaks, CA: Sage Publication.

58

Prochaska, J.O. & DiClemente, C.C. (1986). Toward a comprehensive model of change. In: W.R. Miller and N. Heather (Eds.) <u>Treating Addictive Behaviors</u>. Boston: Plenum Publishing Corporation.

Prochaska, J.O., DiClemente, C.C. & Norcross, J.C. (1992). In search of how people change: Applications to addictive behavior. <u>American Psychologist</u>, <u>47</u>(9): 1102-1114.

Prochaska, J.O., Redding, C.A., Harlow, L.L., Rossi, J.S. & Velicer, W.F. (1994). The transtheoretical model of change and HIV prevention: A review. *Health Education Quarterly 21*(4): 471-486.

Reardon, K.M. with J. Sorenson & C. Clump. (2003). Partnerships with Communities and Neighborhoods. In Barbara Jacoby (Ed.). *Building Partnerships for Service-Learning.* Pp. 192-212. New York: Jossey-Bass Press

Roschelle, J. (1995). Learning in interactive environments: Prior knowledge and new experience. In: J. Falk & L. Dierking (eds.) <u>Public Institutions for Personal Learning</u>, pp. 37-51. Washington, DC: American Association of Museums.

Rosenfeld, S. (1980). <u>Informal education in zoos: Naturalistic studies of family groups</u>. Unpublished doctoral dissertation, University of California, Berkeley.

Rounds, J. (2006). Doing identity work in museums. Curator, 49(2), 133 - 150.

Russell Tytler, R; Duggan, S. & Gott, R. (2001). Public participation in an environmental dispute: Implications for science education. *Public Understanding of Science*, Vol. 10, No. 4, 343-364.

Sass, J. & Blumenthal, C. Evaluating BEST Fit: A Program to Promote Child and Family Health After School. *The Evaluation Exchange*. Volume XII, No. 1 & 2, Fall 2006.

Scott, C. & Dierking, L.D. (2003). The research context. In *Museums & Creativity*. (Eds.) Caban, J. & Scott, C. Sydney Australia: Powerhouse Museum

Sharp, J. (1997). Communities of Practice: A Review of the Literature.

Stake, J.E., and K. R. Mares. 2005. Evaluating the impact of science-enrichment programs on adolescents' science motivation and confidence: The splashdown effect. *Journal of Research in Science Teaching* 42, no. 4: 359-375.

St. John, M., Carroll, B., Hirabayashi, J., Huntwork, D., Ramage, K., & Shattuck, J. (2000). The Community Science Workshops: A report on their progress. Inverness, CA: Inverness Research Associates.

Storksdieck, M., Ellenbogen, K. & Heimlich, J.E. (2005). Changing minds? Reassessing outcomes in free-choice environmental education. *Journal of Environmental Education Research*, 11 (3): 353 - 369

Weiss, H.B., Coffman, J., Post, M., Bouffard, S. & Little, P. (2005). Beyond the classroom: Complementary learning to improve achievement outcome. *The Evaluation Exchange.* Volume XI, (1). Spring. www.gse.harvard.edu/hfrp/eval/issue

Appendix A--The LSIE Evaluation Matrix Questionnaire

Name of your evaluation study:

Authors:

Year completed:

1. Evaluation Type
    [ ] Front-end
    [ ] Formative
    [ ] Remedial
    [ ] Outcome/impact/Summative
    [ ] Other _____

2. What was evaluated? [Check all that apply]
a) Exhibit/Exhibition
    [ ] Permanent
    [ ] Traveling
b) Media
    [ ] TV/Documentary
    [ ] Radio
    [ ] Website
    [ ] Large-screen film
    [ ] Planetarium show
    [ ] Other:
c) Programs
    [ ] Family program
    [ ] Youth program
    [ ] Workshops/Laboratory
    [ ] Demonstrations/Presentations
    [ ] Museum Theatre
    [ ] Outreach
    [ ] Professional development
    [ ] Other kind of program
d) Other _____

e) Brief description of project that was evaluated including   primary audience (e.g. public or professional audience):

f) Duration of the evaluated project:

g) Approximate cost of the evaluation
    [ ] Less than $10,000
    [ ] $10,000 - $25,000
    [ ] $25,000 - $50,000
    [ ] $50,000 - $100,000
    [ ] $100,000 - $250,000
    [ ] More than $250,000
    [ ] Don't know/not available

3. What was/were the major goals of the evaluation? Were there specific research questions? If so, what were they?

4. Who was/were the target audience(s) of the evaluation?

5. Theoretical frameworks, models or theories
a) Was there an underlying theoretical framework of the evaluation?
    [ ] Yes
    [ ] No

b) What was the underlying theoretical framework, model or theory for the evaluation?

c) Why was this underlying theoretical framework, model or theory for the evaluation chosen?

6. What was the design of the evaluation?

7. What methods were used?
[e.g., unobtrusive observation, focus groups, written surveys, portfolio assessments, etc.]

8. What dependent and independent measures were utilized?
[If possible, please distinguish between affective, cognitive, and behavioral measures and list demographic or psychographic variables that were utilized in the study]

a) Independent variables/measures:

b) Dependent variables/measures:

c) How were those measures developed?

[Were they specifically created for the evaluation; pilot-tested, derived from published tools, instruments, methods?]

9. What were major outcomes/findings of the evaluation (broadly defined)?