

Assessing Learning in Informal Science Contexts¹

Michael Brody
Montana State University
brody@montana.edu

Arthur Bangert
Montana State University
abangert@montana.edu

Justin Dillon
King's College London
justin.dillon@kcl.ac.uk

¹ Commissioned paper by the National Research Council for Science Learning in Informal Environments Committee

Assessing Learning in Informal Science Contexts

Abstract

This paper is primarily concerned with the assessment of outcomes in informal settings with a particular focus on science learning. An examination of hypothesized or presumed outcomes (both intended and serendipitous) of informal science learning experiences leads on to a critical analysis of the methodologies and methods used to assess those outcomes across a wide range of contexts. The evidence provided by assessment and evaluation studies from the USA and overseas is used in two ways. Firstly, overarching lessons and trends which emerge are described. Secondly, having clarified what is known about the role of assessment in informal science learning, a prospective research agenda is outlined for consideration. The paper is divided into four sections. Part 1, *Assessment and Informal Settings*, addresses the theoretical underpinnings of assessment in informal settings. In it, we describe relevant theories of learning and assessment. What to assess and various methodologies are also described. Part 2, *Assessment Approaches Used in Informal Settings*, attempts to map well-recognized assessment methods on to the informal learning environment, taking into consideration the unique characteristics of learning in informal settings. We draw, primarily, from the literature related to assessing science learning in formal settings. We address the role of individual and social groups, the effects of assessment on learning and examine how assessment can promote learning outcomes. Part 3, *What We Have Learned* addresses the diversity and complexity of assessing science learning in informal contexts and analyzes the strengths and shortcomings of a sample of studies. Part 4, *Promising Future Directions*, examines the potential of new approaches for assessing science learning in informal learning environments. In addition to matching the unique qualities of the informal science learning experiences to assessment methodologies and methods, we suggest an emphasis on the theoretical underpinnings of an emerging discipline and a systematic approach to future assessment of learning.

Assessing Learning in Informal Science Contexts

Part 1

Assessment and Learning in Informal Settings

- 1.1 Introduction
- 1.2 The challenges of assessing science learning in informal contexts
- 1.3 The range of outcomes of science learning in informal contexts
- 1.4 Identity as an outcome of learning
- 1.5 Assessing and evaluation of science learning in informal contexts
- 1.6 Assessment and learning

Part 2

Assessment Approaches Used in Informal Settings

- 2.1 Overview of approaches
- 2.2 Qualitative/quantitative: a false dichotomy

Part 3

What We Have Learned

- 3.1 The diversity of the informal experience
- 3.2 The complexity of assessment methodologies and methods
- 3.3 Identity as a unique case of complex assessment
- 3.4 Analysis of strengths and shortcomings in assessing science learning in informal environments
- 3.5 Analysis of methodological and data acquisition methods
 - a) Methodological approaches
 - b) Data acquisition methods
 - c) Matrix of methodologies and methods
 - d) Allignment of studies (What?), methods (How?) and purpose (Why?)
- 3.6 A critical review of a selection of studies of science learning in informal contexts

Part 4

Promising Future Directions

- 4.1 Ideology (What?), epistemology (How?) and axiology (Why?)
- 4.2 What has been done that has worked?
- 4.3 What has not been done that has potential?
- 4.4 Conclusions
- 4.5 Discussion

References

Assessing Learning in Informal Science Contexts

Committee's Focus Question: What are the hypothesized or presumed outcomes (both intended and serendipitous) of informal science learning experiences?

Part 1 Assessment and Learning in Informal Settings

1.1 Introduction

This paper is primarily concerned with the assessment of outcomes in informal settings with a particular focus on science learning. An examination of hypothesized or presumed outcomes (both intended and serendipitous) of informal science learning experiences leads on to a critical analysis of the methodologies and methods used to assess those outcomes across a wide range of contexts. The evidence provided by assessment and evaluation studies from the USA and overseas is used in two ways. Firstly, overarching lessons and trends which emerge are described. Secondly, having clarified what is known about the role of assessment in informal science learning, a prospective research agenda is outlined for consideration.

1.2 The challenges of assessing science learning in informal contexts

Science learning takes place in an increasingly diverse set of informal contexts ranging from museums and science centres to Internet use in the home, and from school grounds to farms and community gardens. Learning episodes can be fleeting or substantial, they can be planned or unplanned; they can happen to individuals wandering through a gallery space, or they can be scaffolded by science educators. At one end of a spectrum, informal learning contexts can invoke astonishment and wonder, at the other, boredom or fear. Although we have some understanding of science learning in informal contexts for educating the public, we have only scratched the surface in terms of tapping its potential especially when we realize how little time is spent during our waking lives in formal education (see http://life-slc.org/?page_id=124).

We have, so far, been limited in our understanding of informal science learning by a variety of factors such as the wide range of potential outcomes and by a restricted ability to assess them. Students visiting a natural history museum might have their ideas about evolution challenged, might identify a bird that they see each day, might wonder at the size of a blue whale or might stand under the tail of a model dinosaur and dream of being an scientist. Tapping into these outcomes, some of which might emerge fully-formed some time after an experience, and which might vary from student to student depending on their gender, age, home circumstances, level of education, cultural and ethnic identities, is indeed challenging.

1.3 The range of outcomes of science learning in informal contexts

Research, so far, has found a wide range of assessable outcomes from learning activities in informal contexts. At one level, these outcomes might be categorized as: cognitive (that is, knowledge and understanding); affective (that is attitudes and feelings); social/interpersonal (such as empathy and communication skills); and physical/behavioral (Rickinson *et al.*, 2004). At a somewhat deeper level, Rickinson *et al.*, in a review of research on outdoor learning, go on to report that:

There is substantial research evidence to suggest that outdoor adventure programmes can impact positively on young people's: attitudes, beliefs and self-perceptions – examples of outcomes include independence, confidence, self-esteem, locus of control, self-efficacy, personal effectiveness, and coping strategies, interpersonal and social skills – such as social effectiveness, communication skills, group cohesion and teamwork.
(*ibid.* p. 6)

Although Rickinson *et al.*'s review did not focus on science learning in informal contexts *per se*, many of their findings have implications for this paper. For example, it is reasonable to suggest that any assessment of science learning might benefit from examining science communication skills as a potential outcome.

1.4 Identity as an outcome of learning

Although we have briefly described general outcomes of learning in informal settings above, we found outcomes related to identity to be a special case worthy of more in-depth discussion. Identity as an outcome of learning in informal settings is widely conceived as incorporating a large number of dispositions ranging from feelings (relatively unstable) to attitudes and values (somewhat stable) to beliefs (very stable). Authors in this area agree that identity is an important affective outcome with many complex dimensions that have not been adequately addressed in the research literature.

Affective outcomes related to identity center on personal psychological development, social skills and outcomes related to specific informal learning settings and experiences. Authors theorizing about informal learning contend that identity and related affect should not be separated from cognitive or skill development (Hull & Greeno, 2006; Brody, 2005; Falk & Dierking, 2000).

In most cases, identity is conceived as positional and interactional. In informal learning settings, as in any life activity, there are perceived obligations to interact in communities of people, museums, zoos, work, school, playgrounds, field expeditions, etc. In these informal learning situations participants present and represent themselves to others redefining themselves as they go. Both internal and external discourse leads to the redefinition and the becoming of a different person. Gray (2004) provides a rich description of this redefining of identity in an online community of adult learners. In considering the positional development of identity we see that identity outcomes emerge in relation to the experience and events in informal settings. In relation to museums and construction of identity, Rounds (2006) argues that visitors engage with about one third of the elements of an exhibit and then only limited

attention to those. However, the engagement is sufficient for visitors to engage in what he refers to as ‘identity work’. This is the process through which people construct, maintain and adapt their sense of personal identity and persuade people to accept that new person. Informal settings provide this opportunity for ‘identity work’ without deep cognitive engagement in any particular domain. Rounds characterizes this type of engagement as reflexive activity.

1.5 Assessing science learning in formal contexts

A discussion of common assessment methods used to evaluate learning in formal science contexts provides a guidepost for understanding the acute complexities of evaluating informal science learning. Formal science programs such as those associated with public schools typically use large-scale, standardized assessments for program evaluation purposes. Large scale testing programs provide norm-referenced results that can be used to compare groups of students across school districts, states, nations and countries. It is not unusual for governmental agencies to use these results to make funding decisions and to determine the quality of instructional programs at the local, state, national and international levels.

Examples of standardized, large-scale assessments commonly used for assessing formal learning programs include the National Assessment of Educational Progress (NAEP) and the Third International Mathematics and Science Study (TIMSS). Large-scale assessment systems, such as TIMSS, can efficiently supply data about science and math learning for large groups of students and allow for comparisons across student demographics (e.g., gender, age, and ethnicity) as well as comparisons across countries. However, there is concern that large-scale achievement testing procedures may not adequately tap the student knowledge and reasoning because of their heavy reliance on multiple-choice and short-answer items (Harlow & Jones, 2004). Research suggests that inadequate student performance on large-scale assessments does not necessarily reflect lack of achievement but is more likely to be indicative of students’ inability to discern the intent of the items characterizing these assessments (Murphy, 1995). This assertion is understandable considering the range of approaches and informal techniques that are used to teach science and mathematics concepts. A negative or unintended consequence of the use of these tests would be to suggest that all science and math teachers follow a prescribed, standardized curriculum without regard to cultural differences that exist across countries.

Formative classroom assessments, on the other hand, are created by teachers and other educational facilitators to obtain feedback about student learning. Instructors can use the diagnostic information gathered from informal classroom assessments to identify strengths, target areas for improvement, or guide teaching approaches. Black & Wiliam (1998) found that students learn more and perform better on large-scale assessments when teachers use sound formative classroom assessment techniques to guide their instruction. Self-assessment and peer-assessment techniques are other types of formative assessment methods that are capable of providing the type of diagnostic feedback that can be used to improve student learning (Black & Harrison, 2001).

Self-assessment activities help students to gain a greater awareness of their capabilities and to establish realistic learning goals. For example, students might score their own pre-test for an upcoming instructional unit and, based on their performance, establish their own personal learning goals for that specific unit. This form of feedback is metacognitive in nature because it creates a situation that allows students to become active learners as they engage in efforts to create their own personal knowledge structures. Peer assessment allows learners to co-construct knowledge by engaging in discourse motivated by feedback from other learners. For example, students may be asked to take on the role of the teacher or examiner by scoring homework assignments or quizzes of their peers. This technique has been proven to increase learning because in order to accurately score peer performances, students must thoroughly understand the scoring criteria and the products necessary for proficient and advanced performances.

The assessment methods just described work well in fairly structured learning environments. However, they can be difficult to use in informal settings where individuals have choices about the type of learning activities they want to engage in. How can normative data for informal learning venues be acquired when there is no guarantee that a representative sample of museum or nature center visitors would be assessed? How would we even know what individuals would comprise a representative sample of informal science learners? Self-assessments, consisting of thought provoking questions that accompany exhibits, posted along a nature trail or included with explanations of artwork may be more aligned with methods appropriate for assessing informal science assessments. Yet the accountability requirements of funding agencies still call for the use of formal assessment procedures in the form of fixed-choice surveys and tests of knowledge that generate quantitative results. These assessment practices are not necessarily flawed. However, when used in isolation formal assessments often result in program outcomes that are very narrowly focused and fail to substantially increase our understanding of the variables and their relationships that characterized informal science learning.

1.6 Assessing and evaluation of science learning in informal contexts

If we take learning to include outcomes such as knowledge, understanding, attitudes, beliefs, values as well as interpersonal and social skills, then the opportunities afforded for assessment are substantial. Notwithstanding the challenges touched on above, educators across a wide range of informal contexts have assessed and evaluated outcomes for decades albeit with methods of varying reliability and validity. Later in the paper, the range of methods used to assess informal science learning and the emerging evidence base will be examined but, at this stage, it is worth noting that it is difficult to discuss methods without also discussing the purposes of assessment and the use to which assessment data is put. So, for example, a major reason that informal science educators assess outcomes is to evaluate the effectiveness of exhibits and programs. For the purposes of this paper, we will conceptualize assessment as the measurement of learning outcomes and we shall use evaluation to indicate assessment for the purpose of judging effectiveness.

1.7 Assessment and learning

Models of assessment are usually underpinned by models of learning. Since the middle of the twentieth century, models of learning that have impacted on the design of informal science learning experiences have included behaviorism, personal constructivism, and social constructivism (Hein, 1998). At the same time, a range of other theories of development and learning, including intelligence and IQ, learning styles, and multiple intelligences have impacted on the design of informal science learning contexts and their subsequent evaluation. Another related factor is that different philosophical stances on learning, such as ‘experiential’, and ‘free-choice’, have also had an impact on pedagogy, design and assessment. Consequently, assessment in informal science learning might be conceptualized as having several dimensions including:

Outcome measured: knowledge through to behaviors and feelings (a non-linear dimension);

Purpose: For individual empowerment and information (micro-level) to nationwide comparisons for policy making (macro-level);

Learning model: From behaviorist through to social constructivism;

Methods: Questionnaires, interviews, observations, etc.

Part 2 Assessment Approaches Used in Informal Settings

Committee's Focus Question: What types of approaches have been taken in trying to assess outcomes in informal settings? What have these assessments yielded? What haven't they yielded?

2.1 Overview of approaches

This section presents a critical review of the methods researchers have used to collect evidence of learning in informal science environments. Studies of informal environments are ripe with examples of quantitative and qualitative methods for assessing learning. Quantitative research methods typically involve collecting numerical data that can be analyzed statistically to produce results that can inform decisions about learning. The data for quantitative studies is typically collected from fixed-choice survey questions that require respondents to select a response, or, open-ended responses that have been coded to represent a category of interest. Quantitative studies of informal science variables hypothesized to affect student learning attempt to control for sources of error that may influence the outcome variable of interest (e.g., learning) other than the phenomenon of interest (e.g., museum exhibit). Much of the quantitative research designed to assess learning in informal science environments can be classified as experimental, correlational or descriptive.

2.2 Qualitative/quantitative: a false dichotomy

Qualitative methods have also been used extensively to study the variables that impact on learning within the natural contexts of informal science environments. External variables that would be intentionally controlled when conducting quantitative research are viewed as critical for understanding the complex interplay of variables that influence learning within the bounds of specific informal contexts. Open-ended interviews, observations and reviews of documents are examples of qualitative data sources that are analyzed to gain deeper and more holistic understandings of informal learning environments.

Research conducted at museums and in other informal environments can be daunting because of the complexity and variability represented by the environment and the audience. The challenge for researchers investigating informal science environments is to account for a multitude of confounding, competing and mutually influencing factors to make sense of visitors' learning (Gutwill & Allen, 2006). In the current era of accountability, federal agencies that provide funding in the form of competitive grants (e.g., No Child Left Behind) specify their preference for the use of experimental designs that produce quantitative results for evaluating outcomes. However, several publications by the National Science Foundation promote the use of both quantitative and qualitative research strategies in ways that compliment one another by using multiple assessment methods that produce results that are capable of more accurately informing decision-makers about the accomplishment of program goals (NSF, 1997, 2002).

Experimental designs using random assignment are considered the ‘gold standard’ of social science research by many researchers including those who evaluate grant proposals written to obtain funding for both formal and informal science programs (Feuer, Towne, & Shavelson, 2002). However, our review of the literature related to assessing informal science learning environments revealed that the use of randomized designs is almost nonexistent. One explanation for the lack of this type of research is that it is often not feasible or ethical to randomly assign participants in informal science settings to control and treatment groups. However, attempts have been made to use more rigorous forms of quantitative research to assess the variables in informal science environments that significantly affect learning.

For example, Lebeau, Gyamfi, Wizevich, & Koster (2001) used a randomized, pretest-posttest experimental to study the effects of a novelty-reducing activity coupled with a goal-setting activity on the self-regulated learning of middle school students’ during a science center visit. Their study was based on prior research conducted on free-choice learning environments that suggested that providing information about the setting just prior to a visit reduces novelty and may enhance learning (Falk & Dierking, 1992). In addition, goal setting is critical to sustained exploration and reflection during visits to learning environments such as the science center where this study was carried out. Twenty-seven middle school students were randomly assigned to one of three treatment groups. Students assigned to the first treatment group were given only a map to help them navigate the science center. Students assigned to the second treatment group were provided with a map plus a worksheet activity on which they were asked to indicate three areas they would most like to see, two areas they are not interested in seeing and one area in which they would like to spend the most time. For the third experimental group condition, students were also given the map and the worksheet. However, they were also required to write a question to which they hoped to find an answer as a result of their visit. Students in all three groups completed pre and posttest survey questions related to help-seeking behaviors that were hypothesized to be related to informed goal-setting and their attitudes toward science learning. Results from a One Way Analysis of Variance (ANOVA) found that the three treatment groups did not differ significantly on survey questions written to assess informed goal setting. However, students in the map plus worksheet activity group evidenced higher mean scores on items related to goal-setting than students assigned to the map only group. In addition, results from a Wilcoxon’s matched-pair, signed-ranks test found that the number of students disagreeing with the statement ‘Science is mostly about memorization’ was found to increase significantly from pretest to posttest measures.

Miller (2001) offers a good example of quantitative research that demonstrates the use of correlational and descriptive methods. He used free-choice survey items from the 1997 Science and Engineering Indicators study to determine the categories of use of adults engaging in science activities. Results from a confirmatory factor analysis identified seven items (i.e. number of: public library visits, science museum visits, television shows watched, hours of home computer use, news magazines read, science magazines read and newspapers read) loading on one factor suggesting that the free-choice items assessed a unitary construct appropriate for assessing the use of

informal science-education resources. Miller used these seven items in a follow-up analysis using structural equation modeling to develop a path model hypothesized to explain the influences of age, gender, education level, number of dependent children, occupational interest, issue attentiveness and the use of informal science-education resources on survey respondents' understanding of biomedical and space constructs. Results from this analysis found that educational attainment and college science courses taken had the largest and most significant effects on understanding of biomedical and space constructs. The use of informal science-education resources was found to have small yet significant effects on both science constructs. This research makes a marginal case for the importance of the contribution of informal science resources because the reported effects (.23 for biomedical science and .18 for space science) were small even though they reported as having a significant influence on the science constructs that were assessed.

Research conducted by Falk, Brooks, and Amin (2001) is a representative example of descriptive research that uses data collected from surveys to assess science learning in informal contexts. Their research was conducted to determine how, when, where and why people learn science and technology. During telephone interviews, 1,002 California residents were asked to rate their interest (scale 1 – 10) and knowledge (scale 1-5) in science and technology (scale 1 to 10). Results indicated that most respondents self-reported having a very high interest in science and technology and possessing knowledge level about science and technology equivalent to or slightly higher than the average. In addition, most of the individuals indicated that their main sources for acquiring knowledge of science and technology were divided into three categories: school, work or free-choice. Books, magazines (74%), life experiences (74%), TV (74%), school (68%), museums, zoos (65%), on the job (57%), and family/friends (55%) were identified most frequently as sources that were relied upon for learning about science and technology.

Social scientists have used the case study research method for many years across a variety of disciplines. Researchers, in particular, have made wide use of this qualitative research method to examine the impact of informal science environments on learning situations and provide the basis for the application of ideas and extension of methods (e.g., Storksdieck, Ellenbogen, & Heimlich, 2005). Case study research is an exploration of a 'bounded system' or case over time through detailed data collection involving multiple sources of information rich in context. The object of case study research may be a program, event, or an activity involving one or more individuals rather than a group (Stake, 1995). Content analysis of interviews, observations and responses from open-ended survey questions is used to provide an in-depth description of the phenomenon under study.

A case study conducted by Anderson, Lucas and Ginn (2003) investigated the use of the constructivist learning model as a theoretical framework for understanding the learning processes of visitors to science museums. They conducted an intensive case study of 12 students to study their knowledge transformations resulting from a free-choice science center visit, subsequent classroom-based activities and other personal experiences that the students had at home and during their discretionary time. An in-depth analysis of pre and post-visit semistructured interviews and concept maps suggested that the constructivist learning paradigm was useful for identifying a

number of complex knowledge transformations occurring as a result of the students' science center experiences.

Leinhardt, Tittle & Knutson (2000) investigated the use of diaries as a tool for documenting the learning experiences and meaning derived from museum visits. Participants in this study were 15 adults who were free to visit any museum they wanted for any purpose. Each journal entry was written as soon as possible after the visit and was between three and five type-written pages in length. Participants were also asked to include pictures, cards, flyers or catalogues to support the experiences that were documented. Each diary was read five times by three different analysts and coded according to the purpose of the visit, environmental impact and cognitive processes. Results from narrative analysis of rich, thick diary entries found that the reflective passages written by participants supported the notion that deep levels of meaning derived from museum visitors' experiences are dependent on purpose, the museum environment and level of cognitive processing.

Discourse analysis has been used frequently by informal science researchers to gain a deeper understanding of the cognitive processes of individuals by studying their conversations (e.g., Falk & Dierking, 2000) during free-choice learning experiences. This research method requires that investigators record and analyze transcripts of verbal interactions between informal learning audiences such as those characterized by school groups engaged in after school informal science programs, conversations between parents and children, or the spontaneous dialog engaged in by anonymous participants. Osborne, Erduran & Simon (2004) studied the use of Toulmin's Argument Pattern (TAP) to increase students' understanding of scientific principles and to assess the quality of discourse indicative of deeper and more reflective scientific thinking. Two groups of 8th compared on the number of elements of the TAP framework they used to argue for a scientific claim.

The experimental group (TAP group) in this study was taught to engage in collaborative discourse during science lessons characterized by interactions that supported or refuted a claim based on data, warrants, backings and rebuttals. The comparison or control group of 8th grade students was taught the same science lessons by the same teacher but without the use of the TAP model. Results from descriptive statistics found that after one year, 55% students in the argumentative dialogue group were observed to engage in discourse at the highest two levels of the argumentation framework devised by Osborne, Erduran & Simon compared to only 40% at the beginning of the year for the same group of students. In comparison to the control group, the TAP group evidenced a greater percentage of verbal interactions at the two highest levels of the TAP framework. However, these differences were not found to be significant. The authors suggest that the inability to randomly assign students to groups and the small group sizes may have been responsible for the non-significant differences.

Siegel, Easterly, Callanan, and Wright (2007) devised their own framework based on content analysis methods for analyzing discourse that occurs between families of Mexican descent when engaged in informal science learning in museums and at home. Their study focused on the causal explanations that parents from 40 families of Mexican descent supplied to their children during their investigation of the concept of buoyancy. Videotaped observations of the families were used by the

researchers to code both the content of what families discussed as well as their style of interaction. Categories of parent explanations that emerged from this analysis were: density explanations, density-relevant explanations, other property explanations and no information. The interaction style categories identified were: directive, collaborative, instructional guidance and other. Results from a mixed analysis variance comparing age of child (younger, older), child gender, Parents' schooling (Basic or Higher Schooling), Phase of task (Prediction or Testing) and Explanation Type (Density, Density-Relevant, Other property, No information) found only a significant main effect for explanation type, $F(3,32) = 56.28, p < .001$. Follow-up tests found that parents in general were more likely to give density-relevant explanations than all other types of explanations, as confirmed by follow-up t-tests (density-relevant explanations more frequent than density explanations, $t(39) = 7.65, p < .01$; more frequent than other properties, $t(39) = 7.44, p < .01$; and more frequent than no information, $t(39) = 7.56, p < .01$).

Hogan, Nastasi, and Pressely (2000) examined the depth of reflection and complexity of reasoning characterizing verbal interactions that occur among students as compared to those that occur between peers with support from teachers. Transcripts of eighth grade students taken from a science laboratory investigation studying the characteristics and phases of matter were recorded over a ten week period. Content analysis and discourse maps were used to identify common themes to describe the quality of student dialogs. In addition, a rubric was developed to assess the quality and complexity of student reasoning. Overall results from this study found that teacher-guided student groups engaged in a greater frequency of discussions coded at the moderate and high levels of reasoning complexity than student groups that engaged in discussions with only their peers.

Advancements in computer technologies have created increased opportunities for individuals to engage in a variety of in informal science learning activities (e.g., Knight-Williams Research Communications, 2005). These recent technological developments have been motivated by formal education's need to reach diverse group of learners who are restricted by time and space. Informal science learning programs have been quick to take advantage of the same Web-based technologies that have been fully embraced by public schools, colleges and universities internationally (e.g., Goodman Research Group, 2006; Gray, 2004). Online learning research has contributed a great deal to the development of methods for analyzing electronic discourse (e.g., Henri, 1992; Gunawardena, Lowe, & Anderson, 1997). Garrison, Anderson, and Archer (2000) expanded Dewey's Practical Inquiry model to create a more a framework that more thoroughly assess quality of reflection, they refer to as cognitive presence, that occurs during online discussions. The Practical Inquiry framework created by Garrison et al. provides a much more pragmatic method for assessing online inquiry related both formal and informal science learning than the previous discussed discourse analysis models.

The majority of quantitative research conducted to assess learning in informal science settings coexist with qualitative assessments of learning. Mixed-methods research designs combine quantitative and qualitative approaches by essentially mixing both quantitative and qualitative data in a single study. Combining quantitative and qualitative methods is potentially a powerful way for researchers to

more fully understand factors that influence informal science learning than is possible using either quantitative or qualitative methods alone. Falk and Adelman (2003) collected both quantitative and qualitative data to assess the conservation/environmental knowledge and interest at the National Aquarium in Baltimore. Semistructured interviews consisting of a variety of open-ended and closed questions, as well as detailed field notes, were used to collect data related to the learning experiences and interest of 395 aquarium visitors. Of the 395 individuals completing pretest interviews, 100 also completed posttest interviews. Fifty of the individuals completing post test interviews were tracked through the aquarium by data collectors who recorded their path using a scaled map, and recorded the quality and nature of their interactions with exhibit components (i.e., number of exhibit components visitors interacted with, time spent at each exhibit, quality of interaction, quality of social interaction, facilitation by staff or volunteer). Paired t-tests revealed that across all 100 visitors, there were significant increases in both conservation knowledge ($t_{(98)} = 4.14, p < .000$) and interests ($t_{(98)} = 11.13, p < .000$). Qualitative interview data were categorized by common themes and coded. These results revealed that for entering visitors, thirty-nine percent of visitors were characterized as having minimal knowledge, 55% had moderate knowledge and 6% had extensive knowledgeable. In addition, 14 % indicated minimal interests, 55% expressed moderate interests and 31% showed extensive interests about environmental or conservation issues. Data from interviews immediately after completion of the aquarium visit revealed a different pattern of conservation related knowledge and interest. Analysis of emergent themes found that, 32% of visitors were characterized themselves as having minimal knowledge, 32% indicated moderate knowledge and 33% expressed extensive knowledgeable. In addition, 4 % indicated minimal interests, 32% showed moderate interest and 64% expressed extensive interests about environmental or conservation issues. Overall, aquarium visitors' perceptions of their conservation knowledge and interests were observed as a result of their opportunity to view the exhibits displayed at the National Aquarium.

The research presented in this section is a sampling of the contemporary approaches that researchers have used to assess learning within the context of informal science environments. The complex interaction of external and internal variables that characterize informal science learning environments suggests that multiple research methods which collect multiple sources of data are superior to single method studies for investigating the nature and depth of learning that occur in informal science environments. The mixed-methods approaches recommended by the National Science Foundation provide good guidance to researchers interested in producing sound research outcomes that are well-suited for informing decision-makers about the important factors that influence learning in informal science environments.

Part 3 What We Have Learned

Committee's Focus Question: What do we think we have learned about the particularities of assessing informal learning?

Our analysis is governed by two important factors. First, the diverse nature of informal contexts including the variety of participants, settings, pedagogy and space/time relationships. Second, the complexity of assessment and evaluation as they pertain to methodologies and data acquisition and analysis methods that are influenced by the nature of what is being studied, how it is studied and why researchers engage in their studies.

3.1 The diversity of the informal experience

On the one hand, we might say that there is, in effect, nothing different about science learning in informal settings compared with science learning in formal settings. Participants in events in both categories may or may not learn new scientific ideas or science content, specific skills, and/or develop scientific dispositions or other affective outcomes. That being said, there is a world of difference between formal and informal (not to be confused with a school/out-of-school divide), especially in the degree of difficulty associated with planning and assessing learning outcomes.

In this section of the paper, we categorize the diversity of informal experience as 'micro' (the experience of an individual); 'meso' (the shared experience of a group of people), and 'macro' (the view of experience typically associated with overall program participant outcomes). In addition, we add the 'social/cultural' level of inclusiveness to distinguish experiences that address communities of people focusing on specific ethnic group outcomes to national standardized assessment of overall outcomes for entire populations.

3.2 The complexity of assessment methodologies and methods

In order to make sense of the complex nature of assessment and evaluation endeavors, we have chosen to consider research in terms of the methodologies and methods which have been used in a series of studies. In considering our organization of methodologies and data acquisition and analysis methods, we rely on the conception of stochastic variability within a fractured spectrum of approaches. This choice is a result of our dissatisfaction with either discrete categories (qualitative/quantitative) or a continuum of approaches (descriptive to experimental). We conclude that methodologies and methods overlap discrete categories and that there is no smooth transition between approaches across a continuum. We believe this approach helps make sense of the vast array of assessments and evaluations that have been conducted in informal settings.

In addition, we have concluded that in order to make sense of the various assessment methodologies and data collection methods that researchers must take into consideration *what* is being studied (ontology), *how* it is being studied (epistemology) and *why* it is being studied (axiology). This conclusion reflects the overlap and

interrelationship of context, curriculum, pedagogy and assessment of science learning in informal settings.

3.3 Identity as a unique case of complex assessment

Identity outcomes present researchers and evaluators with levels of complexity that research methodologies and data acquisition methods must take into consideration. The literature on identity in informal learning indicates there are a wide variety of outcomes that are often associated with specific types of informal settings or activities. Regardless of the focus of the informal learning (science, literacy, mathematics, etc.) or the settings (museums, computer/internet, outdoors) identity outcomes are naturally part of the informal learning process. In a sense, we can simply consider identity outcomes to be the result of engaging in various life activities (Hodkinson, 2007). So, the question is, how do we go about assessing typical life experiences? For example, learning at work tends to promote work identities such as responsibility or independence. Outdoor settings and experiences may promote learner identity outcomes related to participants' relationship to nature (Brody, 2005) or natural resources (Brody *et al.*, 2002). Some identity outcomes associated with specific informal learning settings and activities are:

Work: self-discipline, time management, education's value, social interaction, empathy and planning;

Hobbies: identity, self-expression, discipline, social interaction, independence, leadership, test limits (little research in this area);

Media & Information Technologies: visual perception, hierarchical reasoning, communication skills, reading visual knowledge, content knowledge;

Consumerism: personal expression, choices, identity, budgeting,

Peer, leisure and family activities: exchanging perspectives, team-work, group decision-making, communication, responsibility, tolerance.

A closer look at the workplace reveals some of the key principles associated with identity and learning in informal settings. In the workplace, employees are required to learn new skills. As employees acquire these skills, they tend to redefine themselves as individuals and as part of a community. Personal and communal identity shift allowing different opportunities, motivation and participation. Values also shift, leading to changes in conceptions of entitlement and expectations for change, self-efficacy, new careers and job paths. Gray (2004) reports the development of skills and the culture of practice leading to new insight into professional identity and the reconstruction of the collective identity of the community of coordinators of adult learning programs.

The study of identity outcomes in science learning in informal environments must take acknowledge that learning is part of everyday experiences. Learning outcomes relate to cognition, skill and affective development over time and space. Outcomes are typically intertwined and appear dependent on each other. Identity outcomes are dependent on the types of settings, experiences and motivations. Research methods must address the unique aspects of individual, activities and settings.

3.4 Analysis of strengths and shortcomings in assessing science learning in informal environments

Having looked at a wide range of studies addressing the outcomes of science learning in informal settings, we have selected a number of studies to analyze in terms of their methodologies and methods. The studies we selected are positive representations of a variety of methodological approaches. The studies were chosen to represent the body of literature that exists at the time of writing. It is not meant to be numerically representative of the breadth or diversity of the field nor is it an exhaustive selection. For each study we identify the What?, How? and Why? The selected literature is coded and listed on the following page.

Code	Citation
1	Allen, S. (1997). Using scientific inquiry activities in exhibit explanations, <i>Science Education</i> , 81(6), 713-34.
2	American Institutes for Research. (2006, August). <i>Evaluation of FETCH! Activity guide</i> . Concorde, MA: Author.
3	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, <i>Science Education</i> , 84(5), 658-79.
4	Ash, D. (2003). Dialogic inquiry in life science conversations of family groups in a museum, <i>Journal of Research in Science Teaching</i> , 40(2), 138-162.
5	Brody, M. (2005). Learning in nature, <i>Environmental Education Research</i> , 11(5), 603-21.
6	Brody, M, Tomkiewicz, W. & Graves, C. (2002). Park visitor's understanding, values and beliefs related to their experience at Midway Geyser Basin, Yellowstone National Park, USA, <i>International Journal of Science Education</i> , 24(11), 1119-41.
7	Crowley, K., Callanan, M.A., Jipson, J.L., Galco, J., Topping, K. & Shrager, J. (2001). Shared scientific thinking in everyday parent-child activity, <i>Science Education</i> , 85(6) 712-32.
8	Edu., Inc. (2004, June). It's a nano world: A study of use. Findings from a summative report. Ithaca, NY: Author.
9	Falk, J., Moussouri, T., & Coulson, D. (1998). The effect of visitors' agendas on museum learning, <i>Curator</i> 41, 107-20.
10	Falk, J.H. & Adleman, L.M. (2003). Investigating the impact of prior knowledge and interest on aquarium visitor learning. <i>Journal of Research in Science Teaching</i> , 40(2), 163-76.
11	Gerber, B.L. (2001). Relationships among informal learning environments, teacher procedures and scientific reasoning ability, <i>International Journal of Science Education</i> , 23(5), 535-49.
12	Goodman Research Group, Inc. (2006, May). <i>Summative Evaluation of Einstein's Big Idea</i> . Cambridge, MA: Author.
13	Gray B. (2004) Informal learning in an online community of practice, <i>Journal of Distance Education</i> , Spring, 19(1), 20-35.
14	Hodkinson, P., (2007) Learning as Being, Lifelong Learning Institute, University of Leeds, Leeds, Great Britain.
15	Jackson, A. & Leahy, H.R. 'Seeing it for real...?' Authenticity, theater and learning in museums, <i>Research in Drama Education</i> , 10(3), 303-25.

16	Knigh-Williams Research Communications. (2005, August). <i>Summative Evaluation of National Geographic's Strange Days on Planet Earth Television Series and Website with an Adult Audience</i> . Sacramento, CA: Author.
17	Lebeau, R.B., Gyamfi, P., Wizevich, K., & Koster, E.H. (2001). Supporting and documenting choice in free-choice science learning environments. In J.H. Falk (Ed.), <i>Free-choice science education: How we learn outside of school</i> (pp. 133-48). New York, NY: Teachers College Press.
18	Leinhardt, G., Tittle, C., & Knutson, K. (2000, October). Talking to oneself: Diaries of museum visits. Retrieved December 29, 2006 from http://mlc.lrdc.pitt.edu/mlc .
19	Lynch, H. (2007). Learning as Desire, paper presented at ESREA (European Society for Research on Education of Adults) Network Conference on Life History and Biography March 1st-4th 2007 Roskilde University, Denmark
20	Meisner, R., vom Lehn, D., Heath, C., Burch, A., Gammon, B., & Reisman, M. (2007, forthcoming). Exhibiting performance: Co-participation in science centres and museums, <i>International Journal of Science Education</i> .
21	Miller, J.D. (2001). The acquisition and retention of scientific information by American adults. In J.H. Falk (Ed.), <i>Free-choice science education: How we learn outside of school</i> (pp. 93-114). New York, NY: Teachers College Press.
22	Randi Korn & Associates, Inc. (2002, April). <i>MarsQuest Summative Evaluation</i> . Alexandria, VA: Author.
23	Siegel, D., Esterly, J., Callanan, M.A., & Wright, R. (2007, forthcoming). Conversations about science across contexts in Mexican-descent families, <i>International Journal of Science Education</i> .
24	Storksdieck, M., Ellenbogen, K. & Heimlich, J.E. (2005). Changing minds? Reassessing outcomes in free-choice education. <i>Environmental Education Research</i> , 11(3), 353-369.
25	Tenenbaum, H.R., Rappolt-Schlichtmann, G. & Zanger, V.V. (2004). Children's learning about water in a museum and in the classroom, <i>Early Childhood Research Quarterly</i> , 19 (1), 40-58.

3.5 Analysis of methodological and data acquisition methods

Each of the studies listed above was categorized by the authors in relation to the research methodologies and methods of data acquisition used in the assessment or evaluation of science learning in informal contexts. We begin with an explanation of the operational definitions of each of the categories listed in the matrix.

a) Methodological approaches

The assessment matrix represented in Figure 1 is an alphabetical listing of research methodologies and data collection methods. No attempt is made to represent these methodological descriptors as qualitative/quantitative dichotomies or as a continuum of approaches based on complexity. The current form of our proposed assessment framework represents the stochastic variability that characterizes the spectrum of research methodologies and data collection methods that are available to informal science researchers.

To better understand the rationale for the research methodologies represented in the assessment matrix, we need to operationally define our approaches. We can think of the identified research methodologies as representing related research approaches that are used to achieve similar purposes. For example, the experimental methodologies are primarily undertaken to make group comparisons. Designs related to this group of approaches include randomized, quasi-experimental, causal-comparative, pretest-posttest, and single-subject designs. Determining the magnitude of relationship among informal science dependent variables and their influence on one another is the goal of correlational approaches. These methods include multiple regression, discriminant analysis, canonical correlation, path analysis, factor analysis, structural equation modeling and hierarchical linear modeling. Descriptive research approaches on the other hand are undertaken to explain current perceptions, attitudes, behaviors and, in some cases, knowledge of participants engaged in informal science contexts that are designed to promote learning the last bit is rather confusing. Outcomes from analysis of data collected from surveys, interviews and observations are represented by frequencies, percentages and means.

Researchers participating in action research approaches to better understand informal science contexts, use a diverse range of assessments for acquiring data about learning in informal learning contexts. It would not be unusual for action researchers to assess learning by using observations, interviews, surveys and formal tests that tap knowledge, skills and understanding. However, the key feature of action research that sets it apart from other methodological families is in its purpose, that is, to generate outcomes that can be used to improve strategies, practices, and knowledge of the environments within which professionals practice. In comparison to experimental designs, for example, action research is participatory and generates new questions rather than final outcomes for practioners to consider as they engage in a cycle of continuous improvement.

Case study approaches also rely on muliptle assessment methods but are more directed at generating in-depth descriptions of the activities which characterize 'bounded systems' that are uniquely defined by a program, event or activity. Ethnographic studies, on the other hand, use multiple data acquisition methods to describe and improve

understanding of programs, events or activities for entire cultural systems or subsystems. Biographical and historical approaches are undertaken to describe a chronology of events. The life story of Jack Horner, a famous paleontologist, could be written using biographical research approaches. However, a historical approach would be better suited for investigating the chronology of events leading up to the establishment of the Smithsonian museums. Both approaches, however, rely on assessment methods that involve in-depth interviews, dialogs and review of records. Whereas a biography reports the life events of a single individual, phenomenological approaches describe the meaning of the everyday lived experiences for several individuals about a concept or phenomenon (Creswell, 1998). Interactions with museum exhibits or outdoor laboratories are examples of the types of informal science learning experiences that could be studied with phenomenological methods. Describing the unique, yet common, experiences of individuals or groups of individuals participating in informal environments could improve our understanding of variables that have both positive and negative impacts on informal science learning.

b) Data acquisition methods

The assessment matrix lists a diverse range of methods that can be used by informal science researchers to collect data that can be analyzed for the purposes defined by their research methodologies. Again, these data acquisition methods are listed alphabetically to prevent the assumption that one type of data is confined to one methodological approach. Personal meaning maps, concept maps and work samples supply assessment data in the form of outcomes represented by unique, individual products. For example, personal meaning and concept maps are visual representation of an individual's understanding of an abstract idea such as the notion of global warming or the extinction of dinosaurs. In comparison, work samples are more readily defined as projects, artistic productions, portfolios, laboratory activities, writing and other products that individuals produce that demonstrate applications of complex knowledge, skills and problem-solving abilities. In contrast, paper-and-pencil knowledge tests are formal types of classroom or standardized assessments that use fixed-choice response questions to assess discrete knowledge and skills in one or more specific content areas.

Surveys and focus group interviews are directed at assessing perceptions, attitudes, beliefs and knowledge through the use of both fixed-choice and open-ended response items. Although, fixed-choice questions are well-suited for providing quantitative summaries of survey data, open-ended questions are important for obtaining in-depth responses that support the narrow outcomes offered by objective item formats. For example, it is important to know that museum visitors prefer one exhibit over another, however, it is equally important to know the specific details that influenced their overall preferences so that museum curators, for example, can make informed decisions about current and future displays.

Observations, videos and photographs are often used to assess individual and group behaviors that occur within informal science settings. Analyzing behaviors using these approaches is important for understanding how individuals and groups of individuals interact with each other to expand their learning within free-choice environments. Journals, 'think-aloud protocols' and recorded dialogs can be used to assess

both the internal and external conversations individuals engage in when participating in a free-choice activity. These data acquisition methods provide a record of participant conversations that can be examined through discourse analysis to describe the the extent of learning experiences that occurs in informal science settings.

c) Matrix of methodologies and methods

The key purpose for creating the matrix (Figure 1) was to get a sense of the ‘landscape’ of research methodologies and methods that exist within the sample of articles we chose to analyse. The selection of research methodologies and data acquisition methods is critical for effective research and assessment. Figure 1 shows that 10 of the 25 studies are descriptive and, within this group, 4 studies use single method and 6 multi-method approaches. Most of the studies rely on qualitative methods and a number of research methodologies are not represented. In general, we believe that the dominant qualitative descriptive orientation reflects consideration of the many unique qualities of informal settings. The descriptive assessment strategies may indicate an emphasis on learning processes rather than specific learning outcomes.

		METHODS OF DATA ACQUISITION														
R E S E A R C H M E T H O D O L O G I E S		Concept Maps	Dialogue	Focus Group	Interactions	Interviews	Journals	Observation	Paper-Pencil Test	Photographs	Surveys	Talk aloud	Task Analysis	Videos	Work Samples	Tracking Software
	Action Research															
	Biographical															
	Case Study					5, 14					24				5	
	Correlational	9				9		9			9, 21					
	Descriptive/ Interpretative	3	7,13		7,23	1, 3, 6, 15, 19	18	15, 23	11		11, 13, 17			3, 7	3,15	
	Ethnography/ Ethnomethodology		4		4, 20,	4		4, 20,						4, 20		
	Experimental					2, 8, 12, 16, 22, 25		2, 12, 10, 22	16		2, 8, 10,12, 16			2		8,22
	Historical															
	Phenomenological															

Figure 1. Matrix of Assessment Strategies that Hold Promise for Assessing Learning in Informal Setting

Table 1. Alignment in Studies of Learning Science in Informal Environments

ID	Citation	Who (inclusiveness of sample)	What (subject of study)	How (data acquisition methods)	Why (purpose of the study)
1	Allen, 1997	Micro: individual visitors	Effect of inquiry activities within hands-on exhibits on understanding	Individual interviews	Effectiveness of Inquiry activities within exhibits
2	American Institutes for Research, 2006	Micro/Meso: individual 4 th grade students from 10 after school programs assigned to a treatment or control group condition.	The effects of the FETCH! public TV series supplemental program activity guide on the science/engineering inquiry knowledge of 75 4 th grade students.	Pretest and posttest questionnaires and interviews.	Evaluation of the public TV series FETCH! supplemental program activity guide.
3	Anderson <i>et al.</i> , 2000	Micro: 28 school students	Influence of post-visit activities in the classroom on subsequent learning and knowledge construction	Student-generated concept maps; semi-structured interviews at three stages of the study	Theory building: the relationship between learning in school, home and science centre
4	Ash, 2003	Meso: 3 Families	How do families interact with one another? Discussions about biological adaptation	Pre- and post-visit interviews; Video and/or audiotape of visits. Analysis of segments of talk.	Theory testing: application of theories of learning in informal contexts.
5	Brody, 2005	Micro: single case study	How does theory explain this case?	Work samples, field notes and college essay and interview over three years	Theory building
6	Brody <i>et al.</i> , 2002	Micro/Meso: individuals and small groups	Nature of experience at site and affect on understanding and values	Interviews before and after experience at site	Theory building: learning in informal settings Evaluation of visitor materials
7	Crowley <i>et al.</i> , 2001	Meso: small group family	Collaborative scientific-thinking in parent child interactions	Video of interactions	Study the development of scientific literacy
8	Edu., Inc., 2004	Micro: Individual visitors	The effects of the 'It's a NanoWorld' nanotechnology exhibit on visitors learning and awareness of nanotechnology concepts.	Pretest and posttest questionnaires and interviews.	Evaluation of the 'It's a NanoWorld' traveling exhibit.
9	Falk <i>et al.</i> , 1998	Micro: 40 individual visitors	Effect of visitor agenda (motivation & strategy) on learning	i) Tool for measuring visitor agenda (survey; Likert-type) ii) Personal Meaning Mapping	Theory building: effect of visitor agenda on visitor learning

10	Falk & Adelman, 2003	Meso: 100 aquarium visitors	Effect of visitor type (prior interest and knowledge) on conservation related knowledge and interest.	i) Open and closed ended Interview questions ii) Concept Maps iii) Individual Observations	To determine if grouping aquarium visitors by prior knowledge and interest would provide a more accurate of assessment of conservation knowledge and interest
11	Goodman Research Group, Inc., 2006	Micro: Individual viewers and Website visitors.	Effect of a NOVA public TV series Einstein's Big Idea and Website on viewers understanding Einstein's theory of relativity, the important contribution made by other scientists and an awareness of the historical events surrounding the development of this important scientific discovery.	Pretest and Posttest surveys paper pencil and web-based surveys.	Evaluation of the NOVA TV series Einstein's Big Idea and website.
12	Gerber, 2001	Meso: 1,178 students in 7 th , 8 th , 9 th , and 10 th grade.	The effect of student informal science experiences (impoverished vs. enriched) and teaching procedure (inquiry vs. non-inquiry) on science learning and skill development	i) Survey (open and closed ended questions). ii) Knowledge test – constructed response items	Investigated the impact of students' experiences in informal learning environments on science learning and/or skill development.
13	Gray, 2004	Meso: 43 coordinators	To what extent is this experience a community of practice? What is the nature of informal learning?	Review of online forum postings, live chat transcripts, email correspondence, survey and interviews	Theory testing, understand online communities of learners, personal development.
14	Hodkinson, 2007	Micro: 120 adults	What are the relationships of identity and agency in people's lives?	120 subjects interviews over three years and correlation to national survey	Theory building. Describe the relationships of learning, identity and agency
15	Jackson & Leahy, 2005	Micro: individual children	Effect of theater in museum on children's learning	Semi-structured small group interviews, drawing and creative writing	If and how a museum experience affected children's cognitive and affective learning

16	Knigh-Williams Research Communications , 2005	Meso: individual children assigned to a control and treatment group.	Effects of a public the Strange Days on Planet Earth TV series and companion Website on adult viewers learning of environmental concepts and issues.	Pretest and Posttest surveys, paper and pencil knowledge tests and follow-up phone interviews.	Evaluation of the <i>Strange Days on Planet Earth</i> National Geographic TV series and website.
17	Lebeau <i>et al.</i> , 2001	Study 1: Meso: 27 (divided into 3 study groups). Study 2: No information	Study #1: To determine the effects of maps and worksheets on help-seeking behaviors and science learning. Study #2: To describe where museum visitors decide to go (what exhibits they decide to view) and what transitions and connections they make.	Surveys (Likert scale using a Strongly Agree to Strongly Disagree format)	Study #1: To investigate the effects of informed-goal setting on help-seeking and perceptions of science learning. Study #2: To investigate the use of observers traveling with groups to describe museum visitors planned movements
18	Leinhardt <i>et al.</i> , 2000	Micro: 8 individual museum visitors	To describe how diaries were used to describe museum visitors' experiences and to understand the cognitive tools they used to create meaning.	individual diaries	To study the influence of diaries on museum visitors learning.
19	Lynch, 2007	Micro: 120 adults	How does desire affect the creation or rejection for learning and change in people's lives?	120 subjects interviews over three years and correlation to national survey	Theory building, role of desire in learning life
20	Meisner <i>et al.</i> , 2007	Micro & Meso: Individual visitors and small groups	The extent to which the design of exhibits enables particular forms of co-participation or shared experiences	Detailed transcription of short fragments of video – single instances of visitors' performative activities – including talk and bodily comportment	Theory development/design considerations: to develop design sensitivities that exhibition managers and designers may consider when wishing to engender novel ways of engagement and participation with and around computer-based exhibits
21	Miller 2001	Macro: 2,000 Adults over age 18	The relationship of informal science learning to formal science learning.	i. Survey questions of knowledge (open and closed).	To support the supposition that effective formal science learning is an essential prerequisite to the utilization of free-choice learning

22	Randi Korn & Associates, 2002	Meso: 128 adult and children visitor who were grouped according exhibit venue and compared.	The effects of the MarsQuest exhibit on visitors understanding of concepts related to Mars, space travel and planets.	Observations and exit interviews	Evaluation of the MarsQuest exhibit.
23	Siegal <i>et al.</i> , 2007	Meso: 40 family groups	Do children from different family backgrounds have different experiences talking about science in informal settings?	Videotaped interactions of parents and their children during a semi-structured hands-on science task.	Theory building: bridging children's learning environments – home, school, and museum, particularly in traditionally under-represented groups.
24	Storksdiel <i>et al.</i> , 2005	Meso: Three Informal Science Environments.	To describe visitor learning approaches and gains. Secondly, to describe how visitor characteristics influenced learning outcomes and values.	i) Concept Maps (evaluated on a Likert scale) ii) Card sorts (Likert scale) iii) Survey (Likert Scale)	To describe the influence of informal science context and audience on learning.
25	Tenenbaum <i>et al.</i> , 2004	Micro: children	Effect of museum and classroom instruction on children's content knowledge and concept complexity	Pre and post interviews	Development of science learning and dispositions

Citation: Author(s) and date of publication.

Who: The inclusiveness of the study sample as in micro (individual), meso (small group), macro (program), socio-cultural (community) or political (state/national).

What: The ideology of the study or the nature of what is studied.

How: The epistemology of the study or the nature of methods used.

Why: The axiology of the study or the nature of why things are being studied.

d) Alignment of studies (What?), methods (How?) and purpose (Why?)

According to our framework for review and analysis of the selected articles, we are also concerned with the What? (ideology), How? (epistemology) and Why? (axiology) of these studies, and how they contribute to the appropriateness and effectiveness of the assessment of science learning in informal environments. In order to address this issue, we have constructed Table 1 which summarizes the What, How and Why of the selected literature. We also consider these factors in relation to the inclusiveness of the study, that is, the nature of the sample and, ultimately, the unit of analysis. These are the micro, meso, macro and social-cultural categories.

What does the table tell us? In general we can say that when the researchers are clear in describing the sample studied and the reasons for the investigations, there is alignment with the assessment methodologies and data acquisition methods. In most cases the alignment has helped produce effective assessment of learning outcomes.

From the table we are also able to make claims about the theoretical perspectives of the studies, that is, the relationship between the purpose of a study, the theoretical background and the conceptual framework. In the table, we categorize this perspective as theory driven, theory testing or theory building. From the table, we can see significant reference to theories and how they guide methodologies and outcomes. In a few cases, we see actual conclusions drawn about theories of learning in informal settings.

3.6 A critical review of a selection of studies of science learning in informal contexts

What follows is a description of each of the research studies we chose for this analysis. Within the analysis is our judgment of what assessment strategies matched up to the unique qualities of science learning in informal environments. In some cases we describe how these strategies might better address those unique characteristics. The specific qualities that we consider are:

- event centered
- idiosyncratic experiences
- personal/social interactions
- facilitated learning
- emergent experiences
- thinking feeling and acting
- expected/unexpected outcomes
- variance in space and time
- multi-sensory experience
- identity outcomes
- 'free-choice' learning
- physical context
- social context
- motivation, curiosity, interest, expectations
- interactive experience

- unpredictable experience
- multiple entry points (intelligences)
- precursors to engagement, engagement, post engagement
- unique physical setting
- everyday informal, TV, computer, newspapers
- cultural dimensions
- cumulative process
- connecting other experiences
- focus on process of learning
- prior knowledge
- socio-cultural
- physical, personal, social and time
- complexity and variability in setting and audience
- physical complexity
- diversity of audience

We also considered these questions in the analysis:

- How does the study reflect assessment practices that are congruent with the unique characteristics of assessing science learning in informal environments?
- What elements of the study were effective?
- What did the study do and not do?

Allen, S. (1997). Using scientific inquiry activities in exhibit explanations, *Science Education*, 81(6), 713-34.

This study looks at science inquiry activities within exhibits using a descriptive/interpretive methodology involving interview data acquisition techniques. Interviews involved questions and answers and hand drawings of explanations and interpretations of actual exhibit phenomena. This method links the assessment data acquisition to the event-centered, interactional and physical contexts of informal settings. The choice to use a descriptive research methodology seems to be related to the desire to 'focus entirely on individual visitor cognition' (p. 717) which is addressed in think-aloud techniques. Conducting the interviews at exhibits takes into account the physical setting of the events and is congruent with the interactive nature of informal settings.

The researcher's goal for this study was to improve 'science learning in formal as well as informal settings' (p. 716) which places the emphasis of the study on science learning rather than the informal context. Within science learning, the research focuses on scientific inquiry in hands-on museum settings (p. 717). The researcher does take into account the unique qualities of informal settings in the choice of assessment strategies. The other informal characteristic considered is the exhibit, and the study does focus on inquiry within exhibits. The author does 'acknowledge that inquiry is not the only nor necessarily the most important part of exhibits' and that assumptions about exhibits govern the research (p. 716).

In terms of the research questions, the study focuses on the nature of learning. There are four specific research questions (p. 716) which focus on which experiences are more effective, how challenging they are, on the nature of understanding after experience and on the likelihood of revision of conceptual models.

In the conclusions of the study the researcher does not mention the unique qualities of the informal setting. This absence indicates that the focus on science learning and especially inquiry is the main point of the study that might reflect the goal of the study on both formal and informal settings. The word 'museum' is used once in the conclusions in the context of time difficulties and further research. No conclusions are drawn about learning in informal settings (pp. 732-33).

The researcher focused on the unique function of the institutional setting in relation to hands-on inquiry-oriented science learning. While a comprehensive and well-executed study of science inquiry learning in the descriptive family of research, the study could more fully address the unique aspects of experiences in informal settings.

**American Institutes for Research (2006). *Evaluation of FETCH! Activity Guide*.
Concorde, MA: Author.**

FETCH is a competition-based reality TV program for 6-10 year old children that includes animation and live action scenes. Participants are challenged to leave the studio and complete challenges that require the use of science and engineering to answer a question or design a solution to a problem. In each episode, some of the contestants complete challenges while others observe the challenges in the studio and participate in a 90-second game in which they answer questions about what happened. Questions are designed to encourage audience participation and reinforce science concepts featured in the show.

The American Institutes for Research (AIR) gathered data to evaluate the Fetch! Activity guide which was designed to extend the teaching of the TV show. AIR used a longitudinal, pre-/post-test, control and treatment group, experimental design. Children in the control group were exposed to non-science (arts and crafts) activities, while children in the treatment group were exposed to six FETCH! activities in the guide. The participants in study consisted of ten after-school programs in the greater Boston metropolitan area. Five of the programs (two treatment, three control) were in urban locations, while the other five were in suburban towns. Ten facilitators and a total of 75 children participated (45 in the treatment group and 30 in the control group). All the children were in the 3rd, 4th or 5th grades and represented different ethnic groups.

Results from pre-and-post test questionnaires and interviews indicated that 91% of children 'liked or loved' the activities while 93% reported that the activities were good for children of their own age. 95% of the children thought their friends would enjoy the FETCH! activities. Ratings of the Activity Guide by the after-school program facilitators was very high. Their average rating for the educational value was 3.21 on a scale of 1-4, with 4 representing the highest rating. The average facilitator rating for the entertainment value of the FETCH! activities was 3.28 on the same scale. Children in the treatment group showed significantly greater gains in science content than those in the control group.

The strength of this study, from a funding perspective, is that the evaluators used a rigorous, quantitative research design to assess impact of the Activity Guide on students' learning and attitudes. For example, the researchers controlled for prior science and engineering knowledge of students when making control and treatment group comparisons. Although the diversity of the student audience was mentioned, assessment outcomes were not disaggregated by ethnicity or gender preventing evaluation of the differential impact that the activities might have had on different student subgroups. In addition, this evaluation study is heavily focused on quantitative outcomes and does not use the 'voices' of the students or facilitators to describe in more detail the specific learning outcomes experienced or changes in attitudes toward science fostered by the Activity Guide. Detailed observations and recordings of student interactions with one another and the facilitators during the after-school activities could also have supplied evaluators with data that could have been used to more thoroughly assess learning.

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), 658-79.

This article reports part of a larger study into how 11- and 12-year-old students construct knowledge about electricity and magnetism. The research draws on aspects of the students' experiences on a school visit to Sciencentre, an interactive science museum in Brisbane, Australia. The objectives of the study reported were (1) to describe the experiences of two Year 11/12 year-old students during their visit and subsequent participation in classroom activities, (2) to report understandings of the construction of the two students' knowledge about the nature of electricity and magnetism, and (3), to reflect on the implications of the two case studies for classroom teachers, students, museum educators, and the science education community at large.

The study examines the impact of structured pre-, during and post-visit activities on students' museum learning. In terms of the study's methodology, the authors used a hermeneutic cycle approach (Guba & Lincoln, 1989) characterized by the repeated feedback of researcher perceptions to the participants for the purposes of checking, elaborating, and modifying at key stages in the progress of the research. Twelve students, out of a whole class visiting the science center, completed concept maps (students were trained how to construct maps) and were interviewed prior to their museum experiences, after their museum experiences, and after their participation in subsequent post-visit activities. Eight of the 12 were given radio-microphones which recorded their conversations during the visit and the post-visit activities. Other data sets included student worksheets and field notes made by the researchers during the visits. This paper reports on just two of the students.

Prior to the visit, students were shown a presentation to familiarize themselves with the center. During the visits, students spent about 30 minutes in the gallery interacting with the exhibits which were 'stand-alone, hands-on, and phenomenon-based, with little context or no contextual links to real-world application of the scientific principles that they attempted to demonstrate.' Post-visit, the students undertook two activities: firstly, describing their involvement with the exhibits and providing an explanation of how they believed the exhibits worked. Secondly, the students engaged in open-ended practical experiments similar to those they saw at the exhibition.

The authors reported that the activities (pre-, during and post-) 'resulted in students constructing and reconstructing their personal knowledge of science concepts and principles represented in the Sciencentre exhibits.' (p. 677). The two students in the study 'had their knowledge in the domain of electricity and magnetism transformed in many ways not specifically intended by those who planned the exhibits and/or post-visit activity experiences.' (p. 677). The key finding, in terms of this paper, is that 'Many of these changes were of a form that would probably not be detected by traditional classroom-based instruments typically used by teachers to assess student knowledge.' (p. 677). Secondly, pre- and post-visit 'interventions' [seemed] sometimes [to transform] knowledge in both *correct* and alternate ways, despite the best intentions of exhibit designers and the planners of the post-visit activities to provide experiences that would help facilitate knowledge construction in ways consistent with the accepted view of science.' (p. 678).

Ash, D. (2003). Dialogic inquiry in life science conversations of family groups in a museum, *Journal of Research in Science Teaching*, 40(2), 138-62.

The paper takes as its premise the fact that little is known about how families interact during visits to museums, even though they make up half of all visitors. The author claims to ‘demonstrate a new methodological tool that allows a fine-grained analysis of collaborative scientific sense-making, based on family conversations.’ (p.138). Referring to the interactions as ‘dialogic inquiry’, and situating the study in a Vygotskian, sociocultural framework, the author focuses ‘specifically on two aspects of family dialogue: the thematic content that underpins conversations about life sciences and the inquiry process skills [...] that advance or hinder dialogue.’ (p.138).

As the author points out, looking at talk in classrooms is not new ‘but it is relatively new to informal learning research settings which, one can argue, offer a richer context and more free-choice learning opportunities.’ In terms of further justification for the likely success of the study, the author notes that: ‘because museums are rich sources of artifacts, people, gestures, and potential dialogic interactions [...] dialogic inquiry as instruction can take place at any particular exhibit as parents interact with their children, each other, and artifacts.’ (p.139). This perspective is relevant across a wide range on informal contexts.

Part of a larger study, the paper reports on work undertaken at an exhibition on frogs at the San Francisco Exploratorium. Following a pilot study, three families were selected ‘on the basis of appropriate configuration (parents with children from ages 1 – 9 years), their availability, and their interest over time.’ (p.143). As the families moved, at will, through the exhibition, they were video-taped and/or audio-taped. The visits ranged from 43 minutes to 1 hour 25 minutes. Audio-recorded conversations at and between exhibits, as well as pre- and post-visit interviews provide the datasets. In the semi-structured pre-interview, ‘families were asked about their museum-going patterns, their reasons for coming to museums, their goals or expectations for this visit, their interest in the life sciences, and about general areas of interest in the sciences.’ (p.143).

Using Vygotskian theory, a new method of analyzing naturally occurring family dialogue was devised. The author chose ‘an intermediate level of analysis [...] in which each segment [of discourse] provided just enough data to inform the reader about the participants, overall activity, types of mediation within the event, thematic content, and inquiry skills. These representative dialogic segments (RDSs) are typical of the larger conversation; every visit contains many such events’ (p.144). The RDSs were analysed in terms of their context, the evidence of inquiry skills and the development of ideas (using the ZPD concept).

Families were described as using ‘complex negotiating processes’ during their visits (p.153). ‘The understanding that members achieved was a product of several factors including the family thematic agenda, the museum’s thematic agenda, and the inquiry skills used to talk about them’ (p.153). The author noted ‘functional similarities in how families use content themes to make sense of exhibits’ (p.153). This research approach suited for assessing informal science learning it focuses attention on language as the mediator of complex interactions suggesting that RDSs allow for analyzing co-construction of meaning over time and across contexts.

Brody, M. (2005). Learning in nature, *Environmental Education Research*, 11(5), 603-21.

In this paper the author traces work in developing theory of learning to explain how people learn in natural settings. The intellectual roots of the theory in learning in informal contexts, cognition, affective development, experiential and meaningful learning are described and synthesized into a comprehensive theory of learning in nature. The evolution of the theory is traced through several iterations describing the addition of organizing concepts as well as efforts seeking simplicity in its conceptualization.

An in-depth case study of meaningful learning about bogs (fresh water wetland) illustrates how the theory can be applied to explain environmental learning experiences. Multiple data acquisition methods included work samples of ecological field-notes and written college application essay and an in-depth interview. Data was collected over a three-year period following the learning activity. The implications for this theory in terms of environmental education are then discussed.

The motivation for the paper is based on prior identification of a lack of theoretical underpinnings to inform empirical research in the field of environmental education and especially in outdoor settings. The paper is clearly presented in the context of theory building with the author building upon past developmental conceptualizations.

The author claims that in the perceived absence of a comprehensive theory of learning to guide the conceptual and methodological aspects of environmental education research and practice, the theory of learning in nature proposed provides a framework upon which to start building a more robust theoretical perspective on EE activities. The theory takes into account several of the major theoretical perspectives on learning that inform work in other domains such as science learning in informal settings. The theory also takes into consideration the unique aspects of EE outdoor activities. Drawing from Falk and Dierking's work in informal settings, the author builds upon their conceptualization of direct experience, personal and social construction of knowledge over time.

The theory is presented as explanatory, intended to inform and guide EE research and practice. The case study of learning and the bog confirms that the theory of learning in nature provides a coherent explanation of the cognitive, skill and affective development that accompanied the experience and subsequent learning. This research article fills a significant gap in the research literature by simultaneously assessing learning in multiple domains and directly addressing the conceptual development of theory, principles and concepts to help guide future work in learning science in an specific informal setting.

Brody, M, W., Tomkiewicz, W. & Graves, J. (2002) Park visitors' understanding, values and beliefs related to their experience at Midway Geyser Basin, Yellowstone National Park, USA, *International Journal of Science Education*, 24(11), 1119-41.

The study of Yellowstone visitors' experiences at Midway Geyser Basin looked at how individuals and small groups made sense of their experience. The fact that the study took place at the site of the geyser basin ensured the emphasis on the physical setting and the direct experience of being there. In fact, visitors gazed upon the geologic features and organisms, smelled the sulfuric emissions, heard the eruptions and felt the steam on their faces as they described their experience. The study was responsive to multiple entry points for learning because the experience and data acquisition method were multi-sensory involving senses, speech, reading, text, visual and multiple personalized approaches within social groups.

This descriptive study used semi-structured interviews at the beginning and end of the experience. Based on the needs of the funding organization, an assessment of initial understanding and final outcomes was undertaken. This method recognizes the emergent and idiosyncratic nature of learning from the experience. Interviews took into consideration precursors to the experience of walking around the basin including the motivation for being there and events leading up to the moment. Interviews were sensitive to prior informal learning, especially in relation to media information about the geothermal features and associated microorganisms. The focus was on both the process of learning and outcomes taking into consideration prior knowledge and the socio-cultural context of the visitor interviewees.

This study was primarily interested in the nature of the visitors' experience and how that affected understanding and values related to geothermal features and associated microorganisms. In the case of parks and other outdoor informal learning experiences the focus on the experience of visitors within the geyser basin is clearly in line with the spontaneous and event-centered nature of informal learning settings.

Findings of the study indicate a number of relevant concepts that link visitor prior knowledge with intended cognitive outcomes of the visitor experience. The development of several affective outcomes including openness to new scientific ideas and conservation of natural resources for future generations were identified among the sample of park visitors. In general the findings are congruent with the authors' proposed synthesis of learning theories into a comprehensive view of learning science in informal settings. Overall the study had good alignment of the sample, research question and methods, informed by and based on the purposes of the assessment.

Crowley, K., Callanan, M.A., Jipson, J.L., Galco, J., Topping K., & Shrager, J. (2001). Shared scientific thinking in everyday parent-child activity, *Science Education*, 85(6), 712-32.

This study describes the development of scientific reasoning through parent-child relations. It employs a descriptive/interpretive research methodology with interaction analysis, dialogue analysis and video as data acquisition methods. Multiple data acquisition methods indicate a match to idiosyncratic, interactive, multiple entry points and diversity unique characteristics of LSIE.

The broad goal of this research is studying scientific literacy and ‘testing the hypothesis that the guidance of parents is an important bridge between the intentions of the exhibit designer and the experience and knowledge of the child.’ (p. 12) The study seems well situated in the informal learning field particular to museums that specialize in science and discovery. The research is ‘inspired’ by socio-cultural and information-processing theories of how children learn and clearly the work is theory-driven. The study is a collaborative between university and museum interests which helps match the methods to the unique setting of LSIE.

The motivation for this work, although focusing on ‘science’ literacy, is not directly associated or governed by the formal science education processes and institutions. The researchers clearly explicate a well-reasoned informal settings approach including media, internet and traditional venues. The reasons for doing this are well situated in learning in informal settings.

The study focuses on the dynamic, idiosyncratic social interactions of visitors facilitated by the museum experience and the assessment approach is well matched to the experience. The authors include a section on translating the learning model into the research methodologies! What works? When the research takes into consideration the informal setting and derives the methodologies and methods from the view of learning in that setting. What does not work? Taking a well-worn traditional approach and research questions from a well-established domain like science education and shoe horning them into an informal setting.

The conclusions focus first on collaboration and the design of exhibits. They are generalized to other out-of-school settings and the authors comment on the nature of cognitive development through parent-child interaction. The researchers conclude by hoping that the empirical evidence can help guide new theorizing about cognitive development in general. This is a clear indication of theory building and research design based on the unique qualities of LSIE.

Edu, Inc. (2004). *It's a Nano World: A Study of Use. Findings from a Summative Report*. Ithaca, NY: Author.

The Nan biotechnology center, a National Science Foundation supported Science and Technology Center, collaborated with Ithaca, New York's Science center and Painted Universe, Inc. to create the 3,000 square-foot, interactive, traveling *It's a NanoWorld* exhibit. The goal of *It's a NanoWorld* is to introduce early elementary students (kindergarten through third grade) to a world they cannot see and to create a context for future learning about nanobiotechnology. This summative evaluation used pretest and posttest surveys in addition to exit interviews to assess children's ($n = 217$) knowledge, understanding and misconceptions of concepts related to nanobiotechnology concepts.

Results from exit interviews indicated that teens and adults felt that the exhibit was fun, inviting, and allowed them to engage in hands-on learning. Pre-test and post-test questionnaires showed that visitors evidenced overall increased understanding of key science vocabulary such as 'cell' and 'nano'. One unexpected outcome found was that 90% of adult visitors who viewed the exhibits could not give examples of nanotechnology or nanotechnology applications. However, when visitors were provided with illustrated children's books to introduce the NanoWorld exhibit, they showed improved understanding of cells, nano, and nanotechnology as compared to other visitors who did not receive any background information prior to the exhibit. This important outcome suggests that facilitated learning (i.e., cognitive mediation) is crucial for maximizing learning of abstract concepts characteristic of many informal science settings. However, no attempt was made to examine the cognitive and affective outcomes for diverse audience subgroups to determine if there was a differential impact ethnic and gender subgroups.

Results from tracking software found that visitors interacted most frequently and for the longest periods of time with the magnifying tools suggesting that multi-sensory exhibits that were novel (unpredictable) and stimulated visitors' curiosity motivated greater task engagement. Although, the frequency of exhibit visits and time spent at each exhibit were recorded, evaluators did not conduct in-depth interviews with visitors to determine the attributes of the most frequently visited exhibits that attracted and sustained their interest. Likewise, more comprehensive interviews and observations may have supplied more explicit details for assessing the characteristics perceived to be uninteresting for the less visited exhibits would also supply good information about future exhibit designed to create knowledge and understanding about abstract science concepts. An important learning outcome of this evaluation was that the ages where children are most likely to benefit from the use of models was identified by using 'Talk Aloud' interviews to diagnose misconceptions of younger children and solid conceptual understandings of older children.

Falk, J.H., Moussouri, T., & Coulson, D. (1998). The effect of visitors' agendas on museum learning, *Curator*, 41(2), 107-20.

Starting from the position that museum 'visitors not only have an 'agenda' for their visits but that these agendas directly influence visits' (p. 107), the authors attempt to investigate the impact of the agendas on visitor behavior and learning. Such studies, the authors state, 'are virtually nonexistent' (p. 107).

Visitor agendas are described as having two dimensions; *motivation* for visiting and the *strategies* people use when visiting. The authors base their work on six categories of motivation identified from previous studies: Place, Education, Life Cycle, Social Event, Entertainment, and, Practical Issues. Strategies are seen as being on a spectrum with three types, unfocused, moderately focused and focused.

The study took place at the Smithsonian Institution's National Museum of Natural History *Geology, Gems and Minerals* exhibition where visitor learning was assessed using Personal Meaning Mapping (PMM). In this case, the PMM consisted of a piece of paper on which 'gems + minerals' was written in a circle in the middle. PMM was used to 'generate four different equally valid measures of learning' (p. 109).

One in three groups of all-adult groups visiting ($n=3/4$) were approached. 87% of groups accepted the invitation to take part and (individually) completed a PMM before entering the exhibition. Visitors were asked to write down words or phrases which they thought were related to the prompt. Visitors were then engaged in a discussion to examine the reasons they chose the words/phrases and to find out more about their knowledge. Groups were tracked and approached as they emerged from the exhibition. Those who were willing to take part repeated the PMM. More questions were asked about their motivation for attending, their visit plans and their educational background. Visitors rated the six categories of motivation, using a five-point scale, in terms of how important they were on the day of the visit. Forty people completed the second PMM.

The authors comparing visitors' PMM vocabulary pre- and post-visit, and by examining the breadth of understanding. The depth of visitors' understanding was assessed on a scale of 1 (no elaboration) to 4 (significant elaboration). Finally, visitors' mastery was assessed holistically using a scale of 1 (novice-like) to 4 (expert-like). Other variables examined included, time spent in the exhibition, visit strategy, museum crowdedness, time of day, age, social group, and gender. Statistical analyses included analysis of variance, t-tests, Chi squares and correlations.

Most visitors showed 'significant gains in their ability to describe gems and minerals' (p. 114). Conceptual understanding was also improved in most visitors. The majority of visitors failed to demonstrate significant change. 'The research confirmed the hypothesis that an individual's motivation for visiting a museum significantly impacts how, what, and how much he/she learns at the museum' (p. 114). Not surprisingly, those with a strong educational motivation showed 'significantly greater learning than did those expressing a low educational motivation' (p. 115). However, the same was true of those 'voicing strong entertainment motivations' (p. 115). The study concludes that education and entertainment on a continuum, rather they are two separate continua. The authors point to 'the multi-dimensionality of the visitor agenda' (p. 117). They also suggest that the word education is associated with schools so that museum visitors who come to be entertained, do not expect or see themselves as being educated, too.

Falk, J.H. & Adleman, L.M. (2003). Investigating the impact of prior knowledge and interest on aquarium visitor learning, *Journal of Research in Science Teaching*, 40(2), 163-76.

Falk & Adleman investigated how the variability of visitor groups influenced learning at the National Aquarium In Baltimore (NAIB). The authors highlighted the difficulties of relying on constructivist-based learning theories to explain cognitive change because of the wide diversity of visitors attending free-choice educational programs such as science centers, zoos, aquaria, and natural history museum. The study was designed to determine if aquaria visitors with different levels of prior knowledge, experience, interest, motivations and expectations demonstrated different learning outcomes.

This quasi-experiment research was undertaken to examine the effects of visitor grouping on aquarium visitors' environmental knowledge, interest and concerns. A total of 395 visitors participated but only 100 of the original sample completed face-to-face semistructured interviews before and after the visit. Participants were divided into three knowledge groups (extensive knowledge, moderate knowledge and minimal knowledge), and within each knowledge group, by interest/concern (extensive interest, moderate interest and minimal interest). Paired *t*-tests were conducted on results from pre- and post-visit semi-structured interview data that was collected in the form of responses from open and closed questions and Personal Meaning Maps. In addition, results from observations of visitor by groups established for this study were reported.

The study found that there were significant increases across all 100 visitors in both conservation knowledge and interest and concerns. However, significant increases were not found for all visitor subgroups demonstrating that audience characteristics do have an effect on changes in knowledge and interest. These results have broad implications for assessing the learning that occurs in museum-like educational programs because the results obtained from one large group may hide the important changes experienced by subgroups. In this particular study, the investigators found that there were significant group changes for the entire group but not for all subgroups.

According to the authors, constructivist learning theory suggests that prior knowledge, experience and interest are a normative phenomena. This learning perspective does not address the diverse range of knowledge, experience and interest that characterize free-choice learners. Although, this study provides a sound methodological approach for assessing a range of free-choice learners, it is limited to one type of setting. However, the results have broad implications for other informal science settings such as natural history museums, science centers, and zoos. Some of the quantitative comparisons that were made used very small sample sizes leading one to question the validity of the outcomes. Nonparametric data analytic procedures would have been more appropriate in this case. Although, the data collection methods were appropriate for the study, qualitative results from observers were not reported. More detailed descriptions of visitors' behaviors may have shed light on why some subgroups experienced cognitive change while others did not. This study appears to be a reanalysis of data from a program evaluation study of the NAIB that constrains the methodological approaches that are used to build and verify theory (e.g. the contextual learning model) specific to informal science learning environments.

Gerber, B.L. (2001). Relationships among informal learning environments, teacher procedures and scientific reasoning ability, *International Journal of Science Education*, 23(5), 535-49.

Gerber's study aimed to acquire a better understanding of how informal learning experiences promote scientific reasoning ability and influence formal classroom learning. Specifically, the purpose of this quasi-experimental study was to investigate the differences in students' scientific reasoning abilities relative to their informal learning environments (impoverished, enriched), classroom teaching experiences (non-inquiry, inquiry) and the interaction of these variables.

The sample for this study consisted of 1,178 students enrolled in seventh-, eighth-, ninth-, and tenth-grade science classes. The students came from school in rural, suburban, and urban settings and ranged in size from approximately 50 students per grade to approximately 325 per grade level. The Informal Learning Opportunities Assay, a 41-item questionnaire comprised of closed and open-ended questions, was administered to all students and used to group them into either an impoverished or enriched informal learning experience category. Students were also grouped by science teaching method (inquiry vs. non-inquiry) based on interviews with their science teachers. Students across all groups were assessed with the Classroom Test of Scientific Reasoning Skills (CTSR).

The results of two-way ANOVAs indicated that informal learning environments and classroom science teaching procedures showed significant main effects on students' scientific reasoning abilities. Students with enriched informal learning environments had significantly higher scientific reasoning abilities compared to those with impoverished informal learning environments. Likewise, students taught using inquiry-based science methods evidenced higher scientific reasoning abilities compared to those in non-inquiry science classrooms. The findings from this study indicate that classroom activities enriched by informal learning beyond the classroom and inquiry-based teaching correspond to higher scientific reasoning abilities among students. The author concludes that, in general, the extent of informal science experiences and exposure to inquiry-based teaching would promote higher levels of science achievement in formal school settings.

The methods used map well with informal science assessment methods because the researcher attempted to demonstrate a direct connection between science learning that occurs outside the classroom and scientific reasoning abilities. Indirectly, these findings suggest that free-choice learning experiences may have a large influence on the science learning that occurs in school. Although this study did address the diversity of the audience (i.e., grade level), there is no mention of variability among classrooms or settings (e.g., curriculum, class size, prior knowledge) although there is a mention of the teaching approaches used. The data collection methods are quantitative and could have been greatly enhanced by observations and interviews to gain a deeper understanding of how informal science experiences contributed to students' reasoning abilities and ultimately to their science achievement. Additionally, achievement data compared across subgroups would have made a stronger case for the relationship between informal learning and formal learning outcomes. The author argues that both science and non-science oriented informal learning experiences are consistent with constructivist learning theory because they promote cognitive conflict and/or social interactions that are necessary for the development of scientific reasoning abilities among students.

Goodman Research Group, Inc. (2006). *Summative Evaluation of Einstein's Big Idea*. Cambridge, MA: Author.

In 2004, Boston Public TV (WGBH) received partial funding from the National Science Foundation to create *Einstein's Big Idea*, a two hour NOVA docudrama on Einstein and the history of the formula $E=mc^2$. In addition, WGBH created a website and Library Resource Kit to support the intended learning outcomes associated with the series. The Goodman Research Group (GRG) conducted a summary evaluation of the educational and program goals. The four educational themes evaluated were: (1) science is a human endeavor, (2) science thrives on the contributions of outsiders, (3) science impacts society, and (4) scientific advances often generate ethical dilemmas. Pre-/post-test surveys were conducted with regular NOVA viewers ($n=44$), high school students ($n=52$) and their teachers ($n=4$). 1,544 visitors completed the *Einstein's Big Idea* web survey. Most of the web visitors completing the survey were white males, aged 31-45. Post-test surveys were used to assess evaluation outcomes for adults ($n = 743$), young adults ($n = 385$) and child patrons ($n = 888$) who viewed the library resources.

In general, the survey found that the vast majority of respondents expressed favorable opinions of the docudrama, website and Library Resource Kit. In addition, most respondents indicated that they gained knowledge and understanding of Einstein's formula, the current relevance of the formula, the fact that many scientists contributed to the formula's derivation and that science is a human endeavor and a creative process that is dependent on the contributions of diverse groups of people. GRG also assessed the role of women and minorities in science to create an awareness of the diverse groups that contribute to scientific discoveries. In addition, this evaluation study assessed prior knowledge, cumulative knowledge and attempted to predict future learning endeavors through the survey and interview questions. Motivation to learn new information and concepts was assessed by observing the total time visitors interacted with the *Einstein's Big Idea* website. Numbers of visitors accessing individual web pages were also recorded to assess the topics of most interest.

Although, the evaluators were able to collect the data needed to evaluate the program goals, there were missed opportunities for increasing understanding of the factors that contribute to learning theory. For example, although diversity issues were addressed, results from survey respondents were not disaggregated by gender or ethnicity to determine if these groups were impacted differently by the *Einstein's Big Idea* program. Independent observations could also have provided more detail about factors related to curiosity and interest that motivated some participants to visit the website and its accompanying pages longer than others. Furthermore, website visitors were not asked to make comments about their learning experiences or engage in interactive dialogues with other visitors or web site facilitators missing an important opportunity to collect data on the type and depth of social interactions that influence science learning in informal settings. Analyzing discourse from visitor conversations during the TV series and when visiting the website could provide a deeper understanding of the complex interaction of factors that influence learning in informal contexts.

Gray, B. (2004) Informal learning in an online community of practice, *Journal of Distance Education*, 19(1), 20-35.

This research study addresses the question, 'What role can online communities play in meeting the informal learning needs of a professional association?' The motivation for the study is based on the fact that organizations and professional associations are increasingly examining the potential of online networks to enable members to share knowledge and engage in on-going workplace learning and professional development. The stated purpose of the study is to better understand to what extent participant's experiences in an online environment constituted a 'Community of practice' and the nature of the learning that occurred.

The study focused on 43 coordinators of the Alberta (Canada) Community Adult Learning councils who voluntarily participated in a WebCT based online community for one year. Contributing to the need for professional development of the participants was the fact that they were geographically isolated and their positions typically had a high turnover rate. The author was immersed in the study as both moderator for the online environment and as the sole researcher.

This is a well-grounded empirical study based on the theoretical construct of communities of practice (COP). COP is an anthropological perspective that examines how adults learn through everyday social practices rather than focusing on the context or environment that are designed to support learning. Recent work in COP has focused on the use of innovative technologies and electronic platforms to support communities in the context of the increasing geographic distribution of employees and the global nature of work. This is an interpretive study based on multiple data acquisition methods including: review of online discussion postings, live chat transcripts, email correspondence of all participants, a participant survey including seven open ended questions and individual on-site interviews with 11 participants.

Results of this study suggest that the online environment functioned as a community of practice where online participation served as a tool for learning situated in the context of everyday experience. Participation became important in defining the identity of the practice itself. In particular in relation to identity and meaning, a key characteristic of learning communities is that they provide members with a medium for creating identity and understanding their work. Through their participation in the online community, participants explored identity questions especially as they related to their positions as coordinators. This aspect was especially apparent in people who were new to their job and the online environment. The author also claims that the development of collective identity was particularly important for this occupational group.

This study brings together both well-defined research methodology and data acquisition methods focusing on the contextual distance learning in an informal setting. Outcomes regarding the development of identity in informal settings are well documented.

Hodkinson, P. (2007). *Learning as Being, Lifelong Learning Institute, Leeds; University of Leeds.*

'Learning Lives: learning, identity, and agency in the life-course' is a 3.5 year project which started in 2004. It involves small teams of researchers at four UK universities. The aim is to deepen understanding of the meaning and significance of learning in the lives of adults, particularly in relation to identity and agency. To do this they are examining a range of learning experiences from the perspective of the adult learners themselves, set within the context of their unfolding lives. There are two strands to the data collection, involving the integration of three different methodologies. The first strand is a qualitative study of more than 100 people, drawn from different walks of life, different parts of the country, and of different ages, genders and ethnicities. The second strand is quantitative, combining two data acquisition methods based on interviews: life-history research and longitudinal research, in which interview of subjects occur several times over three years.

Each individual subject has been interviewed first about their life history, being asked to tell their own story with as little prompting as possible, but in the knowledge that the project was about learning. In later interviews, the stories have been further explored, with more intervention from the interviewer where necessary, to build on the earlier interview(s). As well as elaborating their life-history, successive interviews elicit information about ongoing events in people's lives. Thus the project explores the temporal context of learning retrospectively and as 'work-in-progress'. The life-history project encourages people to look back over their lives, whereas the longitudinal research allows them to provide a series of 'snapshots' tracking the ways their learning is lived over the research period.

Life-history research is dependant on the ways in which people reconstruct the past through the narration of their life stories. The past, in turn, impacts on the present. The past affects the way the present is understood, and the way it is presented. A person's current learning relates to past learning, building on previous experience or reworking previous understandings. The longitudinal research is important both because of its interaction with the life histories and because it provides the possibility of reviewing learning almost as it happens. Although there are a large number of reports from this research group, two were selected for review in this paper.

Jackson, A. & Leahy, H.R. (2005). 'Seeing it for real...?' Authenticity, theatre and learning in museums, *Research in Drama Education*, 10(3), 303-25.

This study assessed the effectiveness of theater within the museum experience and compares and contrasts the different learning outcomes of theater and non-theater sessions over a period of two months. It employed an experimental research methodology with interaction analysis, observation and work samples as data acquisition methods. The two-month research period reflects the long-term learning in informal settings. The multiple methods within the methodology match well to characteristics of informal settings such as complexity, diversity and variability of the audience as well as multiple entry points for participants. The sample is small but with a greater responsiveness to the diverse views and reactions of the participants than is usually possible in large-scale studies.

The researchers' motivation is related to the desire to open a field of inquiry beyond the framework of 'evaluation which is primarily instrumental concerned with measuring effects against prescribed learning outcomes' (pp. 305-6). The researchers note that 'there is comparatively little open-ended, disinterested research into the effects of educational policies and programs' (p. 305). The study was motivated by the desire to see if, and how, theater in the museum setting contributed to, or detracted from, children's cognitive and affective learning during their visit. This approach is congruent with the need to investigate the experiences, outcomes (both expected and unexpected) in the light of an informing theory of learning. Although experimental, the purpose of the research is not to determine which strategy is better or more effective, but rather to compare the strengths, weaknesses and outcomes of each. Although not committed to a particular theory of learning, studies such as this contribute to the very nature and purpose of learning in museums: namely, what theories of knowledge and learning underpin contemporary practice (p. 305).

This research is conceptualized within unique aspects of theatrical experiential learning in informal settings. The authors question the predominant paradigms of museum learning assessment and traditional approaches to evaluating prescribed outcomes. The authors discuss issues of learning in informal settings based on connections made through historical fiction and theater rather than traditional archaeological artifacts.

Knight-Williams Research Communications (2005). *Summative Evaluation of National Geographic's Strange Days on Planet Earth Television Series and Website with an Adult Audience*. Sacramento, CA: Author.

Strange Days on Planet Earth was a four part television series and outreach program produced by Sea Studios Foundation (SSF) for National Geographic Television and Film with funding from the National Science Foundation. The project was comprised of three primary components: a broadcast series, website, and a national consortium of informal learning institutions. 98 participants (18-55) were randomly assigned to either the series-only condition ($n = 48$) where they only watched the four television programs or the series plus website ($n=48$) condition where individuals also visited the *Strange Days on Planet Earth* website. Both groups were found to be equivalent with respect to gender, race, level of education, interest in and knowledge of the environment, or frequency of viewing science/nature shows and the PBS channel. Each group completed pre-/post-test questionnaires to assess the appeal, entertainment value and clarity of the television series and website. Participants in each group also completed pre-/post-test knowledge tests to assess viewer learning. Frequencies and percentages of responses from fixed choice survey items and content analysis of open-ended survey and items indicated that the TV series appealed to viewers, and had a substantial impact on their knowledge, attitudes, and beliefs about a wide range of environmental issues related to invasive species, global warming, predators and water pollution.

No significant differences on pre-/post-test surveys or knowledge tests were found for comparisons made between the series-only and the series plus website groups. With respect to subgroup differences (i.e., gender, education, occupation, frequency of viewing science programs, interest and knowledge of environment) across both comparison groups, one difference that emerged was that more frequent visitors of these programs rated the series higher on the elements of storytelling, level of visual excitement, likelihood of recommending and clarity. The only other subgroup difference was found for viewers who felt they were less knowledgeable about the environment. They rated their learning from the series significantly higher than those who felt more knowledgeable. In general the series was highly regarded by both men and women, individuals of varying ages, educational backgrounds, television viewing habits and knowledge of the environment.

The strength of this evaluation study lies in its rigorous research design involving random assignment to control for extraneous sources of error. In addition, the diversity of the viewing audience was assessed and comparisons were made to determine if the *Strange Days On Planet Earth* TV series and website had a differential impact on subgroup science learning and attitudes. In addition, the researchers assessed specific learning outcomes and attempted to determine the influence of prior knowledge on the type and depth of learning that occurred in this free-choice learning context.

Evaluators also included pretest and posttest survey questions that assessed viewer interest, engagement and inspiration to take action promoted by the *Strange Days* series and Website. Evaluators also assessed the retention of environmental concepts by conducting telephone interviews on a small sample of participants two weeks after viewing the TV series.

Lebeau, R.B., Gyamfi, P., Wizevich, K., & Koster, E.H. (2001). Supporting and documenting choice in free-choice science learning environments. In J.H. Falk (Ed.), *Free-choice Science Education: How we Learn Outside of School*. New York, NY: Teachers College Press, pp. 133-48.

This study examined the effects of goal-setting activities on the plans, interests, expectations, efforts and learning outcomes of public school students during informal learning activities. The study was based on research that suggested that providing information about the setting prior to a visit reduces novelty and enhances learning (Falk & Dierking, 1992). A randomized, pre-test-post-test approach was conducted to study the effects a novelty-reducing activity coupled with a goal-setting activity on the self-regulated learning of students. 27 students, aged 11-14, enrolled in a summer remedial program in a New York City school, were randomly assigned to one of three experimental groups. Students assigned to the first treatment group were given only a map to help them navigate the science center. Students in the second treatment group were provided with a map plus a worksheet activity on which they were asked to indicate three areas they would most like to see, two areas they were not interested in seeing and one area in which they would like to spend the most time. Students in the third group were also given the map and the worksheet and they were also required to write a question to which they hoped to find an answer as a result of their visit. Students in all three groups completed pre- and post-test survey questions related to help-seeking behaviors that were hypothesized to be associated with informed goal-setting and attitudes toward science learning.

The three treatment groups did not differ significantly on the survey questions written to assess informed goal setting. However, students assigned to the 'map plus worksheet activity' group evidenced higher mean scores on items related to goal-setting than those assigned to the 'map only' group. The number of students disagreeing with the statement 'Science is mostly about memorization' was found to increase significantly from pre-test to post-test measures. The authors suggest that a pre-visit goal setting activity increased students' willingness to ask for help necessary in order to focus on learning resources that were available to them, rather than wandering aimlessly with little attention to improving their understanding of the context.

This study is a good example of investigating the kind of variables that influence the learning process of visitors to ISIs. The research addressed the physical complexity of ISIs that can certainly be overwhelming to visitors of all ages. What this study does not examine is the influence of goal-setting activities on the diverse and complex characteristics of audiences. Although, the authors assert that goal-setting is essential to learning, participants were not assessed by knowledge tests related to the worksheet activities that they completed. The authors' original intent was to randomly assign 41 students to one of three experimental groups. However, 14 of the original students that were expected to participate were absent for either the pre- or the post-visit assessments. The unintended outcome points to the difficulties faced by informal science researchers who feel compelled to engage in experimental designs that require random assignment to satisfy the evaluation requirements of governmental and private bodies.

Leinhardt, G., Tittle, C., & Knutson, K. (2000). Talking to oneself: Diaries of museum visits. Retrieved December 29, 2006 from <http://mlc.lrdc.pitt.edu/mlc>.

The authors investigated the use of diaries as a tool for assessing the nature of learning that occurs in museum settings. More specifically their intent was to study visitors' diaries to gain a better understanding of general experiences and to identify the cognitive tools they used to construct meaning from museum interactions. This research effort is premised on sociocultural learning theory which suggests that meaningful knowledge is co-constructed through conversations with others. The authors assert that conversations prior to, and after, museum visits provide visitors with opportunities to co-construct knowledge with other visitors about features and issues related to the specific free-choice context of interest. They extend sociocultural learning theory by proposing that internal visitor conversations, written in diaries, can provide the same degree of knowledge construction that is normally experienced with fellow visitors.

The diary entries of 15 adults visiting museums of their choice were analyzed. Each 3-5 page journal entry was written as soon as possible after the visit. Participants were also asked to include pictures, cards, flyers or catalogues to support the experiences that were documented. Each diary was read five times by three different analysts and coded according to the purpose of the visit, environmental impact and cognitive processes used.

Results from narrative analysis of rich, thick diary entries found that the reflective passages, written by participants, supported the notion that deep levels of meaning derived from museum visitors' experiences are dependent on: purpose, the museum environment and level of cognitive processing. The study aligns well with informal learning research methods because it examines the complexity and variability in physical, personal and social contexts of free-choice learning environments across time and space. Each of the participants' diaries resulted from several experiences, at different times, with museums of their choice. Although the methods employed focused on learning processes, they do not explain much in the way of participants' prior knowledge to help us better understand their relative and cumulative cognitive gains. Individual, semi-structured pre and post-visit surveys may have supplied data that would have further supported the cognitive gains experienced based on the meaning of experiences diarists reported.

Lynch, H., (2007) *Learning as desire*. Paper presented at ESREA (European Society for Research on Education of Adults) Network Conference on Life History and Biography, March 1-4, Roskilde University, Denmark

In this paper the author draws from more than ten life stories, seeking to describe the concept of desire and how that influences individual adult learning. The author draws on the theoretical construct of Boltanski and Thevenot's sociology of critical capacity (1999). This approach leads to particularly relevant insight into the decision-making process revealing operationalized desires. The author builds her interpretation of the life stories seeking to map the differing contexts in which decisions are made.

Condition	Characteristics
Domestic	This realm encompasses values of family, community and tradition
Civic	Public benefit, common will, the good of all and equality
Inspired	personal growth, creativity and spontaneity
Market	material wealth, competitiveness and short term gain
Industry	productivity, efficiency, functionality
Opinion	Recognition by others, fame, celebrity status

The results of the multiple life stories reveal that learning by adults is based on a desire that is composed of a 'geography of values'. Elements of the 'learning desire' landscape identified in this paper are constancy, shifting desire, conflicting desire, ability to control life choices, embodiment, proximity and significant others, mobility, momentum, agenic capacity, imagination and space.

The strength of this article lies in its basis of long-term in-depth descriptive data collection. Identity outcomes related to adult learning are well documented. New categories of identity outcomes related to specific activities and experiences are proposed. The study is firmly based in theory building and emerging research methodology.

Meisner, R., vom Lehn, D., Heath, C., Burch, A., Gammon, B., & Reisman, M. (2007, forthcoming). Exhibiting performance: co-participation in science centres and museums, *International Journal of Science Education*.

Given the lack of knowledge about the type and quality of interactions between visitors and computer-based exhibits, the authors use novel computer-based exhibits ‘to explore how people, both alone and with others, interact with and around the installations’ in the context of the Energy Gallery at the London Science Museum. Unusually, the authors of the study include researchers and museum practitioners.

Focusing on the problems of educating visitors about complex scientific phenomena, the researchers ask, ‘how do you make the seemingly tedious engaging?’ More specifically, the authors ask: ‘What are the appropriate media to convey such content? And, what are the desired behaviours to encourage engagement with the exhibits?’ Video-based field studies captured visitors’ performative activity rather than relying solely on visitor talk. Visitors’ verbal and bodily conduct are used to examine the social organization of visitors’ ‘performances’.

Methodologically, the authors identify ethnomethodology and conversation analysis which are used ‘in conjunction with Goffman’s [...] insights on face-to-face interaction.’ Video cameras were wall-mounted or set on un-personned tripods so as to minimize possible feelings of being observed. The authors ‘subject[ed] the [20 hours of video recordings to detailed scrutiny to uncover how action and interaction emerge from and are part of the context in which they are occurring. Here, ‘context’ refers not only to the physical environment but to the unfolding nature, or moment-by-moment production, of the activity that arises’. ‘The analysis involves the detailed transcription of short fragments of video – single instances of discrete phenomena, here, visitors’ performative activities – including participants’ talk and bodily comportment.’

The study finds that visitors may ‘exploit certain design features – such as multiple interfaces, large screens and the various spaces around exhibit components – to configure their actions in elaborate and embellished, and thus more noticeable, ways that can attract and hold other people’s attention...’ The authors’ position is that further studies of visitor ‘performances’ ‘may bring to light both how an exhibit works and what it is about, and as such, might become ‘talking-points’, occasioning the verbal forms of social interaction which are increasingly considered to be critical resources in shaping people’s experience of and learning from exhibits.’

The authors ‘reflect upon the extent to which the design of exhibits enables particular forms of co-participation or shared experiences, and to develop design sensitivities that exhibition managers and designers may consider when wishing to engender novel ways of engagement and participation with and around computer-based exhibits.’ Additionally, the authors argue that ‘by focusing on performances as a particular kind of communication, this paper wishes to extend the current focus on conversation to include looking more broadly at the creation of shared experience. The study ends with a pertinent question for researchers and museum practitioners. ‘Is there a conflict between always wanting visitors to be both deeply engaged at an exhibit and actively conversing with one another? Or, might it be that shared experiences, including but not limited to conversations, allow visitors to create a memorable experience upon which to build both in the moment at hand and in the future?’

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 Outside of School*. New York, NY: Teachers College Press, pp. 93-114.

Miller conducted this correlational study to determine if there was a relationship between formal science and informal science learning environments (Why). He specifically sought to answer the questions, 'Is there a relationship between formal and informal science learning when examining free-choice survey items taken from the 1997 Science and Engineering Indicators study (What).

Miller first compared results of respondents biomedical and space knowledge to their informal science knowledge. Results from this analysis found that education and the number of college level science courses are positively related to the understanding of biomedical and space science construct. The second part of the study was undertaken to conduct an exploratory factor analysis of the seven free-choice learning items (i.e. number of: public library visits, science museum visits, television shows watched, hours of home computer use, news magazines read, science magazines read and newspapers read) from the 1997 Science and Engineering Indicators study. Results from the analysis found that all seven items were highly correlated indicating that they assessed one dimension of informal science learning.

The items identified in the factor analysis were used to develop a path model based on the responses of 2,000 adults of proportional gender representation and from across varying levels of education, college courses taken and science related occupations (Who). This correlational analysis was hypothesized to explain the influences of age, gender, education level, number of dependent children, occupational interest, issue attentiveness and the use of informal science-education resources on respondents' understanding and knowledge of science, technology, space and environmental issues. The total effects of level of educational attainment and the number of college courses taken had a significant influence on the use of free-choice science education resources and better-educated Americans with some college are significantly more likely than other Americans to use the resources. Results from a structural equation modeling analysis found that that informal science-education resources had small yet significant effects on students understanding of both biomedical and space knowledge.

The author concludes that there is an important and reciprocal relationship between formal and informal science learning experiences of the adults who responded to the NSF 1997 Science and Engineering Indicators study. Miller does not attempt to explain the relationship between the variables investigated and a learning theory that supports the importance of their study.

This study provides a good example of a method for examining the influence of prior and coexisting knowledge on informal science learning across a varied free-choice audience. However, the study is limited because it does not address the complexity and variability of specific free-choice experiences or the socio-cultural influences of free-choice learning. In addition, the assessment of understanding is constrained to one period of time therefore failing to address the issue of cumulative knowledge. However, the number of college sciences course taken might be considered as an indicator of cumulative informal science knowledge since informal science learning was found to be highly associated with formal science learning experiences.

**Randi Korn & Associates, Inc. (2002). *MarsQuest Summative Evaluation*.
Alexandria, VA: Author.**

This report presents findings from the summative evaluation of the Space Science Institutes of Boulder, Colorado's National Science Foundation funded *MarsQuest* traveling exhibit. Data collection took place at the Park Place Mall in Tucson, Arizona and the Hampton Air and Space Museum. Randi Korn & Associates (RK&A) used timing and tracking data collection strategies to determine the length of time that visitors spent at the *MarsQuest* exhibit as well as the time spent viewing each of the individual components. The individual *MarsQuest* displays consisted of touch exhibits, mechanical interactives, computer interactives, models, videos, and panels. In addition, RK&A conducted post-visit interviews at both the Tucson ($n = 30$) and Hampton ($n = 20$) sites..

52% of the respondents were male and 48% were female. The majority (56%) of visitors across both sites were aged between 22 and 44. Visitors at the Tucson venue ($n = 26$) spent on average 55 minutes in the exhibit as compared to an average of 28 minutes for Hampton *MarsQuest* visitors ($n = 99$). RK&A also used Serrel's 'Sweep Rate Index' (SRI) to further analyze the time visitors spent in the exhibit at both sites. The SRI is calculated by dividing the exhibition's square footage by the average total time spent in the exhibition. The lower the SRI, the more time visitors spent per square foot of space. The SRI for the Tucson visit was 85 as compared to the Hampton SRI of 163 indicating that visitors moved more slowly through *MarsQuest* in the Tucson venue than visitors attending the Hampton venue. The SRI results indicate that Tucson visitors were more interested and engaged in the *MarsQuest* exhibit than those Hampton *MarsQuest* visitors.

Another unique data analysis method employed for this summative evaluation was the use of Serrell's 'Percentage Diligent Visitor Index' (%DV). The %DV is obtained by calculating the percentage of visitors who stopped at more than one-half of the exhibits. The higher the %DV the more thoroughly the exhibition was used. The %DV across all *MarsQuest* displays for the Tucson site was 42 percent as compared to 8 percent for the Hampton site. These results suggest that the Tucson visitors spent more time at each individual *MarsQuest* exhibit than visitors at the *MarsQuest* Hampton exhibit. The most frequently visited *MarsQuest* component was the Interactive Programmable Rover. Visitors were found to spend the most time at Imagination Theater.

This evaluation also assessed the number of adult-child interactions occurring at both the Tucson and Hampton *MarsQuest* exhibits. 100% of the visitors at Tucson exhibit ($n = 22$) engaged in adult-child interactions during their visit while child-adult interactions were observed for only 68% of visitors at the Hampton ($n = 64$) site. The interactive and educational qualities of the *MarsQuest* exhibit were beneficial for adults and children. Interview results further found that most well-liked exhibits were the programmable rover and Imagination Theatre which provide added support for the interest that encouraged the amount of sustained engagement time recorded for those exhibits.

Little formal data was collected to compare knowledge and understanding before and after attending the exhibits. The small number of interviews and general questions about learning do not provide valid information about the cognitive gains that were made. Nor are we informed about the depth of reflective inquiry encouraged by the *MarsQuest* exhibit. Although demographic data was reported by gender, the authors made no attempt to contrast male and female perceptions.

Siegel, D., Esterly, J., Callanan, M.A., & Wright, R. (2007, forthcoming). Conversations about science across contexts in Mexican-descent families, *International Journal of Science Education*.

Siegel *et al.*'s study is part of a larger research project that investigates parental use of causal explanations (Tenenbaum, Callanan, Alba-Speyer, & Sandoval, 2002). The study asks 'whether children from different family backgrounds have different experiences talking about science in informal settings'? The study is situated in the literature on parent-child 'everyday' conversations which, they argue have been suggested as a source of children's early science learning. The theoretical view of learning is a sociocultural one which enables the authors to examine how 'parents' language use provides children with potentially useful information about how they mentally carve up domains'. One of the unusual aspects of the study is that the authors link 'the theme of exploring family science talk [...] in cognitive developmental research [with] science education research, and [...] sociocultural research' by focusing on interactional style. Specifically, the authors focus on the relation between 'parents' schooling and both their explanatory talk in science-related activities, and the styles of interaction they use with their children.' The study aimed to develop a theoretical understanding of the links between parental schooling and family talk but does not make direct claims about specific aspects of the theories and how they inform each other.

The participants in the study were 40 families with different schooling backgrounds from a group underrepresented in science education, that is, Mexican-descent families. In Study 1, the authors focused on family interactions while discussing buoyancy. Using videotaped observations of the families, the researchers coded both the content of what families discuss as well as the style of interaction. The analysis involved variance and correlational techniques and looked at 'how the structure of a task influences family interactions, by looking at two different phases of the buoyancy task (testing and prediction), and comparing interactions in that task with conversations in a more open-ended visit to a set of science-related exhibits at a children's museum.

Regardless of their schooling, parents had similarly rich explanatory conversations with their children in the buoyancy task. Across the two activities, 'but especially in the museum setting, parents with higher schooling tended to be more directive with their children. Thus, parents' experience with formal schooling in our studies does seem to relate to their later interactions'. This finding supports the work of Rogoff and colleagues (Rogoff et al., 1993; Rogoff & Toma, 1997). The authors comment: 'In a more open-ended setting, but where the focus was on science, parents with higher schooling continued to behave as they did in the structured task, perhaps using school-like strategies for discussing the exhibits with their children.' The authors note that the 'findings of our studies add some support to the idea that open-ended activities such as museum visits may encourage more collaborative discussions about science between adults and children.' The researchers' claims emphasize museum visits and experiences and address the unique qualities of LSIE. One of the unique qualities of LSIE that could easily have been addressed in this research would be recognizing the importance of related informal learning experiences through the association of common everyday events that inform children and parents knowledge of buoyancy such as boats, ships and swimming.

Storksdieck, M., Ellenbogen, K., & Heimlich, J.E. (2005). Changing minds? Reassessing outcomes in free-choice education, *Environmental Education Research*, 11(3), 353-69.

The case study research conducted by Storksdieck, Ellenbogen & Heimlich was undertaken to describe the learning impacts of three, free-choice, informal learning programs and to establish guidelines for designing learning experiences that accommodate the varied worldviews and attitudes of learners (Why). The overarching research question posed was ‘Which groups of visitors gained knowledge and in what ways?’ (What).

The first case study focused on self-selected, environmentally-aware audiences visiting a traveling exhibit on biodiversity that was supported by the National Science Foundation. Personal Meaning Maps and pre- and post-exhibition interviews, using a card sorting task, were conducted with 62 visitors selected from a random sample of the general public attending the exhibit. Results suggested that visitors’ understanding of biodiversity issues improved significantly. The second case investigated the impact of hotel water conservation (guests selected whether or not to have linens and towels laundered daily). Analysis of 200 response cards of guests from 16 hotels were analyzed and, in general, most participants indicated positive perceptions of hotel water conservation practices. However, most guests also indicated that the hotel conservation practices did not make a difference environmentally. A nature center program partnered with 16 Baltimore city schools was the third case examined. Matching pre-visit and post-visit surveys were collected from 149 middle and high school students. Although changes in knowledge were small, the majority of secondary students expressed positive perceptions of the nature program and of working with elementary children (Who and How).

This multi-case study suggests that discrete knowledge outcomes are too narrow to assess the impact of free-choice environments on informal learning. Storksdieck *et al.* argue that behaviorist models of learning which focus on assessing changed behavior by measuring discrete knowledge outcomes are not suitable method for evaluating the complex outcomes associated with free-choice learning (Theory). They assert that informal learning outcomes must be expanded to include a broader array of indicators focused on incidental learning; unintended improvement in skills; reinforcement of previously held knowledge; acquisition of attitudes and beliefs, program attraction and cumulative learning.

The value of this study is that it addresses the complexity, diversity and variability of informal science learning environments across time and space by investigating three cases with three very different types of visitors. The authors report quantitative results in the form of both descriptive and inferential statistics to demonstrate learning gains and positive changes in interest and attitudes. However, very little is reported in the way of in-depth descriptions from open-ended questions to inform the study about specific experiences that were responsible for the reported changes in learning, attitudes and interest. Rich, thick descriptions of visitor experiences would have provided a deeper understanding of the complex interaction of variables that had positive influences on learning for each free-choice context that were studied.

Tenenbaum, H.R., Rappolt-Schlichtmann, G., & Zanger, V.V. (2004). Children's learning about water in a museum and in the classroom, *Early Childhood Research Quarterly*, 19(1), 40-58.

This study looked at how children from low-income homes learned about water as a result of combined museum and classroom instruction. The study employed an experimental research methodology using pre-/post-test interviews with children and a teacher questionnaire as the data acquisition methods. The researchers interviewed control and experimental groups before and after they visited museum exhibits. The research methodology was guided by Fischer's skill theory which is congruent with the informal focus on learning in multiple contexts and assessing learning processes and unexpected cognitive outcomes.

The researchers claimed influence over this study from government agencies' interest in low-income, under-represented groups in science, science learning and achievement and museum learning especially as it combines informal and formal settings. These perspectives are combined in the authors' new conceptualization of a framework for learning and development in context based on socio-cultural and constructivist perspectives. The research questions focused on children's knowledge acquisition, positive affect, complex understanding and correct judgment focusing on science learning rather than informal settings.

Interviews incorporated questions and hands-on activities with water. They were conducted in a school hallway before any instruction. Although this approach matched the event-centered nature and interactional social aspects of informal settings, it was decontextualized and does not connect well with the unique experiences of informal settings. The experimental group had both classroom and museum instruction on water, one water lesson before the museum and one lesson after the museum. It is not exactly clear what instruction the control group had about water. However the researchers report that most teachers taught two classroom lessons on water.

The researchers discuss the outcomes of the study in relation to their hypotheses about children learning science. The children who visited the museum exhibits and participated in the before and after classroom activities showed more content knowledge and more complex concepts. All students showed affective development. The researchers claim that they cannot determine if learning took place in the classroom, museum or both and that the interviews did not truly assess changes in children's reasoning abilities or deep conceptual knowledge, in fact they focused on knowledge and, to some extent, comprehension of science concepts.

Based, as it was, on a sound experimental research design and data acquisition method, this study corresponds to several of the unique characteristics of science learning in informal contexts. It allowed for the assessment of unexpected outcomes, assessing with hands-on experiences, event-centered learning and assessment that was sensitive to emergent ideas and focused on affective development.

Part 4 Promising Future Directions

Committee's Focus Question: What do research efforts thus far indicate for promising future directions?

4.1 Ideology (What?), epistemology (How?) and axiology (Why?)

It is evident from the review of the literature that there exists a multitude of data collection methods available for use within a range of research methodologies. There is, however, a need for researchers to be clearer about issues such as purpose, methodology and methods. Simply classifying approaches as qualitative or quantitative divide is simplistic and invokes a false dichotomy. Surveys, for example, can be comprised of open-ended or fixed choice questions so the data can be analyzed in ways which are 'quantitative' and 'qualitative'. Even the text from the open-ended questions can be coded to yield categories that can be analyzed quantitatively.

In relation to future directions, we believe that it is important for the committee to consider the concept of consequential validity. Some researchers suggest that this to be the most important type of validity to consider and that all other forms of validity contribute to it (Messick, 1989). The point here is that the method should be driven by the 'purpose' of the research and the audience that it is directed towards. However a single method study is no more or less appropriate than its fit to the circumstances and purpose of the research/evaluation study. Certainly, some assessment and evaluation methodologies and methods can, and do, match up to the unique qualities of learning science in informal environments, however, this match is inconsequential in light of the issue of validity. The validity of the research conducted (and methods used, including assessments and other data collection procedures) is dependent on the purpose of the research and how the research relates the appropriateness of its use to the important research questions that are under study.

In order to move assessment of forward, researchers (including evaluators) must explain why they chose a certain approach/or approaches. Consider the different research methods you might select for a grant proposal. You will propose specific research methods to evaluate the outcomes based on the funding priorities. These could be very different from methods selected to generate greater understanding of learning processes or build new theory. In reviewing the research, we often found that the authors did not provide enough information to judge the appropriateness of the research methods used. There was often an insufficient description of the research design, a failure to adequately describe the participants, a lack of technical and development information related to instruments, and poor detail in describing data acquisition methods.

One observation that we have made from our reviews is that research studies often fail to clearly articulate their purpose using a common research language. We believe this is true for much of the research in many disciplines. For example, a purpose might be to *describe* visitors' experiences and learning as a result of their engagement with an exhibit. The researcher's purpose for describing the experience would indicate that a case study approach using interviews and observations might be an aligned strategy for conducting this type of research. On the other hand the

purpose of the research might be to *describe* visitors' experiences and determine if there is a *relationship* between visitor interest and time spent engaged at an exhibit. In this case a case study approach in collaboration with a correlation approach exploring the relationship between time of engagement (collected from observations or self-reports) and assessed interest for a larger group of visitors would be appropriate

In many cases we are left to deduce the intended purpose of the research. In much of the literature we found the research purpose (Why?) seem to be so broad that any research methodology and data collection method could be used to answer the research questions (the What?).

In the future, we suggest that researchers interested in informal science learning clearly communicate their purpose, using descriptors that are linked to, but not limited to, certain methodologies. When the purposes are clearly established, then the data acquisition and analysis should be aligned with one another. In the case of the research we have been reading, we often had to work backwards from the conclusions to determine what was the purpose of the study.

4.2 What has been done that has worked?

In the future, we suggest that researchers take into consideration that the purpose of research studies is in part based on the researchers' relationship to theory. Most of the work we reviewed was theory-driven. Few studies took the results and then returned to the theory and explained how the new knowledge extended the original theory to new settings (theory testing) or how the knowledge helped reorganize the theory in more meaningful ways (theory building). We believe addressing the role of theory is important because the theoretical underpinnings help determine the strategies and outcomes of the studies.

From our review of the literature, it is clear that the underlying theories of science learning in informal environments are emergent and not fully defined. Some studies that we reviewed directly addressed the theoretical underpinnings and conceptual frameworks upon which the study was based. In these cases, the authors made claims about theory in relation to the data and results. Most made reference to theory in the belief that it would inform practice in informal settings. Some authors described relevant theory and then did not address how the results informed the proposed theories. If we consider the Ash study (2003), the author addressed the theoretical underpinnings early in the paper, and began the conclusion section with a discussion of how the outcomes reflected the specific theory and then went on to explain how the work informed the theory. This approach is responsive to the unique characteristic of learning in informal environments in that it contributed to the emergent theoretical base.

On the other hand, Anderson *et al.* (2000), in the introduction to their paper, referred, in a very limited way, to the theoretical framework of the research when they stated that 'students construct knowledge'. The research objectives: to describe experiences; to report understanding of construction; and to reflect on implications for teachers, are bereft of a coherent theoretical framework. In the 'significance' section of the paper, the authors reviewed studies into the effect of exhibits on visitors and, again, briefly mentioned that knowledge was constructed. Then, the authors

moved quickly on to methodology, procedure, analysis and findings. In the discussion, the authors devoted several paragraphs to the fact that children in the sample constructed and reconstructed personal knowledge. The descriptor, 'Theory driven' applies to this study.

We believe that paying attention to the underlying theoretical construct of a research (assessment or evaluation) project is important because it informs the study in terms of selecting methodologies, and methods of assessing learning as well as interpreting the results. Most of the papers that we reviewed attempted to extend theory from diverse fields to informal learning environments. Most of the articles examined factors and influences that are related to broader learning theories or their features (sociocultural issues, social interactions, constructivism, and the contextual learning model, etc.) relative to learning science in informal contexts. However, the authors did not explain their intent to examine the use of these theories to explain learning in informal contexts,

A major confounding factor in considering the relationship to theory as an important criterion, is that much of the work is tied to public and private funding. If we consider the journal publications of Falk and Dierking's work in informal settings in relation to the contextual model of learning, we often see a disconnect between theory, methods and claims. This disconnect may be attributable to the need to meet funding expectations rather than building the contextual learning model into a theory of learning in informal settings. Although they have proposed the contextual model of learning, often their research articles might be read as reflecting the more mainstream experimental empirical positivistic approaches favored by public and private funding agencies. In those studies, the conclusions often do not relate back to how the theory informs the outcomes or how the outcomes inform the theory.

We propose that one of the unique characteristics of science learning in informal environments is that it is an emerging field and thus theory driven, theory testing and theory building are critical to the development of the field. Research that is responsive to the theoretical underpinnings of the study may take this into consideration and use methods and derive claims that are responsive to the informal context. The alternative, taking existing theories and trying to make them work in an informal setting, may result in claims of little value in that context.

In the special case of identity, we know that affective identity outcomes are the result of many situations and happen frequently thus contributing to the complexity of assessing them. Several examples provide insight into what works with assessment of these complex personal and social outcomes. In-depth analyses of an individual, such as a case study with three years of multiple data sources, provides convincing evidence of identity development for an individual (Brody, 2005). Investigations of groups of people (Gray, 2004) who share motivation and goals and are involved in a community of learning activities provide convincing evidence of identity development for an individual and the group because records are publicly constructed, recoded in on-line environments, focus on shared meanings and are collected and analyzed over time.

It is important to consider variety in terms of both research methodologies and data acquisition methods. Researchers report the use of both quantitative and qualitative data acquisition methods often in combination. The most effective and

revealing studies concerning identity incorporate multiple methods in both quantitative and qualitative approaches. Among the data acquisition methods reviewed we found interviews and observations as well as quantitative data on patterns of behavioral observations. In terms of research methodologies the literature focusing on identity utilizes case studies, biography, descriptive and ethnographic approaches.

4.3 What has not been done that has potential?

Dierking reported to the American Zoo and Aquarium Association that there is little to no systematic research on the impact of zoos and aquaria on visitor conservation knowledge, awareness, affect or behavior (Dierking 2001). We, too, found no systematic approach to assessing science learning in informal environments generally. We suggest that the idea of an assessment system be considered as a recommendation for future work in the field. The idea of an assessment system to inform us about learning is not new. Wilson and Sloane (2000) have described an embedded assessment system, the so-called BEAR system (Berkley Evaluation and Assessment Research) applied to science learning in the middle grades. In the case of learning in informal contexts, Schauble *et al.* (1997) proposed a framework for organizing a cumulative research agenda.

We do not want to give the impression that an assessment system is a specific linear process or set system. It is not, it is a set of tools from which people engaged in assessment and evaluation in the area can choose. We believe the issue here is that not everyone has all the tools in their box and even if they did, they would not necessarily know how to use each of them. So the challenge is to provide a toolbox with a good set of instructions for use in your own context.

There are, however, several potential problems associated with the idea of an assessment system in the informal science context:

1. ISIs often commission external agencies or individuals to undertake the assessment/evaluation. In doing so they may lose control of and/or fail to appreciate what the assessment is telling them. We would argue that they should have control but it would require that every institution has access to an independent assessment/evaluation advisor.
2. Researchers in universities might be interested in researching what goes on in informal contexts motivated purely by curiosity. They are less likely to be willing to see themselves as needing a system (although they would frequently claim to be working systematically).
3. There is a danger that rather than measuring the important, what becomes important is what can be measured.

An emerging research methodological approach to investigating identity in informal settings is the Learning Lives Research Project (LLRP) (<http://www.education.leeds.ac.uk/research/lifelong/>). In the LLRP the first data acquisition method is a qualitative study of 120 people in the UK, drawn from different walks of life, different parts of the country, and of different ages, genders

and ethnicities. The second data acquisition approach is quantitative, combining two data sets based on interviews and survey, and the third is the fact that it is longitudinal research, in which interviews of subjects occur several times over three years. The LLRP refers to this approach research as Life History Research. Life History Research methodology has the potential to inform and contribute to the field of learning science in informal settings especially as it relates to identity and agency (Hodkinson *et al.*, 2007; Hodkinson, 2007; Lynch 2007 A broad conception of identity has the potential to better inform the conceptualization of affective outcomes in learning science in informal settings. It can take the field beyond the traditional conceptions often derived from the literature in science education by better addressing what actually motivates people to participate in these experiences and activities. It contributes to the argument for deeper, richer assessments over longer periods of time than is often the case in research in the formal education field.

4.3 Conclusions

In our review of the literature we have found that many strategies work in assessing science learning in informal environments, some work better than others in specific situations, but nothing works if it is not grounded in theoretical frameworks that inform learning in informal settings. Any of the methodologies that we have listed in the matrix have advantages and disadvantages. What we have seen is a reliance on descriptive work that is typical of a domain as it begins to develop its theoretical framework. Other methodologies such as biography or phenomenology are suitable to answering certain types of questions but appear to be under-utilized. These under-represented methodologies should be considered further by researchers. For example, biographical studies of outstanding informal science learners such as Jane Goodall (National Geographic/Animal Behavior) or Jack Horner (Museum of the Rockies/Dinosaur Evolution) may give insight into the cumulative outcomes of learning in informal contexts. The key factor remains that in the most effective studies, methodology is selected in terms of the conceptual underpinnings of learning in informal settings.

In terms of methods of assessing science learning, it does not appear there is any evidence that single vs. multiple methods or one method vs. another is of any specific relevance unless it is considered in terms of the actual informal experience. Methods are ‘best’ or most appropriate when assessing learning in direct relation to the informal experience. This understanding has led us to a tentative set of characteristics that should be considered when selecting methods. These include:

- Event-Centered Nature of the Assessment:
 - o Does the assessment mimic, replicate or take into consideration the actual informal learning experience including the physical, personal and social context of the event?
 - o Is the assessment multi-sensory?
- Idiosyncratic Nature of Each Person’s/Group Experience:

- Does the assessment recognize that each person's experience is different, take this into consideration and allow for idiosyncratic learning outcomes?
- Does the assessment recognize socially-constructed knowledge and assess it in appropriate ways, i.e. in a social context?
- Is the assessment interactive?
- Does the assessment recognize that outcomes continually evolve toward greater complexity as learners engagement progresses from everyday experiences to informal events
- Emergent Nature of the Experience:
 - Does the assessment recognize that outcomes continually evolve toward greater complexity as learners engagement progresses from everyday experiences to informal events.
- Integration of Thinking, Feeling and Acting:
 - Does the assessment take into consideration the full experience of the learner?
- Expected and Unexpected Outcomes:
 - Is the assessment sensitive to alternate, idiosyncratic and unexpected learning outcomes?
- Variance in Space and Time:
 - Does the assessment take into consideration precursor, engagement and post-engagement or cumulative outcomes?
 - Is prior knowledge assessed?
- Complexity and Variability of Setting and Audience:
 - Is the assessment designed to ascertain differences between and among people and groups?
- Motivation, Curiosity and Interest:
 - Does the assessment take into consideration learner dispositions, before, during and after the event?
- Related Everyday Informal Learning:
 - Are related outside learning experiences, such as watching TV or reading newspapers taken into account when assessing outcomes of a specific informal science learning experience?
 - Does the assessment connect to learning experiences?
- Identity as Complex Outcomes:
 - Do assessments take into consideration personal and social learning as it relates to everyday life experiences?
 - Can specific science identity outcomes be related to specific types of science experience and activity?

It is unrealistic to think that researchers would take all of these characteristics into consideration, but to be effective, they need to be clear about what they are addressing and what they are not addressing. Any of the characteristics will be better addressed in studies that are driven by theoretical understanding of learning in informal settings.

In addition to the identifying the need for stronger theoretical bases for studies, our other conclusion is that ideology (the What?), epistemology (the How?) and axiology (the Why?) must be aligned within each study. The first step is for researchers and evaluators to recognize and describe these aspects of their work in sufficient detail so that the conclusions may be judged in terms of their relation to the alignment. In the future, assessment and evaluation studies should be conceived, proposed and funded in relation to this criterion.

4.5 Discussion

As museums began to reorient themselves towards an educative role (for example, by providing better interpretation and creating education/learning departments), a niche opened for researchers to study what happened during visits and to evaluate aspects of exhibit design. The growth in museum and visitor studies during the last few decades, while beneficial to the sector, tended to create a formal/informal separation. With the increasing popularity of science centers and museums, researchers have crossed the virtual barrier between the formal/informal sectors and appear to be moving to a consensus that learning is learning wherever it takes place. Both sectors have much to learn from each other and partnerships between informal learning institutions and universities seem to offer enormous potential for intellectual cross-fertilization and for developing and testing theories in practical contexts.

The difficulties faced by researchers, often evaluators trying to find evidence of the impact of exhibits, are well established. Visitors do not want to spend much time doing pre- and post-tests, the time they spend at exhibits is relatively short, the environment is often noisy (making data collection difficult), and it is hard to attribute learning to specific moments during visits. Rather than wringing our hands about the challenges of researching in informal contexts, maybe we should be reflecting on the difficulties faced in researching in formal, usually school-based contexts. Close examination of many school-based studies shows that many of them suffer from the same methodological and conceptual challenges that are normally associated with informal studies.

Turning now to nomenclature. It is clear that the term 'free-choice' has its limitations. The problem is two-fold. Firstly, learning happens all the time and we do not really have much choice in it sometimes. We can have free-choice visiting but free-choice learning does not capture what is happening. Secondly, many of the young people visiting ISIs are not there of their own free will. If free-choice refers only to visitors who come to ISIs of their own volition, then it might be an appropriate term. But what then is the point of delineating visitors who are their of their own free-choice and those who are not? In any class of schoolchildren there are those who would rather be nowhere else but learning math and those who would

rather be anywhere else. It is time to look more closely at the reasons why we delineate between schools and other sources of learning such as museums, science centers, the Internet, etc?

A related issue is the confusion over evaluation and research. Crudely, assessment and evaluation both use research methods. They are both types of research. It might be more helpful to consider the purpose of a study. Carrying out research to build theories or to test them is not evaluation. Carrying out research to see if one type of wording on a sign works better than another is evaluation. Evaluation is about judging the value of something. That is not the purpose of all research though it is the purpose of some. Rather than decry evaluation as not being research (or 'proper' research) let us use the same criteria for judging its worth as we do for other types of research.

Politics, it is said, is the art of the possible, but politicians who expect researchers to provide evidence-based 'scientific' research have a limited understanding of what is actually possible. Scientists make knowledge in a number of ways: a theoretical physicist does not do science in the way that an industrial chemist does. A geologist does not do the same type of science as does an epidemiologist. So, calling for research to be done 'scientifically' is a somewhat meaningless, reductionist and simplistic notion. We judge the quality of research, in the end, by the validity of the knowledge it produces rather than by how it is produced. Denying research findings that appear to be valid because they were not produced experimentally is intellectually bankrupt and morally dishonest.

We continue to move to a situation where schools of the future look to museums and science centers for inspiration. Our knowledge of the learning that happens in ISIs and of the lifelong impact of museum and science center visits should encourage us to dig deeper financially and intellectually. The history of science and science education is hung on the timeline of creativity, and museums and science centers answer the question, 'What's next?'

References

- Aittola, T. (1999). *Possibilities of Informal Learning Inside and Outside the School*. http://www.erill.uni-bremen.de/lios/sections/s2_aittola.html: 8.
- Allen, S. (1997). Using scientific inquiry activities in exhibit explanations, *Science Education*, 81(6), 713-34.
- American Institutes for Research. (2006). *Evaluation of FETCH! Activity Guide*. Concorde, MA: Author.
- Anderson, D., Lucas, K.B., Ginns, I.S., and 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), 658-79.
- Ash, D. (2003). Dialogic inquiry in life science conversations of family groups in a museum, *Journal of Research in Science Teaching*, 40(2), 138-62.
- Black, P. and Harrison, C. (2001). Self- and peer-assessment and taking responsibility: the science student's role in formative assessment. *School Science Review*, 83(202), 43-9.
- Black, P. and Wiliam, D. (1998). Assessment and classroom learning, *Assessment in Education*, 5(1), 7-41.
- Boltanski, L and Thevenot, L (1999) The sociology of critical capacity, *European Journal of Social Theory*, 2(3) 359-77.
- Brody, M, Tomkiewicz, W., and Graves, C. (2002). Park visitor's understanding, values and beliefs related to their experience at Midway Geyser Basin, Yellowstone National Park, USA, *International Journal of Science Education*, 24(11), 1119-41.
- Brody, M. (2005). Learning in nature, *Environmental Education Research*, 11(5), 603-21.
- Creswell, J. (1998). *Qualitative Inquiry and Research Design: Choosing Among Five Traditions*. Thousand Oaks, CA: Sage.
- Crowley, K., Callanan, M.A., Jipson, J.L., Galco, J., Topping K. and Shrager, J. (2001). Shared scientific thinking in everyday parent-child activity, *Science Education*, 85(6) 712-32.
- Delhi, K. and Fumia, D. (2002). *Teachers' Informal Learning, Identity and Contemporary Education 'Reform'*. NALL Working Paper. Access ERIC: FullText. Toronto Canada, Ontario Inst. for Studies in Education, Toronto. New Approaches to Lifelong Learning: 31.
- Dierking, L.D. (2001). American zoo and aquarium association multi-institutional research project. *The Informal Learning Review*, 16-17.
- Edu, Inc. (2004). *It's a Nano World: A Study of Use. Findings from a Summative Report*. Ithaca, NY: Author.
- Falk, J. and Dierking, L.D. (2000). *Learning from museums*. Walnut Creek, CA: AltaMira Press.
- Falk, J., Moussouri, T., and Coulson, D. (1998). The effect of visitors' agendas on museum learning, *Curator* 41, 107-20.
- Falk, J.H. and Adleman, L.M. (2003). Investigating the impact of prior knowledge and interest on aquarium visitor learning. *Journal of Research in Science Teaching*, 40(2), 163-76.
- Falk, J. and Dierking, L.D. (1992). *The Museum Experience*. Washington, DC: Whalesback

- Feuer, M. J., Towne, L., and Shavelson, R. J. (2002). Scientific culture and educational research, *Educational Researcher*, 31(8), 4–14.
- Garrick, J. (1998). *Informal Learning in the Workplace: Unmasking Human Resource Development*. New York: Routledge.
- Garrison, D.R., Anderson, T., and Archer, W. (2000). Critical inquiry in a text-based environment: Computer conferencing in higher education, *The Internet and Higher Education*, 2(2-3), 87-105.
- Gerber, B.L. (2001). Relationships among informal learning environments, teacher procedures and scientific reasoning ability, *International Journal of Science Education*, 23(5), 535-49.
- Goodman Research Group, Inc. (2006). *Summative Evaluation of Einstein's Big Idea*. Cambridge, MA: Author.
- Gray, B. (2004). Informal learning in an online community of practice, *Journal of Distance Education*, 19(1), 20-35.
- Gunawardena, C.N., Lowe, C.A., & Anderson, T. (1997). Analysis of a global online debate and the development of an interaction model for examining the social construction of knowledge in computer conferencing. *Journal of Educational Computing Research* 17(4), 397-431.
- Gutwill, J. and Allen, S. (2007, forthcoming). *Dealing with complexity in visitor studies: Practical research and evaluation approaches*. Book chapter in J. Falk (Ed.). Lanham, MD: Alta Mira Press. (This paper was sent by the liaisons to us for our use in writing this paper).
- Harlow, H. and Jones, A. (2004). Why students answer TIMSS science test items the way they do. *Research in Science Education*, 34, 221-38.
- Hein, G.E. (1998). *Learning in Museums*. New York: Routledge.
- Henri, F. (1992). Computer conferencing and content analysis. In A. Kaye (Ed.), *Collaborative Learning Through Computer Conferencing: The Najaden Papers* (pp 117-36). Berlin: Spring-Verlag.
- Hodkinson, H. (2007), *Learning as Becoming Through the Lifecourse: Contrasting and Challenging Relationships with Identity*. Leeds: Lifelong Learning Institute, University of Leeds.
- Hodkinson, P. (2007) *Learning as Being, Lifelong Learning Institute, Leeds: University of Leeds*.
- Hull, G.A. and Greeno, J.G. (2006). Identity and agency in non-school and school worlds. In Z. Bekerman, N. Burbules, & D.S. Keller (eds) *Learning in Places: The Informal Education Reader*. New York: Peter Lang, pp 77-97.
- Jackson, A. and Leahy, H.R. 'Seeing it for real...?' Authenticity, theater and learning in museums, *Research in Drama Education*, 10(3), 303-25.
- Knight-Williams Research Communications. (2005). *Summative Evaluation of National Geographic's Strange Days On Planet Earth Television Series And Website With an Adult Audience*. Sacramento, CA: Author.
- Lebeau, R.B., Gyamfi, P., Wizevich, K., and Koster, E.H. (2001). Supporting and documenting choice in free-choice science learning environments. In J.H. Falk (Ed.), *Free-choice science education: How we learn outside of school* (pp. 133-48). New York, NY: Teachers College Press.

- Lebeau, R.B., Gyamfi, P., Wizevich, K., and Koster, E.H. (2001). Supporting and documenting choice in free-choice science learning environments. In J.H. Falk (Ed.), *Free-choice Science Education: How we Learn Outside of School* (pp. 133-48). New York, NY: Teachers College Press.
- Leinhardt, G., Tittle, C., and Knutson, K. (2000). *Talking to Oneself: Diaries of Museum Visits*. Retrieved December 29, 2006 from <http://mlc.lrdc.pitt.edu/mlc>.
- Lynch, H. (2007) *Learning as Desire*. Paper presented at ESREA (European Society for Research on Education of Adults) Network Conference on Life History and Biography March 1st-4th 2007 Roskilde University, Denmark
- Malcolm, I. and Field, J. (2005). *Researching Learning/Working Lives: Issues of Identity, Agency and Changing Experiences of Work*. Paper presented at the 35th Annual SCUTREA Conference, University of Sussex, England, UK.
- Meisner, R., vom Lehn, D., Heath, C., Burch, A., Gammon, B., and Reisman, M. (2007, forthcoming). Exhibiting performance: co-participation in science centres and museums, *International Journal of Science Education*.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd. ed., pp 12-103). New York: MacMillan.
- Miles, S. P. A., Stauber, B., et al. (2002). *Communities of Youth: Cultural Practice and Informal Learning*. Burlington, VT: Ashgate Publishing Company.
- 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 outside of school* (pp. 93-114). New York, NY: Teachers College Press.
- Murphy, P. (1995). Sources of inequity: understanding students' responses to assessment, *Assessment in Education*, 2(3), 249-70.
- Randi Korn & Associates, Inc. (2002). *MarsQuest Summative Evaluation*. Alexandria, VA: Author.
- Rennie, L.J., Feher, E., Dierking, L.D., and Falk, J.H. (2003). Toward an agenda for advancing research on science learning in out-of-school settings, *Journal of Research in Science Teaching*, 40(2), 112-20.
- Rickinson, M., Dillon, J, Teamey, K. Morris, M., Choi, M.Y., Sanders, D., and Benefield, P. (2004). *A Review of Research on Outdoor Learning*. London: National Foundation for Educational Research and King's College London.
- Rounds, J. (2006). Doing identity work in museums, *The Curator*, 49(2), 133-50.
- Schauble, L., Leinhardt, G. and Martin, L. (1997). A framework for organizing a cumulative research agenda in informal learning contexts, *Journal of Museum Education*, 22(2-3), 3-8.
- Siegel, D., Esterly, J., Callanan, M.A., and Wright, R. (2007, forthcoming). Conversations about science across contexts in Mexican-descent families, *International Journal of Science Education*.
- Stake, R.E. (1995). *The Art of Case Study Research*. Thousand Oaks, CA: Sage
- Storksdieck, M., Ellenbogen, K. and Heimlich, J.E. (2005). Changing minds? Reassessing outcomes in free-choice education, *Environmental Education Research*, 11(3), 353-69.
- Tenenbaum, H.R., Rappolt-Schlichtmann, G. and Zanger, V.V. (2004) Children's learning about water in a museum and in the classroom, *Early Childhood Research Quarterly*, 19(1), 40-58.

- The Learning in Informal and Formal Environments (LIFE) Center: An NSF Science of Learning Center. (n.d.). Retrieved January 21, 2007 from http://life-slc.org/?page_id=124.
- Toma, C. and Rogoff, B. (1997). Shared thinking: Cultural and institutional variations. *Discourse Processes*, 23(3), 471-98.
- Wells, E. (2005). Emphasizing main idea and higher order concepts to improve teaching and learning of non-native plants and biodiversity, *ARexpeditions: Action Research Journal*, Montana State University, Bozeman, MT found at: <http://arexpeditions.montana.edu/articleviewer.php?AID=84&PAGE=1>.
- Whitehead, J. and McNiff, J. (2006) Action Research Living Theory. London: Sage.
- Wilson, M. and Sloane, K. (2000). From principles to practice: An embedded assessment system, *Applied Measurement in Education*, 13(2), 181-208.