Listening to Young Children: Assessment and Research Techniques for Very Young Visitors

Christine Massey
Swarthmore College
Swarthmore, Pennsylvania

Introduction

In recent years there has been a lot of growth in museums that cater to families with very young children. Zoos, aquaria, and science museums all have significant numbers of young children visiting them, and museums designed especially for children have been one of the fastest growing segments of the museum community. Understanding and serving the needs of an audience that includes small children is clearly necessary; however, many of the methods most commonly used in visitor studies – for example, written questionnaires, focus groups, and interviews – often are very difficult to implement with children. The intent of this paper is to survey methods from other fields, particularly developmental psychology, in order to gain insight into how one might allow young children to be direct participants in visitor evaluation and research studies, with a special emphasis on investigations involving learning.

The systematic study of children and their development has a relatively short history. Prior to this century, one might have seen the occasional case study report, or philosophical speculations about children's perceptual abilities, moral evolution, or thought processes. In the early decades of this century, much attention was paid to infants' and children's motor and physical development, and rigorous measurements were made, especially by Gesell, who established maturational norms. With the rise of mental testing, studies of children's intellectual functioning became more widespread, but attempts to conduct standardized tests of children's cognitive performance and achievement before roughly school age were not very productive nor predictive. Indeed, one could argue that it wasn't until the Swiss psychologist Jean Piaget and the Soviet psychologist Lev Vygotsky introduced new methods and theories that revolutionized the scientific study of children's development, that one could say that psychology was finally "listening to young children." (See Cairns, 1983 for an overview.)

Jean Piaget has probably been the most influential investigator of the cognitive development of infants and young children that the field of psychology has known. The creativity, comprehensiveness, and insight with which he explored and theorized about how cognitive functioning unfolds

from infancy through childhood has yet to be matched, and the legacy he has left to the field is a complex one. On the one hand, he was typically the first person to raise most of the significant topics and questions that continue to dominate the field to this day. His observations of his own three children as infants set the agenda for hundreds of subsequent studies of infancy. The surprising results of his investigations with older children, including his classic conservation experiments, took American psychology and education by storm in the 1950's and 60's. Although his methods were often unorthodox, he was simply brilliant in revealing patterns and phenomena in children's thinking and development that no one had ever noticed or considered before. Among the most familiar aspects of Piaget's theory is his hypothesis that cognitive development proceeds through a sequence of four general stages. This is not the time or the place to review Piaget's theory, but I mention it because research with children in the second of Piaget's stages, which he termed the "preoperational period," ranging from about ages 2 through 8, has been the center of much of controversy and bears on our topic today of listening to young children.

In recent years, Piaget's work has been criticized and, in some cases, contradicted. His overall theory has not survived unscathed, and new discoveries and developments have led in new directions. Technological and methodological advances have allowed us to study infants in ways that were unknown in the years during which Piaget developed his theory, and contemporary views of infant development are somewhat different from Piaget's depiction. And a new look at young children — children between the ages of about 2 and 7 — has led to a reassessment of their knowledge, skills, and abilities, as well as developmental processes more generally (see Carey, 1985 or Gelman & Baillargeon, 1983 for reviews). Even for the pre-eminent Piaget, the preschool age group turned out to be a tough nut to crack.

In the stage theories that have dominated developmental psychology until recently, the children in this age group have largely been treated as foils for older children. They were most often described in terms of knowledge and abilities they lacked, tasks that they failed, and cognitive operations that were unavailable to them. But it has turned out that these descriptions often seriously underestimated preschool children and stemmed from difficulty in developing appropriate methods for working with them (Gelman, 1979). These children are too old to use the habituation and looking-time paradigms that have been successful with infants. But they are too young to complete tasks that require a lot of verbal ability or the use of written or other symbolic systems. Piaget's interviews with children, for instance, are thought by many to rely too heavily on the child's ability to interpret complex verbal questions and descriptions and to verbalize their own ideas and thinking processes. Open-ended interview questions and written questionnaires — often the methods of choice for evaluation research — are extremely poor choices for use with younger children. Either the

children simply cannot perform the task at all and so the investigator drops them from the sample, or their performance is so poor that misleading (and often disheartening conclusions) are drawn about them: they didn't learn anything, the content was too hard for them, they didn't pay attention, and so forth. One of the key lessons of developmental research with this age group is that it is easy to demonstrate performance differences between older and younger children. It is much more difficult to assess the underlying competence of young children and to explain *why* they perform differently from older children on many tasks.

Choosing Methods for Studies with Young Visitors

So what advice might we apply to evaluation and research methodology with young children in mind? One place to begin is with a full analysis of what your purposes are in undertaking some form of research or evaluation. What questions are you most interested in answering?

Are you evaluating a particular exhibit or program?

Do you want to assess whether visitors have learned something from using an exhibit or from some other museum experience?

Are you interested in *optimal* or *typical* performance at an exhibit?

Do you want to know what the visitor is bringing *to* the exhibit or program — in terms of interest, knowledge, and prior experience?

Are you studying visitor behavior more generally — how families spend their time, interact, make choices, use resources, direct attention, and so forth?

Your answers to these sorts of questions are the first factor to establish when deciding on methods. Our assumption in this session is that you also have reasons to include young children in your investigations. What then are the advantages and disadvantages of various research methods with respect to the kinds of research questions they are suited to addressing and to their implementation with young children?

Just Ask Them: Self-reports, Interviews, and Questionnaires. It's often tempting to assume that the best way to get information from someone, particularly if the kind of information you want is not externally observable, is simply to ask them. Unfortunately, even for adults, this method often is not as useful as it first seems.

• Self-reported behavior. Self-reports of behavior are so often unreliable that one should generally either avoid them or supplement them with independent measures. There is often a significant gap between what people think they do and what they actually do, and the larger the time frame, the worse people seem to get. People's constructions of their own actions and

experiences are fascinating in themselves and may separately be of interest to an investigator, but one should generally not assume that self-narratives faithfully or objectively reproduce past behavior. Young children in particular usually lack metaconceptual awareness of their own behavior and may have trouble adding frequency or time scales to their accounts of what they do. They often cannot accurately report how often or when or for how long they might do something. (Adults may not be much better.) Parents reports about their children may also be selective, inaccurate, or incomplete. If self-reports are selected as a method, specifically anchored questions (e.g., Have you been to our children's zoo since the duck pond was added?) or general ratings (e.g., Would you say your family visits the children's zoo very frequently, occasionally, or rarely?) may be better than requests for quantified accounts (e.g., How many times have you been to the children's zoo this year?).

• Interviews and questionnaires. Asking people to answer general or open-ended questions, whether verbally or in writing, may introduce variables that you cannot control or assess. Some of these include the participant's vocabulary knowledge, facility in answering various types of questions (e.g., multiple choice or checklist vs. open-ended), interview length, support available from others (which may complicate the extent to which a child is really speaking for himself or herself), and motivation to take time to complete the interview or questionnaire, especially if there are attractive alternative activities nearby.

Given the variability in visitor populations, literacy skills and verbal expressive abilities may be very hard to predict. It is often hard to be sure that a child is answering the question you thought you asked. Sometimes subtle differences in the way a question is phrased yield large differences in the kinds of answers children give. Other times, it seems that no matter what question you ask or how you phrase it, all you are going to get out of a particular preschooler is something like "I liked the snake." Even if young children are interested and are able to understand your questions well enough, you should generally count on getting mostly very short, unreflective answers. General explanations of exhibit content (i.e., "why" questions) are difficult for young children. It is usually easier to ask a child to make a choice or a prediction and then invite them to tell you more about why they made that particular response. It requires experience and pilot testing to phrase questions in such a way that they are productively related to the research question and that they are responsive to the visitor and their state of understanding, but not too general or too leading. Finally, developing a reliable coding scheme for open-ended responses also requires expertise.

When I develop interviews for young children, I am always careful to keep in mind the limitations of the method and to tailor it as much as possible to the youngest age group I'll be working with. Often there is no substitute for actually talking with a child, but I generally avoid choosing a

method that consists exclusively of open-ended questions, whether written or verbal, with no concrete referents or performance component.

Let Them Show Rather than Tell: Interactive Assessments

The techniques considered in this section involve asking the participant to *do something* that is relevant to the learning or behavior you are interested in. This is often a very effective method for children, but it must be well-designed and you have to be clear about what you want to assess, completing a detailed analysis of the exhibit or experience you want to evaluate. Frequently these kinds of performance assessments are the method of choice in much other developmental research with this age group. This is often a very good way to assess visitors' competence with the content or performance relevant to a particular exhibit or activity. You might ask children to sort or classify items in a relevant way, to perform an action, predict an outcome, detect a performance error, or generalize to a new situation. For example, if you want to assess whether children understand how a particular apparatus functions, you might ask them to make it work for you or you might change some elements of it and ask the children to predict what will happen. Or you might use it yourself but make a deliberate mistake in the way you use it and see if the children can identify your mistake and correct it. All of these methods are much more likely to engage young children and to reveal more of their competence than an open-ended question like "tell me how this works" or "tell me what this shows." These sorts of methods when combined with appropriate control groups can be used as learning assessments or for front-end evaluation.

Watch What They Do: Observational Methods

Some kinds of research questions are better suited to observational methods rather than face-to-face assessments. If you can identify externally observable behaviors that address the research question you want to answer, then direct observation may give you more detailed and more reliable information than interviews, interactive assessments, or questionnaires. Observational methods are especially appropriate when you are interested in typical or spontaneous behavior. Another advantage is that it gives you time and sequence data, should you want that. But there are some common pitfalls to beware of when planning an observational study.

One of the most common errors I've seen is to get the data first and then decide what one is looking for. In the very beginning phase of a research program, when you are exploring the phenomena and trying to define a research agenda, it may be worthwhile to do some open, free-form observations, but these data will lack the validity and the reliability that are required if you want to make generalizations or draw firm conclusions.

Once you have decided on observation as a method, you will have to decide whether you are going to do live or recorded (e.g., videotaped) data collection. Live observations are often very effective and efficient once you

have a good coding scheme and trained coders. The difficulty of carrying out live observations is affected by how well one can see and hear what is going on and by how many dimensions of an interaction you want to capture. Investigators sometimes try to get a little bit about everything for every member of an interacting group — social, cognitive, physical, and so forth. The results can be disastrous. I suggest that a more constructive strategy may be to do several studies, each of which goes after one thing very clearly, rather than trying to get everything at once. For example, one study might analyze only the content, form, and pragmatics of questions a family asks at an exhibit. Observational studies require significant amounts for observers to learn the coding system and practice the method until they can use it reliably and efficiently. Experienced, well-practiced observers using a reliable, sensible coding scheme generally yield high-quality data. This requires an up-front investment in planning a well-focused study, devising a good coding scheme, and practicing the data collection procedure until it is smooth and efficient (Bakeman & Gottman, 1986). But it saves a lot of time and frustration later.

In designing an observational study, it is also important to understand appropriate sampling techniques (see Altmann, 1974). For example, a mistake to avoid is observing only those visitors who catch your attention as doing something interesting. This severely biases the database and will cause you to overestimate what are likely to be unusual or low frequency behaviors. For example, if most visitors use an exhibit for less than 30 seconds and are minimally involved but you don't bother to record those people because they don't seem very interesting, your data will yield a very skewed view of what is happening at that exhibit. Working from your research objectives, you must decide in advance the "rules" that make someone eligible for being observed and then stick to them if you want to make any generalizations about how representative the observed behaviors are of exhibit visitors in general.

In my experience, recording observations often sounds more appealing than it actually turns out to be. Video often gives you both more and less than you actually want. It is a naive belief that the video will capture everything — audio is often hard to hear; unless you have a good camera person, important action will take place out of the frame; people walk in front of the camera; actors are turned away from the camera at crucial times, and so forth. It is a common experience that audio or videotaped data make the most sense to someone who was present when the recording was made—which implies to me that the recording is capturing less than one might expect and the role of the live observer is doing more. This also yields a serious practical problem: if you record and then code video later, if the coder was present at the recording, you will be relying in undetermined ways on the coder's ability to "fill in the blanks." There is simply *no good way* to assess whether the coder does this accurately. It is likely that coder will introduce inaccuracies, especially if much time has passed between

recording and coding and especially if the coder has observed many similar interactions (a common phenomenon in cognitive psychology called "repisodic blurring"— repisodic being derived from "repeated episodes" (Neisser, 1981).

Assuming that an audio or video recording is of good enough quality, you still have a big task ahead in terms of coding it. Researchers typically spend many coding hours for each hour of video or audiotape. The importance of knowing what you are looking for and having a clear, well-devised coding scheme that independent observers can use reliably is just as important for recorded as for live observations. (In my experience, people often end up choosing to use recorded data because they are simply deferring the work of developing the coding scheme—not because they really need recorded data as the raw data.)

There are times, however, when having a recorded dataset is extremely useful. The ability to replay interactions is most helpful if you are after very subtle phenomena (and the recording quality is high and consistent) or if you are coding many different aspects and levels of the same interaction. (You still have to go through process of checking reliability for independent coders.) Videotaped data can also be useful in training new observers, though the process of coding from recorded data and from live action are somewhat different. Transcripts and video archives can be a good resource for reanalysis; you can go back and ask new questions or use the database to generate and evaluate new hypotheses for further investigation. However, one should avoid falling into a circular method in which the same database is used to generate and then test new hypotheses. Videotapes and transcripts can be invaluable for communicating and demonstrating the phenomena you are studying. They provide a good source for examples to show in presentations or to summarize in papers by providing a supply of documented anecdotes or case studies.

A final word on observations with young children: the pace of change in behavior sequences tends to be very rapid and episodes are short. Young children are generally in the company of other people, and there are frequent short cycles of interaction with their companions and with objects around them. You need tracking rules that allow you to decide what you are supposed to be observing at a given time (e.g., sticking with a target child, tracking a child and adult or peer as an interacting unit, coding all instances of a particular behavior, coding any use of a particular exhibit, etc.). You also need a coding system that is agile enough to keep up with the pace of behaviors, letting you capture what is significant out of the rapid stream of behavior.

As these considerations indicate, working with young children presents some special challenges. It requires a researcher or evaluator to remain flexible and creative while also keeping a clear fix on his or her objectives and standards. But those who are willing to make the effort to really listen to young children have the reward of discovering that they are one of the

most charming, entertaining, and fascinating audiences one could ever work with.

References

Altmann, J. (1974). Observational study of behavior: Sampling methods. *Behaviour, 59*, (3-4), 227-267.

Bakeman, R. & Gottman, J. M. (1986). *Observing interaction: An introduction to sequential analysis*. New York: Cambridge University Press.

Cairns, R. B. (1983). The emergence of developmental psychology. In W. Kessen (Ed.), *Handbook of child psychology, 4th Ed.: Vol. 1. History, theory, and methods*. New York: John Wiley & Sons.

Carey, S. (1985). Are children fundamentally different kinds of thinkers and learners than adults? In S. Chipman, J. Segal, & R. Glaser (Eds.), *Thinking and learning skills: Vol. 2. Research and open questions*. Hillsdale, NJ: Lawrence Erlbaum.

Gelman, R. (1979). Preschool thought. *American Psychologist, 34* (10), 900-905.

Gelman, R. & Baillargeon, R. (1983). A review of some Piagetian concepts. In J. H. Flavell & E. M. Markman (Eds.), *Handbook of child psychology, 4th Ed.: Vol. 3. Cognitive development*. New York: John Wiley & Sons.

Neisser, U. (1981). John Dean's memory: A case study. *Cognition, 9*, 1-22.