

## Should The 51% Solution Have A "Caution" Label?

Harris H. Shettel  
Rockville, Maryland

*Things should be made as simple as possible,  
but not too simple. (Albert Einstein)*

When Steve Bitgood asked if I would be willing to write a commentary on Beverly Serrell's article on the 51% Solution (this issue), my first reaction was an ambiguous "Let me think about it." From the very first time I was made aware of her ideas on this subject several years ago, Beverly and I have had many "deep" discussions and shared long memos relating to her approach to determining exhibit effectiveness. More recently, in a talk I gave at the Raleigh VSA conference in 1994 ("Do We Know How To Define Exhibit Effectiveness?"), I made specific reference to the 51% Solution and summarized some of the problems I saw in its adoption. My "tone" was essentially negative, as reflected in this quote from my talk: "While I admire anyone who is willing to stick his or her neck out and say 'Here is what I think characterizes a good exhibit,' it would be unfortunate if we accepted those criteria prematurely as really distinguishing between effective and ineffective exhibits." (By the way, after my talk Beverly publicly challenged me to a mud wrestling match at the next VSA conference in St. Paul. However, she did not show up at that conference, giving as her reason the fact that she is working on an up-dated version of her excellent book on how to write effective labels. That's going pretty far to avoid a good fight!)

As I thought more about what Beverly is trying to do, and on comments she has made to me about her work on the 51% Solution (e.g., a "first step," a "work in progress," "changes will be made"), I decided that I should cast my earlier comments in a somewhat broader context. In fact, the very title of her paper, "The 51% Solution Research Project..." indicates that it represents an *ongoing* investigation of a novel approach to measuring exhibit effectiveness to see if, or how well, it "works," and make changes as necessary. (Hence, the title of this paper, reminding those who are participating in her study that they are being asked to collect certain kinds of visitor data in certain special ways that have yet to be fully developed or verified.)

Beverly is also careful to point out what her approach is not, as she helpfully does in the body of her article. For example, it is not being proposed as a "solution" in the sense of solving all of our exhibit evaluation problems, but a "solution" in the sense that it uses certain kinds of visitor data from three broad dimensions (time, use, and impact) in order to produce an index of effectiveness (at least 51% on each

dimension). "Solution" = "Mixture." Also, this approach was never intended to be applied to other than completed exhibits *in situ*. Thus, it would not be appropriate for use in front-end or formative evaluations, only for summative evaluation. These are helpful reminders to those who are contemplating its use.

There are several "plusses" to the approach that Beverly is investigating. For one thing, it is based upon the three conceptual pillars of summative exhibit evaluation: Attracting power, Holding power, and Communicating power. It has been accepted almost from the very beginning of serious visitor studies that a "good" exhibition (or exhibition element) is one that attracts visitors, holds them long enough to get its messages across, and actually gets those messages across. Unobtrusive visitor tracking studies and pre-post interview/questionnaire data consistently reveal a veritable gold mine of valuable information about these three independent (orthogonal) variables for a specific exhibition or element thereof, including what could be done to reduce any problems or weaknesses that may have been identified in any of these three areas.

But, as Beverly reminds us, the methods that have evolved over the years for collecting these kinds of summative data are cumbersome, expensive, and time consuming, which, she believes, accounts for their (relative) infrequent use in the real museum world. I would agree that if it were possible to collect similar data in a "simple yet rigorous" way (her words), we might see a greater utilization of summative-type studies.

However, data collection methodology is not the only problem with our current methods. When we get through collecting our summative data we still do not have an acceptable way of relating our findings to a standardized effectiveness scale, one which would allow cross-exhibit comparisons to be made. We are faced with a serious unanswered question: On a scale ranging from "excellent" to "lousy," just where does a particular exhibit stand relative to other exhibits? We know how many visitors the exhibit *did* attract and *did* hold, but how many visitors *should* it have attracted and held? We also know how many messages visitors did receive, but how many should they have received? As I noted in my Raleigh paper, the sad truth is, we really don't know. (Saying 100% is an unrealistic copout. It will almost never happen.) And so we "eye ball" the data and say, "I don't think enough visitors are going over to area X," or "Too few visitors stay very long at element Y."

(This argument, it should be clearly noted, is not the same as saying that we do not know how to find out if modified "version two" of an exhibit is better or worse than "version one" of the same exhibit. Ten minutes is greater than five minutes, and ten messages is greater than three messages. These are ordinal measures that allow valid judgments of relative effectiveness to be made within a single exhibit. Such measures do not, however, allow com-

parisons to be made between any one exhibit and other exhibits, which requires measures arranged on an interval scale, such as ratios or percentages.)

And so, bottom-line-wise, if it were possible to simplify the collection of summative time and impact data without compromising their inherent value, and, at the same time, standardize the collection and interpretation of these data so that meaningful effectiveness comparisons could be made across exhibitions of all types and sizes, it would represent a breakthrough of major proportions, both practically and theoretically, for the visitor studies field. Showing that this is indeed possible is the challenge Beverly faces in carrying out her research project.

While I am not privy to the complete research agenda that Beverly intends to employ in her analysis and refinement of this "unique ... methodology", there are several issues that I think will need to be addressed if this quest for "useful information for making decisions about exhibition development and evaluation" is to be successful.

It goes without saying that testimonials from "satisfied users" of the 51% Solution, while nice to get (and Beverly has several), should not be confused with data-based, empirical support. For a person who has never systematically observed and recorded visitor behavior in their exhibit spaces, it is almost always an "eye opener" (and not usually a pleasant one). "I had no idea that hardly anyone went to that side of the room," or "People walk past the gem cases like they weren't there!" Tracking data has always had this capacity to destroy naive assumptions about how much visitors "enjoy" our exhibits. Also, once attention has been directed to these problem areas, there are often fairly obvious "fixes" that can be made (e.g., move an exhibit element to a more prominent location, provide better lighting, move the tall exhibit behind the shorter one). Results are almost always positive and immediate. Thus, the question is not "Does the 51% approach provide useful information related to these kinds of basic insights about visitor behavior?" (it obviously does), but, "What is there that is uniquely advantageous about this approach as compared with the way many of us do things now, that would recommend its use?"

Looking first at the instructions sent to participants for the collection of time and use data, there appear to be mixed signals. While Beverly asks her observers to count only 2 to 3 seconds at an exhibit element as constituting a "stop," her tracking data collection forms suggest recording visitor behaviors at each element they stop at (read, point, touch, etc.), along with various kinds of observable demographic data (gender, age), and the path the observed visitor takes from one stop to another throughout the exhibit area. Thus, the 2 to 3 second rule does not appear to save the observer any time or make the recording task any easier than the more traditional ways of collection and record-

ing visitor time/use data. In fact, the tracking data forms that Beverly asks her participants to use look remarkably similar to the ones that I (and others) have been using for many years.

What contributes to my confusion is that the only data that is required to be sent to Beverly is the total number of potential stops (elements) in the exhibition, the number of stops actually made for each visitor (as defined by the 2 to 3 second rule), and the total time spent in the exhibit area for each visitor. With these data Beverly can then compute the average number and percentage of stops made by the observed visitors as well as the average total time spent by them in the exhibit area.

However, it would be just as easy and no less time consuming for the observer to record the total time spent by the visitor at each element since they have to "be there" anyway to record visitor behaviors and the overall path taken. This total "element" time figure would appear to be a more sensitive measure of element usage (along with the behavioral data), and the one most often used in past tracking studies (called *holding time*, or *holding power* if a ratio of time spent vs. time required is used).

My question is, why is readily available and potentially valuable data/information being ignored in this study? If the answer is to simplify the data collection process (among the more important advantages claimed for the 51% approach), why even bother to ask observers to record visitor demographics, behavior, and path through the exhibition, in the first place? It is true that such information might be useful to have, as Beverly clearly states in her article, but participants are not told how to use it and Beverly doesn't want it!

I would think that one would want all of the available data at this early stage of the study, on the basis of which one could then determine empirically at what point reductions or simplifications in those data significantly reduce their usefulness. As it stands now the 2 to 3 second rule appears to be an established figure rather than a working figure. I don't think that I am the only person who will want a solid answer to the question "How did you select 2 or 3 seconds as being an adequate representation of the actual time spent at an exhibit element?"

"Total time on task," by the way, has been found to be one of the most useful predictors of educational or training effectiveness, and has been used for this purpose in countless studies. It usually correlates quite well with outcome measures of various kinds, including cognitive, affective, and behavioral. Adding more hours to the school day or more school days to the school year are notions based on this well-established (but far from perfect) "connection." Thus, using average holding time or (better yet) average holding power at an exhibit element as one reasonable indicator if its ability to communicate its messages to visitors (the "bottom line" of most of our studies) is not based on a personal whim but on many years of research-based findings. Those who suggest new

data collection methods have at least as much of an obligation to support and validate them as do those who "defend" the "old" methods. Beverly's study has the potential to do this, but only if the appropriate data are collected and analyzed.

Beverly does use "total time on task" when it comes to the overall exhibition. Thus, she includes in her requested data set the total time each sampled visitor spent in the exhibition. Most would agree that such a figure has meaning only in the context of other factors such as the size, content, complexity, and density of a particular exhibition. One must take into account at least some of these differences if the total average time in any one exhibition is to be interpretable on a "universal" scale (which, as noted, is one of the important goals and advantages of using the 51% Solution).

To achieve this aim, Beverly has selected a single figure as the basic datum - the overall square footage of the exhibit area. This figure, divided by the average time visitors spent in the exhibition area, is the basis upon which the notion of "square feet per minute" is derived. This figure is intended to tell us how quickly (or slowly) the average visitor moves through that space. An upper limit of 300 sq.ft./min. has been established as the cutoff figure. Thus, an exhibition in which at least 51% of the visitors move through it at 300 sq.ft./min. or less (slower), is considered an "effective" or "acceptable" exhibition in terms of this variable.

This has been one of the hardest parts of the 51% approach for me to understand. It is certainly true that a square footage figure is a lot easier to obtain for an exhibition than the linear footage (pathway) that would take the visitor past all the exhibit elements. It would thus be the "figure of choice" if it could be shown to relate to the way exhibits are actually "used" by visitors. I remain skeptical on this point.

Rate of speed in physics is always computed on a linear basis, not a "square" basis. Beverly now refers to the "square feet per minute" figure as a "sweep rate" measure rather than a speed measure. Given the wide variety of "sight lines" that exhibits of various kinds have, I fail to see how this new terminology helps to make the 300 sq.ft./min. calculation any more understandable or meaningful.

For example, let us visualize a room of 6000 sq.ft. that is lined with dioramas, and the same room filled with exhibit cases or "elements" that include videos, interactive games, hands-on devices, etc. It is hard for me to imagine that visitors would "treat" these two spaces the same way or that the distance they would actually travel in them would be comparable.

I happen to be working with an exhibition that is 6000 sq. ft. I have a very detailed floor plan showing the actual location of all the exhibit elements, drawn to scale. Using a string I was able to quickly trace a path that would pass by all the elements. An easy conversion showed that this path is 864

linear feet. If this same overall exhibit space were lined with dioramas around the wall, the linear path would be approximately 300 feet; the same square footage, but certainly not the same "task" for the visitor.

Now let's look at the 300 sq.ft./min. figure that is offered as the upper limit for an "okay" sweep rate visitors can take through all exhibits. If one converts this figure to a (linear) miles per hour figure, we arrive at a speed of 3.4 miles/hr. In this linear form (the only one I can conceptualize) this is just slightly below a fairly brisk and steady walking speed. In her article, however, Beverly says that this pace allows visitors to "stroll through the exhibit, look around, stop occasionally, and sometimes stop long enough to look closely, read, and/or interact with an exhibit element, or interact socially..." If applied to the two configurations described above, visitors would be in the "okay" group if they went past all the dioramas (300 linear ft.) in about one minute, and through the actual exhibit that is planned for this 6000 sq. foot space in just under 3 minutes!!

If we use the 6000 sq.ft. figure only and ignore what is going to be put in that space, the 300 sq.ft./min. figure would give visitors as little as 20 minutes to do whatever there was to do, perhaps too much time for some (rare) events, but most likely much too little time for the vast majority of events. Since only 51% of the visitors are required to meet even that level of "exposure," we can predict that many exhibitions will meet this criterion.

It seems to me that while the total time data calculation has achieved a kind of mathematical comparability (6000 sq.ft. is 6000 sq.ft. no matter where it is located), it has lost its connection with meaningfulness. The real exhibit being planned for the 6000 sq.ft. space is a very dense combination of objects, labels, hands-on activities, videos and an introductory orientation film (the latter of which will itself require about 10 minutes to view) and, as such, is very typical of exhibits being produced these days. The exhibit planning staff expects (unrealistically, I believe) that the average visitor will spend between 1 and 1.5 hours in the space. If a visitor read, looked and did everything there was to do, I would estimate closer to 3 hours would be required (even more unrealistic!).

It should be noted (again) that the application of the 300 sq.ft./min. rule to this exhibit would almost guarantee that at least 51% of the visitors would be in the acceptable category, even though I (and everyone else connected with the exhibit) would consider those visitors who took even 30 or 40 minutes to traverse the exhibition to be in a "problematic" (too fast) category. Under this scenario, the exhibit would probably be declared a success by the 51% Solution time criterion, but should actually be declared less than that by any kind of realistic performance-based criteria (what it takes to reasonably "do" the exhibition properly). In one case the museum staff throws a party; in the other case, they (should) go back to the drawing

board to find out what went wrong and what can be done about it.

(Readers can easily do their own reality check "at home." Select one or more exhibitions you are familiar with and for which you know the square footage, and compute the "acceptable" average time figure, using the 300 sq.ft./min. rule: sq.ft. divided by 300 = time in minutes. If you know the linear distance through any of the exhibitions, do the same calculation using that figure. Do the time figures make sense? If not, why not?)

The lack of sensitivity in the element time data, by the way, makes it very risky to use these data to correlate with other measures, especially impact measures. One needs to have a range of data (actual time spent at elements), not dichotomous data (stopped, did not stop) to make such computations. This problem does not apply to the "total time in exhibit" data.

In discussing these time issues with Beverly, she has said that I am completely unrealistic and naively idealistic (or words to that effect) in expecting visitors to spend hours in a single exhibit. Based on the vast amount of time data collected over the past 20 years or so, I am well aware of the fact that visitors will not do so for the vast majority of exhibits as presently conceived and designed. However, I am not the one who expects these long visitation times, it is those who do the exhibits who "expect" them, often, I am sure, without even realizing it. However, only by taking their "expectations" seriously, can we show them that they must drastically modify the scope and/or design of their exhibits. By "allowing" visitors to completely miss 49% of the elements in an exhibit and to stop for only 2 to 3 seconds at the remaining elements, would seem to me to be letting off the exhibit developers much too easily (although Beverly says that most of the exhibits she has studied don't even reach *her* levels!). Under these conditions, improvements in exhibit conception and design will be shooting at a lower than desirable mark, missing the opportunity to find out just how "good" exhibits *can* be.

Finally, in a landmark study of this kind, one must be certain that one has a firm grasp on the dependent variable, which in this case is the intended impact of the exhibition on the intended visitor. How long visitors spend and what they do in an exhibition are of little interest or value in and of themselves. They obtain meaning largely because we believe that, along with a host of other variables, they are related to, and influence the attainment of, the goals and objectives of the exhibit.

The forms sent out to participants in this study contain two questions related to the impact of the exhibition on visitors. (This sample of visitors, by the way, is recruited at the beginning of their visit and are not taken from those who are part of the unobtrusive observation group). The first question is "What is the main idea of this exhibit?" with two suggested prompts "To show. . ." and "To make people. . ." The second question is "What

is one new idea you are taking away with you?" along with the prompts "I never knew..." or, "I never realized that. . ." and/or "It reminded me that. . ." When visitors complete their written answers to these questions, they are asked if there is "Anything else?" they wish to add.

Since these are exhibit-related kinds of questions, they avoid the necessity of requiring pre-exhibit data to be compared with post-exhibit data (gain data). They could be answered only by those who had seen the exhibition. They also avoid the necessity of designing questions specifically for each exhibit being evaluated. They are very easy to administer (but not, I would think, easy to accurately and reliably score). Thus, this phase of the 51% approach meets the criteria of "easy to learn and easy to use," as well as providing comparability across exhibits (same questions).

However, it is in the all-important area of the meaningfulness of these standardized questions to the wide variety of goals and objectives of the many different kinds of exhibits being studied, that I think additional thought needs to be given. Once these questions are "validated" I think everyone would be more than happy to use them in their exhibit evaluation studies; until that time, I would be hesitant to claim that as little as one word can really tell us "the degree to which the exhibition achieves its stated objectives with its intended audience." Here again, we are being asked to accept surrogates for more comprehensive impact measures before it has been demonstrated that they are indeed adequate for this role.

Despite my somewhat lengthy list of "concerns," I think that Beverly is attempting to do something worthwhile and is headed in the right general direction. Simplification of data collection and standardization of results are two worthy goals, and have been recognized as such for a number of years. Some of us have the wounds to show how contentious these subjects can be. There are those who also believe that we can approach these worthy goals only slowly and painstakingly. Beverly obviously does not want to wait, and has jumped in forcibly and visibly with her own version of how to achieve them. She knows that in doing so she has made herself a very visible target. And, I am sure that she knows that being a javelin catcher is always a lot harder (not to mention more dangerous) than being a javelin thrower. But we really need both.

I have offered my comments over the years (usually solicited) in the spirit of constructive criticism, leading to what I hope will be more reliable, more valid, more generalizable, easier to obtain, and, overall, more helpful visitor data. I also hope that as many data collectors as possible will join in this enterprise. I urge you to send your data to Beverly so that she can have as large a data base to work with as possible. But also send her your comments, your questions, and ideas. Beverly is offering all of us the unique opportunity to share in the evolution of what may become an important contribution to our field of study.