

# WHY BEHAVIORAL INTERVIEWS SUCCEED IN STUDIES BUT FALTER IN THE FIELD: THE MILES TO GO BEFORE WE SLEEP.

Dr. Tom Janz  
Chief Scientist and CEO  
Global Human Asset Systems, LLC



And lead author of the book:

Behavior Description Interviewing: New, Accurate, Cost Effective. (1987, Prentice-Hall).

Most staffing strategies end with an interview that leads to offers of employment. Many others include screening interviews and multiple social as well as assessment interviews. Most large organizations train their recruiters and hiring managers in one or another form of behavioral interviewing. Many hiring managers use behavioral interview guides that contain suggested questions, follow-up probes, and even behavioral guidelines for rating answers. All would appear to be well in the woods known as Behavioral Interviewing.

Yet I argue that the research-proven promises of strong decision power for behavioral interviewing have not been kept in practice. I base that assertion on the following personal observations made over the 20 years since my Behavior Description Interviewing book first appeared. I call it my Poor Practices List:

1. Few field interviewers take anything but cursory notes—that they themselves find difficult to read 20 minutes after they jotted them down.
2. Field interviewers rarely rate each interview answer, summing the ratings to form a total score.
3. Field interviewers often fail to hear when candidates answer behavioral questions with a non-behavioral answers, allowing (and even encouraging) them to go on about their preferences and advice instead of collecting specific past performance to job-related challenges.
4. Many interviewers fill silence that occurs during an interview by either describing what they are looking for (telegraphing the ideal answer) or moving on to the next question.
5. Field interviewers rarely seek confirmation, or even the means of confirmation, for the specific examples of past performance that they collect. But then, neither do the interviewers involved in the research studies. So while confirming behavioral answers could increase decision power, failing to seek confirmation can't add to gap between research findings and field practice.

Beyond these five limitations of interviewing practice, behavioral interviewing is normally applied to the 4-6 finalist candidates, relying on some combination of resume sorts, telephone screening interviews, or Boolean text searches to pick the 5-8 candidates who make it onto the short list from the 10 to as many as 400 that initially respond. These alternative screening methods have low decision power, resulting in high screening error

rates of two types. The false positives add labor cost to a selection strategy by lengthening the short list to 5-8 from the 2-3 required if there were an efficient way to accurately pick the performers from the pretenders at the top of the funnel. At least false positives can be caught in final decision behavioral interviews—if only they enjoyed the full decision power found in studies. The false negatives lower the average performance value of those hired by falsely ruling out superior talent early in the process.

## How big is the gap?

Hiring decision power is measured by the correlation between interview score and a measure of performance on the job. A correlation is an abstract statistical number, but it directly reflects the proportion of potential talent value that the selection strategy (the interview in this case) delivers to the job. So a value of .50 for a staffing strategy means that this strategy captures 50% of the potential talent value that would have been captured by a strategy that hired ONLY the very best talent from those that applied. We are dealing with predicting the job performance of people over a period of from under a year (in the case of telemarketers and restaurant staff) to over 20 years (in the case of workers and managers in your local power company). So when someone tells you their staffing strategy (interview or test) produces decision power (correlation) in the 80s, 90s or “is 100 percent accurate”—they are pulling your leg to put it kindly.

Selection scientists (Industrial Psychologists) have studied the hiring decision power of all the common hiring decision tools in independently reviewed research studies. Taken separately, tests of mental ability come out on top, averaging a decision power of .55 across hundreds of studies. Of course, some jobs require less cognitive skill than others, and the decision power for those jobs is much less—as low as the high teens. For other jobs that require a ‘big brain’ (such as systems analyst, electrical engineer, or rocket scientist) the decision power tops out around .70. At the other end of the scale, research into staffing strategies that involve graphology, numerology, and astrology consistently finds that these séance vs. science-based approaches deliver decision power near 0 (as should be expected).

When it comes to interviews, a lot depends on the type of interview. Traditional, unstructured interviews average a decision power (even after being corrected for downward research biases) of around .2. Unstructured panel interviews do better at .35. The good news is that behavioral interviews do almost as well as mental ability tests—averaging .53 across all the published studies. So when we ask: “How big is the gap?”, we mean the gap between the .53 found for research that studies the power of behavioral interviews and the level of actual decision power that employers receive when they use behavioral interviews to make hiring decisions in the field.

It would be nice if we could directly assess the decision power of interview-based staffing decisions made in the field, but that is precisely why there is a gap. It’s a lot like the uncertainty principle in quantum physics, where the very act of measuring a particle causes it to change. Instead, I will draw on experience and related research to estimate the gap. Initially, I will focus on the first two causes for the gap listed above—[1] poor note

taking and [2] the failure to rate each answer and combine the ratings to evaluate the candidate. Then I turn attention to the lack of confirmation details and the placement of the interview at the bottom of the funnel.

Early research on the interview by Ed Webster found interviewers to be in a hurry to remove doubt from the process and reach quick decisions about candidates. So they don't need to take notes or evaluate every question, because they have already made up their minds. Careful interviewers following their behavioral interview training ask questions from their interview guides, probe to make sure they acquire a specific, behavior description (or intention), take readable notes, and rate each answer against the behavioral anchors provided in the guide. When the doors close in the field, interviewers often do the minimum--- ask a couple of questions from the guide. When they revert to the Ed Webster practices (take poor if any notes and evaluate the candidate on one scale), they damage the decision power potential of the behavioral interview by from 10-20 points. Thus .53 becomes .33 - .43, in my view.

Turning to the fifth item on the poor practices list above, how much additional decision power could be gained if there was a way to confirm each of the candidate's answers? Behavioral interviewing practice in the currently reported research does not collect such confirmations. While confirming each answer presents practical problems, the benefits of having independent answer confirmation are obvious. I estimate that it would add at least 5 - 7 points of decision power, raising the .53 to from .58 - .60. Thus a best practices field approach that combined great note taking, careful rating of each answer, and collecting confirmation for most answers, could easily move the needle on hiring decision power from the mid thirties up to the high 50s.

Finally, the gap between the potential **value** delivered by a staffing strategy is a function of decision power times funnel power. Funnel power is a direct function of the number of candidates evaluated at each decision point. If there is only one candidate per open position, funnel power = 0. Hiring better talent when you have more candidates to choose from makes intuitive sense. The funnel power formula merely translates that common sense into numbers. The relationship between funnel power and number of candidates per decision point is complex mathematically, but well known. Some common values appear in the following list:

Candidates/ Decision	Funnel Power
2	.81
3	1.09
5	1.41
10	1.75
20	2.06
50	2.89
100	3.5

Putting the behavioral interview down at the end of the hiring funnel, where it applies to just 3-5 short list finalists, reduces the funnel power and thus the value of the staffing strategy. So if there was some feasible way to move the greater decision power of a behavioral interview, that could further leverage the value of a rigorous behavioral interviewing staffing strategy by from 30% to 150%.

Summing up, “How big is the gap between what is now, and what could be, for behavioral interviewing in the field?” I see a 20-30 point gap in decision power and another 30 to 150% gap in staffing strategy funnel power. Those are pretty big gaps in my book.

### Does the Gap Matter to Business? How Much?

Only if money matters, and we all know the answer to that. To quantify HOW MUCH it matters, let’s examine four examples: [1] security guard, [2] collections agent, and [3] convenience store manager, and [4] department store manager. Another article in this series explores the science and the math behind the numbers that follow in the table below.

<b>Position</b>	<b>Top of Funnel</b>	<b>Salary</b>	<b>Average Tenure</b>	<b>Per Hire Gap Impact in \$</b>
Security Guard	3	\$ 15,000	1.5	\$ 1,850
Collections Agent	8	\$ 24,000	.75	\$ 3,100
Convenience Store Mgr	20	\$ 50,000	5	\$ 38,000
Department Store Mgr	150	\$70,000	16	\$ 214,000

The *Top of Funnel* column tables the number of people who respond with interest by: walking in, applying online at the corporate career site, calling in to an 800 number, faxing in their resume, or following up on an internal referral. *Per Hire Gap Impact in \$* column tables the increase in the performance value per hire that could be achieved for that position if behavioral interviewing were practiced to its full potential vs. practiced the way it normally plays out in field settings.

The gap impact in dollars varies so dramatically between security guard and department store manager for reasons that make good sense based on the well established (and mathematically proven) utility formula. The dollar impact of hiring terrific vs. terrible security guards is not so high where the financial consequences of hiring a top vs. terrible department store manager could ruin your whole annual revenue forecast for that store. While hiring aggressive vs. mild mannered collections agents has considerably greater dollar impact than hiring mistakes for security guards, their shorter tenure cuts into the gap impact for them. For the convenience store chain, the store manager impacts not only the hiring of store staff, but also is empowered to adjust the store inventory to meet local retail conditions. Not only do convenience store managers have greater performance dollar impact and greater tenure, there are also many more candidates to choose from, and they stay at least 5 times as long as the first two positions in the table—thus the 10X increase in gap impact. The annual store revenue for the convenience stores runs around \$1M but it is many times that for department store managers. On top of that, a strong

employer brand attracts even more candidates and they stay an amazing 16 years with the company. Financially, closing the gap for department store managers makes a lot of sense and quite a few dollars per hire—even though the department store chain collects those dollars over a 16 year period.

## How can we Close the Gap?

There is a substantial gap between the promise and the performance of behavioral interviewing in supporting hiring decisions in the field. There are miles to go before we can close the gap and take a well earned rest. As usual, there is good news and there is bad news on the topic of closing the gap. Starting with the bad news first, strong arm tactics that could include requiring all post-eligibility screening decisions to be carried out via behavioral interviews, surveying candidates and ‘secret shopper’ candidates-- followed up by enforcement from the hiring police could do the job, but would cost a fortune and deliver as much resentment as results.

Now for the good news: internet delivered intelligent software offers better answers. See the link to **① [Accelerated Interviewing™](#)** to discover how these shortcomings have been overcome.