

CRITERION-RELATED VALIDATION THROUGH BENCHMARKING
An Introduction to Benchmarking and a Case Study
Sales Position for Light Manufacturing Company
2005

Benchmarking is the process in which the competencies necessary for job success are determined by comparing the results of individual assessments of a sample of successful incumbents on a specific job with those from samples of both average and poor or struggling performers on that job. The differences in the patterns of results among the three groups can then be used as a benchmark against which to compare the characteristics of applicants for that job on a variety of tests—a criterion-related validation process.

Benchmarking can help an organization to better understand the requirements that make for job success. Moreover, it is arguably the most common method of establishing the validity of an assessment process. Despite not being generally understood as such, benchmarking clearly identifies those characteristics associated with success on a particular job and thus is criterion related.

Since benchmarking compares the results obtained from a test or a test battery with the current levels of performance of job incumbents, benchmarking thus is a form of concurrent validity. While it is not as strict a test of the predictive value of the test, most test users understand the need for some measure of validity in their organization and benchmarking is realistically seen as a suitable alternative. And since it does have criterion relatedness, it does provide an essential determinant of whether or not using the testing instrument or test battery is legitimate.

Conducting a benchmark

Benchmarking is a particularly useful tool in establishing both the personal/inter-personal characteristics and the level of cognitive ability associated with success on a specific job. Using benchmarks to establish which characteristics make for success provides objective data, eliminating the guesswork and speculation about what is necessary for success on the job.

While a job analysis plays an important role in determining the competencies required for job success, it is often difficult to translate the results of a job analysis into the psychological test results that are hallmarks of success. This is especially true of personal/interpersonal characteristics and cognitive

ability. Most line managers who are usually involved in a job analysis rarely possess the level of expertise to make such determinations. Thus identifying these characteristics through testing is necessary to select those who will succeed on the job. As a result, in order to make certain such factors are considered, we recommend including the cognitive and personality measures from the Reliant Clues™ assessment in all benchmarking studies. Even at the entry level, whenever the job is critical to the organization, such tests are necessary and should be included.

Developing the benchmark for a specific job involves the following specific steps:

(1) Conduct a job analysis by asking a panel of persons knowledgeable about a job to identify independently the knowledge, skills, and abilities each of them assumes to be important for success on the job. This group may involve incumbents, first-line supervisors, managers, and any others with knowledge of that job. This information can be collected through interviews, questionnaires, or e-mail.

(2) Analyze the content of the job analysis to identify the level of knowledge, skills, and abilities required. This analysis should identify which of these competencies are most important for success on this job, providing a foundation for understanding the characteristics associated with job success.

(3) Identify the measures, including tests, which can best measure the most critical competencies.

(4) Identify three groups of incumbents. One group should be composed of individuals who are top performers on the job, a second composed of average performers, and a third group of poor or struggling performers. It should be noted that, in some organizations, there will be resistance to identifying anybody as a weak or poor performer. We have used the device of simply identifying the three groups by number or letter to obfuscate the groupings. How these groups have been composed, of course, should not be made public, and especially to those involved. It is important that, in order to avoid bias and favoritism, job success be clearly defined by objective performance criteria, such as dollar value of sales, amount of rework, absenteeism, disciplinary actions, and the like. If such performance measures already exist in personnel records, it is preferable that they be used. Using such existing records not only saves time and effort, but also these measures usually have a high degree of acceptance with the organization's management.

(5) Administer the selected tests to all the individuals in the three groups, ensuring that each distinctive, important competency or characteristic identified by the job analysis is tapped. Those tested should simply be told that the testing is part of an effort by management to better understand the factors involved in success on that job—which, in fact, it is. As we noted above, some measure of job-related personality characteristics and a measure of cognitive ability should always be included among these measures. In many situations, these two measures, such as those included in Reliant Clues™, will constitute a sufficient source of data for the benchmarking.

(6) Examine the three sets of data—one from the top performers, a second from the average group, and the third from the marginal or poor performers—to determine which of the test scores most clearly differentiate the most successful incumbents from the least. Because this sixth step is so

important for understanding the benchmarking process, the procedures involved in comparing the three groups will now be discussed in some detail. The data from the three groups must be entered into a spreadsheet, such as the Excel document shown in the table to the right. In addition to several demographic variables, each of the Reliant™ cognitive and personality scale scores are included in deciles (tenths). It can be noted that, in this actual case

Perf	Yrs												% Fit to							
		Rating	W	Co	Age	Education	Race	Gender	Cog	Org	Asse	Rules	Ext	Stab	Team	Good	I	Benchmark	Cog Flag	Rules Flag
A	15	45	4-year College	White	Female	8	2	8	2	8	4	5	3	100%						
A	10	45	4-year College	White	Male	9	4	7	5	7	2	5	2	100%						
A	21	41	4-year College	White	Male	9	1	10	2	2	2	1	1	93%						
A	10	45	4-year College	White	Male	9	1	10	5	1	3	1	1	93%						
A	10	55	4-year College	White	Male	7	1	10	4	1	6	6	1	93%						
A	15	55	4-year College	White	Male	10	4	10	6	9	4	3	2	93%						
A	10	35	4-year College	White	Male	10	1	8	5	3	4	1	3	93%						
A	2	28	More	White	Male	9	5	7	6	9	5	8	2	86%						
A	9	33	4-year College	White	Male	9	4	8	6	7	10	6	1	86%						
A	1	25	4-year College	White	Female	7	1	10	1	3	1	1	1	79%						
A	3	47	4-year College	White	Male	6	3	6	3	7	10	1	5	79%						
A	20	50	4-year College	White	Male	9	1	10	10	8	1	4	7	84%					High Rules	
A	5	47	High School	White	Female	3	2	8	2	5	2	9	2	84%				Low Cog		
A	3	47	4-year College	White	Female	8	5	4	10	1	6	3	5	57%					High Rules	
A	15	33	4-year College	White	Male	4	4	4	6	7	8	5	6	57%				Low Cog		
B	15	60	2-year College	White	Female	9	4	10	3	6	2	2	1	100%						
B	2	45	2-year College	Asian	Male	9	1	10	4	2	6	9	1	86%						
B	2	35	4-year College	White	Male	7	2	6	1	7	3	1	2	86%						
B	3	39	2-year College	White	Female	6	3	8	1	9	4	1	4	79%						
B	2	35	4-year College	White	Female	6	1	6	6	2	1	3	4	71%						
B	8	36	2-year College	White	Male	1	4	9	2	7	6	2	1	71%				Low Cog		
B	8	41	4-year College	White	Male	4	4	9	5	6	4	6	3	71%				Low Cog		
B	8	39	2-year College	White	Female	7	6	6	1	5	10	2	3	71%						
B	10	60	4-year College	White	Male	3	2	10	6	5	6	4	2	71%				Low Cog		
B	10	44	4-year College	White	Male	7	5	8	8	3	4	3	1	64%					High Rules	
B	15	63	2-year College	White	Male	6	2	10	9	5	6	1	3	64%					High Rules	
B	10	42	4-year College	White	Male	6	2	7	2	4	3	4	1	64%				Low Cog		
B	9	46	High School	White	Female	4	1	8	5	2	3	7	1	57%				Low Cog		
B	8	33	High School	White	Female	6	7	7	1	9	9	8	2	57%						
B	12	58	4-year College	White	Male	7	1	9	10	1	1	7	2	50%					High Rules	
B	30	64	2-year College	White	Male	7	10	1	4	10	8	3	3	50%						Low Assert
B	4	40	4-year College	White	Male	4	10	5	4	10	8	4	9	43%				Low Cog		
B	2	35	2-year College	White	Male	6	1	10	1	3	1	10	1	43%				Low Cog		
B	10	37	2-year College	White	Female	4	2	1	5	3	3	6	9	36%				Low Cog		Low Assert
B	8	41	4-year College	White	Female	6	2	9	10	3	1	2	5	29%				Low Cog	High Rules	
C	8	40	High School	White	Male	8	2	5	1	4	6	9	7	71%						
C	20	42	High School	White	Female	2	3	8	4	7	6	1	1	71%				Low Cog		
C	7	69	4-year College	White	Male	6	1	8	9	7	4	1	3	64%					High Rules	
C	2	48	4-year College	White	Male	9	2	5	8	1	6	1	8	57%					High Rules	
C	2	50	2-year College	White	Male	4	2	5	6	3	3	2	6	57%				Low Cog		
C	6	53	2-year College	White	Female	5	2	8	1	9	2	1	7	57%				Low Cog		
C	3	35	4-year College	White	Male	9	6	3	6	9	7	4	6	50%						Low Assert
C	10	49	High School	White	Female	4	3	10	10	5	1	4	5	36%					High Rules	
C	20	52	4-year College	White	Male	6	5	10	10	4	1	1	3	29%				Low Cog	High Rules	
C	15	54	4-year College	White	Female	4	3	1	6	9	1	7	6	21%				Low Cog		Low Assert

involving the sales force of a large light manufacturing company, there were 15 individuals in the top group, 20 in the middle group, and 10 in the struggling or bottom group. The different sizes of the groups were a function of the actual distribution of performance indicators found in the company's personnel records.

Demographic considerations

In any benchmarking study, the demographics of the three groups needs to be examined carefully to determine if any demographic factor is linked to job success. Since all of the employees employed in this sales job were Caucasians, race obviously could not be a factor in differential job performance. Turning to gender, Group A—the top group— contained four women (26.7%), Group B—the middle group—had six women (30.0%), and Group C—the bottom group—had four women (40%). The differences among these three groups were not significant and, in any event, given the small numbers would be difficult to interpret as determinants of job success.

The mean age of Group A was 42.1, of Group B 44.2, and of Group C 48.2. The differences in mean ages between the groups were not statistically significant. The mean age of Group C was skewed by one 69-year old employee; excluding this single case from consideration reduces the mean age of Group C to 45.9. Thus it would not appear that age was a significant determinant of job performance.

The mean length of employment of these employees at this company was also included in the data set in order to determine if length of employment played an important role in job success. The mean length of employment for Group A was 9.9 years, for Group B was 8.7 years, and for Group C was 9.3 years. The differences among the years of employment between these three groups were not statistically significant and it would appear that differential experience with the company also was not a significant determinant of job success.

Thus we can conclude that none of the differences in job performance could readily be explained in terms of demographic differences between the three groups. Our attention then needs to be turned to the psychometric data from Reliant Clues™.

Reliant Clues™ data

Each of the Reliant Clues™ scores, except for the Good Impression score, was given unit weighting (that is, each of the scales was given the same weight) in a goodness of fit design to determine if job performance could be identified by

the Reliant Clues™ scale scores. Initially Group A was compared by inspection with Group C for differences on each of the Reliant Clues™ scales. The table quickly reveals that Group A had many more high scores on the Cognitive Ability Scale than Group C. The Cognitive Ability scores for Group A ranged from 3 to 10 with a mean of 7.8 while the range for Group C was from 2 to 9 with a mean of 5.6, indicating that Cognitive Ability is an important determinant of success in this sales job.

In a similar fashion, we were able to determine that high scores on the Rules scale and low scores on the Assertiveness scale of Reliant Clues™ were related to poor job performance. A combination of using these three markers and a cutoff score of 72% made it possible to identify 100 percent of the poor performers and 73 percent of the top performers. Overall, the benchmarks differentiated the top and bottom performers with 84% accuracy. Clearly, these data provide strong support for the concurrent validity of Reliant Clues™ within this employment situation.

In summary, the above case study represents a recent application of this benchmarking methodology to the sales force within this customer. The largest difference between the two groups was on cognitive ability, with the most successful sales people having considerably higher scores. Up to this point the sales force had been selected on the basis of extroversion—the capacity to successfully engage others in social interactions and be resilient to rejection. While extroversion scores were high for both groups, there was only a slight difference in the means.

(7) The obtained pattern of differences revealed by the benchmark then can be used as a template for selecting new employees. Such an application would surely lead to better workforce performance and fewer poor performers on the job.

One question that is frequently asked is why three groups are necessary for benchmarking. The most important reason is that we sometimes have found that there are differences between the average and bottom groups, but not between those two groups and the top group. And, we have had instances where the opposite has occurred. Thus using three groups increases the probability of detecting differences that do exist at various levels of performance.

Additional reasons for benchmarking

This case example illustrates one of the important benefits of using benchmarking; namely, it provides management with information about the nature of a specific job and the competencies necessary to do that job successfully.

This is often new information and can be helpful to management in identifying reasons for generally poor performance; that is, limitations in their present workforce.

As we began this paper, we noted that benchmarking is a form of validation. Benchmarking is a criterion-based process for evaluating the validity of Reliant Clues™ in the user's workplace. And it should be noted that the results of benchmarking should provide the necessary evidence of the validity of the assessment process in the face of legal challenges, an important consideration in this litigious age.

Problems with benchmarking

The biggest challenge with benchmarking is obtaining agreement from current management as to which employees are top performers versus bottom performers. In many cases, the criteria used to determine performance are not consistent from one manager to the next. It is critical that agreement be reached as to the criteria for success and that the mechanism for ranking current employees against those criteria is consistent. Otherwise, it is difficult to conduct a proper benchmarking study.