

2006



Beware of “Barnum and Forer Effects” in Organizational Assessments

James Houran, President, 20 20 Skills™ Employee Assessment
Rense Lange, President, Integrated Knowledge Systems, Inc.
Gene A. Ference, President, HVS-The Ference Group

HVS INTERNATIONAL NEW YORK

372 Willis Ave.
Mineola, NY 11501
+1 516.248.8828 (ph)
+1 516.742.3059 (fax)

September 2006

New York San Francisco Boulder Denver Miami Dallas Chicago Washington, D.C. Weston, CT Phoenix Mt. Lakes, NJ
Vancouver Toronto London Madrid New Delhi Singapore Hong Kong Sydney São Paulo Buenos Aires Newport, RI

Abstract. – Feedback from organizational assessments of poor quality can appear specific, meaningful and valid due to psychological factors called Barnum and Forer Effects. These are statements that are so general or vague that they can apply to many people, and hence, are often rated as highly accurate by individuals. Even assessments with high psychometric quality may provide overly vague feedback or “cookie cutter” statements that are essentially useless. This article will help professionals evaluate their current organizational assessments for the presence of these shortcomings and to better understand the benefits of specific, actionable and evidence-based feedback that is provided by Item Response Theory (IRT) testing methods.

Although organizational surveys and assessments can be important components of Best Practice recruitment and training processes, many executives and human resource professionals avoid using them. Often there are understandable reasons for this, including lack of awareness of suitable instruments, perceived time constraints in completing quality research and cost-to-benefit issues. And then there is outright skepticism about the accuracy and usefulness of assessments in the first place.

What is not common knowledge is that professionals may still miss the mark even when these issues are resolved. The psychometric quality of an assessment – that is, its reliability and validity – may be excellent, this does not guarantee that the feedback generated by that assessment is also high in quality. In the service-hospitality industry and beyond we frequently hear criticisms that the feedback from surveys and assessments is often too general or “cookie cutter” to be useful and actionable. That is a reasonable complaint about some assessments. What makes matters worse is that illegitimate or useless feedback can appear specific, meaningful and legitimate. This is explained by what social scientists call *Barnum* and *Forer Effects*. Unfortunately, professionals sometimes do not realize that the assessment on which they depend for recruitment and employee training suffers from these limitations.

Barnum and Forer Effects – or “One Size Fits All”

The *Barnum Effect* is the name given to a type of subjective validation in which a person finds personal meaning in statements that could actually apply to many people. Psychologist Paul Meehl is credited with coining the expression, which apparently is in deference to circus man P. T. Barnum’s reputation as a master psychological manipulator who often claimed that “we have something for everybody” and “there’s a sucker born every minute.” It is not difficult to see why assessments with illegitimate or useless feedback might be perceived as valid instruments. Specifically, if Barnum statements

appear on a feedback report that a person believes has been specially prepared for him or her based on a realistic looking assessment, recipients often agree with such statements thereby giving validity to the assessment itself.

Related to the Barnum effect is the *Forer Effect*. Psychologist Bertram R. Forer found that people tend to accept vague or overly general personality descriptions as uniquely applicable to them, without realizing that the same description could be applicable to nearly everyone. Thus, the Forer Effect refers to the tendency for people to rate sets of statements as *highly accurate* for them personally even though the statements could apply to many people. The difference between the Barnum and Forer Effects is that the former describes a vague statement, whereas the latter describes how people react psychologically to a Barnum (or vague) statements.

In his now classic 1940s study, Forer² administered a “personality test” to his students, ignored their answers, and gave *each* student the (same) above profile that was borrowed from a newsstand astrology column. He then asked these students to assess the accuracy of “their” profile on a scale from 0 to 5, with “5” meaning “excellent”, “4” meaning “good,” and so on. The class average evaluation was a striking 4.26. Forer’s classic experiment has been replicated hundreds of time with psychology students and the average is still around 4.2 out of 5, or 84% accurate^{1,6}.

Personnel managers are also known to be susceptible to Barnum and Forer effects, even though these professionals should recognize these effects by training⁷. This explains why organizational assessments may be seen as highly accurate and contributing to a company’s bottom line when in reality that feedback might be illegitimate or useless.

As an example, consider the personality evaluation shown on the next page as if it was given to you as part of the recruitment process at an organization —

Bookman's Appreciative Inquiry Index (AII)TM

CONFIDENTIAL and In-Depth Analysis for: Janet

Assessment completed by:



James Houran, Ph.D.,
Bookman Institute of Recruitment and Executive Training

Date: August 21, 2006

The following profile was derived from comparative determinations based on the responses to the AII questionnaire analyzed by the above test administrator using empirical descriptions and anecdotal evidence. This report is a guide that was designed to assist in the employee screening, selection and training process. The report should not be used in isolation but always in conjunction with both an interview and a process whereby a person's experience, education, qualifications, competence and trainability can be assessed.

Type: Strong Perceiver/ Deep Feeler

Profile: You have a strong need for other people to like you and for them to admire you. You have a tendency to be critical of yourself. You have a great deal of unused capacity which you have not turned to your advantage. While you have some personality weaknesses, you are generally able to compensate for them. Disciplined and controlled on the outside, you tend to be worrisome and insecure inside. At times you have serious doubts as to whether you have made the right decision or done the right thing. You prefer a certain amount of change and variety and become dissatisfied when hemmed in by restrictions and limitations. You pride yourself as being an independent thinker and do not accept others' opinions without satisfactory proof. You have found it unwise to be too frank in revealing yourself to others. At times you are extroverted, affable, sociable while at other times you are introverted, wary and reserved. Some of your aspirations tend to be pretty unrealistic.

Note: This profile is a guide that was designed to assist in the HR screening, selection and training process. The report should not be used in isolation but always in conjunction with both an interview and a process whereby a person's experience, education, qualifications, competence and trainability can be assessed.

Forer convinced many people he could accurately assess their character traits with a profile like this – even though the personality profile he gave was not derived from a legitimate personality assessment but rather was borrowed from a newsstand astrology column and presented to his students without regard to their sun sign. Likewise, many online and offline assessments on the market today are not guaranteed to meet professional testing standards grounded in modern testing theory⁴, yet almost all assessments have a host of satisfied customers who are convinced they are legitimate and accurate. As for the validity of “Bookman’s Appreciative Inquiry Index (AII)TM,” please see an important note at the end of this article!

Why We Fall Prey to Barnum and Forer Effects

The most common explanations given to account for the Forer effect are in terms of hope, wishful thinking, vanity and the tendency to try to make sense out of experience. Forer’s own explanation was human gullibility². Likely, there is a little bit of truth in all of these explanations. People tend to accept claims about themselves in proportion to their *desire* that these be true rather than in proportion to their empirical accuracy. We tend to accept questionable – even false statements about ourselves – if we deem those statements positive or flattering enough. We will even give extremely liberal interpretations to vague or inconsistent claims about ourselves in order to help us make sense out of them.

Human beings experience apprehension and anxiety when faced with ambiguity and uncertainty; it is a common and natural reaction given that our brains are hardwired to make sense of the world around us and the information we collect. Therefore, people often psychologically “fill in the blanks” and provide a coherent picture of what it is seen, heard and otherwise perceived, even though a careful examination of the evidence would reveal the data to be vague, confusing, obscure, inconsistent and even unintelligible. Consistent with these ideas, mathematical models of survey data suggest that our belief systems help us find meaning in chaos, thereby coping intellectually and emotionally with ambiguity and uncertainty^{3,5,6}.

The fake personality profile shown above nicely illustrates how a candidate or hiring professional can readily accept the validity of an illegitimate personality report on face value given the contextual pressure to believe it – for example, the academic sounding jargon and an intimidating test name, imposing trade marks, words like “confidential” and “in-depth,” a name and signature of a perceived authority figure and the clinical and professional look of the feedback itself. Plus, favorable assessments are “more readily accepted as accurate descriptions of subjects’ personalities than unfavorable” ones. But, unfavorable claims are “more readily accepted when delivered by people with high perceived status than low perceived status.”

There have been numerous studies conducted on Barnum and Forer effects. Dickson and Kelly¹ examined much of this research and concluded that overall there is significant support for the general claim that Forer profiles are generally perceived to be accurate by

participants in the studies. Furthermore, there is an increased acceptance of the profile if it is labeled “for you,” while personality variables such as *neuroticism*, *need for approval*, and *authoritarianism* are positively related to belief in Forer-like profiles^{1,8}.

Fortunately, people can generally distinguish between statements that are accurate (but would be so for large numbers of people) and those that are unique (accurate for them - but not applicable to most people) when they are provided with the appropriate information and education. In other words, professionals must carefully consider the feedback from assessments and keep that information in proper perspective.

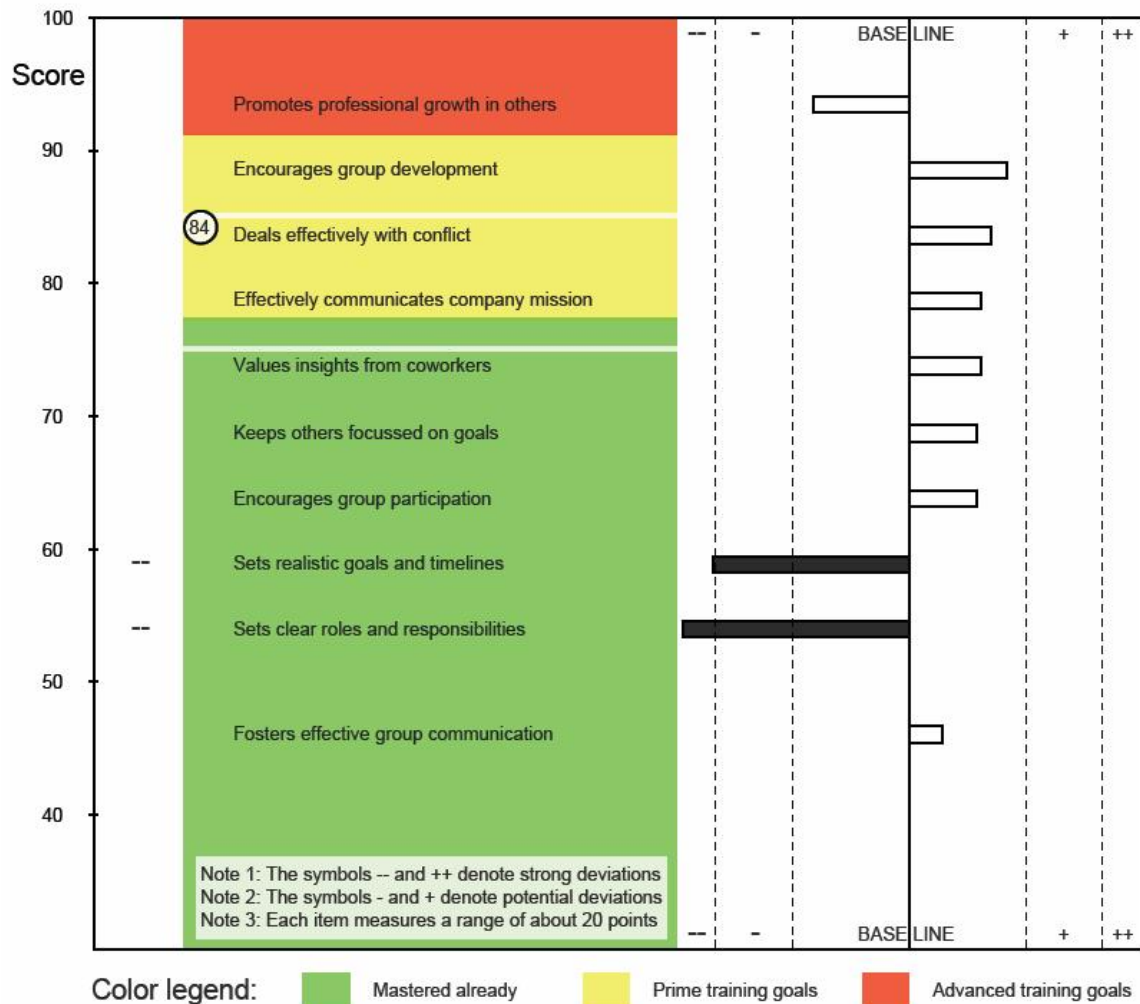
Moving on from Poor Assessment and Feedback

The truth is that no organizational assessment or survey is perfect; even those of the highest psychometric quality can only produce feedback that is based on mathematical extrapolations (as reliable as these may be). Excellent assessments will occasionally be “off the mark,” but the idea is to generate quality feedback that is as specific as possible and is “on the mark” more often than not.

We have found that an *Item Response Theory* (IRT) approach to tests and measurements is the best practice solution towards achieving this level of reliability, especially when the results of the assessments have real-world implications⁴. However, hiring professionals – not psychological instruments – have the ultimate responsibility for comprehensively evaluating a candidate or employee. Assessment feedback should only be one component of a broader process to ensure proper due diligence on candidates during the screening, selection and training process. Accordingly, we believe that in addition to ranking test takers, assessment systems should assist HR professionals by pointing to the right questions to be addressed in live behavioral interviews and follow-ups.

Figure 1.

As the following Action-Map(TM) illustrates, Group Process refers to a set of attitudes and behaviors that foster a collaborative work environment characterized by high morale and employee commitment to tasks and goals. Relative to most test takers, Janet's score of 84 is considered to be MODERATELY HIGH.



As an example of such assessment systems, we note that the 20 | 20 Skills™ assessment from HVS provides hiring professionals with complete information concerning all reporting factors being measured by this instrument. As can be seen in Figure 1 for the “Group Process/ Team Building” subscale, the 20 | 20 Skills™ assessment revolves around of a series of Action-Maps™ that contain the type of information referred to above.

The example shows the actual data of an anonymous respondent (dubbed “Janet”) who obtained a score of 84 out of 100 on this competency. Please note the following main features:

- The Action-Map™ plots the skills that define “Group Process” at the location (Y-axis) corresponding to the score needed to possess the skill listed in the boxes (or, in other cases, perform a particular action, or solve a particular problem). A scale of 0 to 100 is used. Based on extensive testing, scores of 85 or higher are regarded as “high,” scores from 75 to 84 are “moderately high” and scores below 75 are “somewhat low.”
- The test takers’ scores are computed so as to use the exact same scale, and thus the map also gives an unambiguous interpretation of all test scores. In the example, Janet’s score of 84 exceeds the location of all statements in the green section – i.e., “Values insights from coworkers,” “Keeps others focused on work,” “Fosters effective group communication.” This indicates that Janet almost certainly masters all of these aspects of Group Process. By contrast, Janet’s score falls below the location of “Promotes professional growth in others” (red section), reflecting that Janet almost certainly lacks this particular skill. Finally, three statements are listed in the yellow section, indicating that we cannot say with sufficient certainty whether Janet masters or does not master these particular aspects of Group Process.
- The above implies that Group Process is characterized by a quantitatively ordered hierarchy of skills such that test takers’ scores directly indicate which skills they likely possess or not possess. Individual variations may exist, however, and this is indicated by the bar graph in the right side panel of Figure 1. This graph plots the (standardized) deviations from what one might expect for someone with a particular score. For instance – as is indicated by the black negative bars – given her score of 84, Janet does not “Set realistic goals and timelines,” nor does she “Set clear roles and responsibilities.” The smaller deviations indicated by the white bars can be ignored.

From the information summarized above managers and hiring professionals can deduce training information (e.g., it is less necessary to train Janet on the skills in the green box than in the yellow box), as well as specific and actionable information that can be utilized effectively during behavioral interviews. For instance, if Janet was a new applicant, one might want to find out what her issues (if any) are with “goals and timelines.” A number of possibilities exist. For instance, Janet might have misunderstood the question, or she might have interpreted the question in light of counterproductive practices in her current place of employment. Similarly, it seems advisable to address issues concerning roles and responsibilities with Janet during an interview or follow-up.

The 20 | 20 Skills™ approach of providing overall scores for competencies as well as detailed information in Action-Maps™ guarantees that feedback is customized, specific

and actionable for each applicant or incumbent. However, in contrast to generic Barnum-type statements, the 20 | 20 Skills™ feedback reflects actual empirical information as obtained via advanced IRT methods.

Our reliance on IRT and related methodologies extends to other areas – including survey research, marketing research, cost-benefit analyses, etc. – because this approach is Best Practice for research in the tests and measurement field. And it is an approach that yields considerable insights and practical outcomes when it is combined with the expertise of managers and human resources professionals. A customized and collaborative initiative is the key to successful implementation.

So, it seems that our fake “Bookman’s Appreciative Inquiry Index” was legitimately on the mark about one thing, which is mentioned in its footer. That is, standardized assessments should never be used in isolation but always in conjunction with both a behavioral interview and a process whereby a person’s experience, education, qualifications, competence and trainability can be assessed.

Note. Astute readers will notice that the fake “Bookman’s Appreciative Inquiry Index” profile was named in deference to the font that was used to create it. Special appreciation goes to Mark Keith (HVS Executive Search – Hong Kong) for this. Similarly, the Bookman Institute of Recruitment and Executive Training is a fictitious affiliation!

About the Authors

James Houran holds a Ph.D. in Psychology and recently joined HVS to head the 20 | 20 Skills™ assessment business. He is a 15-year veteran in research and assessment on peak performance and experiences, with a special focus on online testing. His award-winning work has been profiled by a myriad of media outlets and programs including the Discovery Channel, A&E, BBC, NBC’s Today show, Wilson Quarterly, USA Today, New Scientist, Psychology Today, and Rolling Stone.

Rense Lange holds a Ph.D. in Psychology and a Masters’ in Computer Science. He is one the world’s foremost expert in tests and measurement and applied Item Response Theory and Rasch scaling, and Computer Adaptive Testing (CAT) in particular. In addition to serving on the faculty of the University of Illinois, the Southern Illinois University School of Medicine, and Central Michigan University, Rense has worked for ten years as the lead psychometrician at the Illinois State Board of Education and he is the Founder and President of Integrated Knowledge Systems, Inc.

Gene Ference holds a Ph.D. in Industrial-Organizational Psychology and is one of the highest respected Industrial Psychologists and Management and Organizational Development Specialists in the industry with over 35 years of experience in building peak-performing cultures and developing brand engagement strategies. His work has directly assisted clients in successful applications for the Malcolm Baldrige National Quality Award, Employee of Choice, Best Human Resources, Employer of the Year, and

Fortune 100 Best Companies to Work For, as well as the quality of work life and service culture awards of Great Britain, Brazil, The Netherlands, Mexico, Australia, and Singapore.

For information on the Best Practice 20 | 20 Skills™ assessment system, contact:

James Houran, Ph.D.
jhouran@2020skills.com
516.248.8828 x 264

For information on Best Practice Organizational Assessments & Professional Coaching and Training Workshops, contact:

Gene A. Ference, Ph.D.
gference@hvsinternational.com
203.266.6000

References

- ¹Dickson, D. H., & Kelly, I. W. (1985). The 'Barnum effect' in personality assessment: a review of the literature. *Psychological Reports*, 57, 367-382.
- ²Forer, B. R. (1949). The fallacy of personal validation: a classroom demonstration of gullibility. *Journal of Abnormal Psychology*, 44, 118-121.
- ³Houran, J., & Lange, R. (2004). Redefining delusion based on studies of subjective paranormal ideation. *Psychological Reports*, 94, 501-513.
- ⁴Houran, J., & Lange, R. (2006). State-of-the-art measurement in human resource assessment. *HVS Journal*. 28th Annual NYU Hospitality Industry Investment Conference, New York, NY, June 4-6.
- ⁵Lange, R., & Houran, J. (1998). Delusions of the paranormal: a haunting question of perception. *Journal of Nervous and Mental Disease*, 186, 637-645.
- ⁶Lange, R., & Houran, J. (1999). The role of fear in delusions of the paranormal. *Journal of Nervous and Mental Disease*, 187, 159-166.
- ⁷Stagner, R. (1958). The gullibility of personnel managers. *Personnel Psychology*, 11, 347-352.
- ⁸Thiriart, P. (1991). Acceptance of personality test results. *Skeptical Inquirer*, 15, 116-165.