

Occupational Managers vs. Professional Assessors

by

Richard C. Joines, President

June 2004

Management & Personnel Systems, Inc.

Walnut Creek, CA 94595

1-925-932-0203

Occupational Managers vs. Professional Assessors

Richard C. Joines, President
Management & Personnel Systems, Inc.

The Issue

Does the professional testing and assessment literature support the contention that assessors must be managers in the same occupational area as the assessees?

To use a few examples: Must a police sergeant or lieutenant assessment center use police managerial personnel as the assessors? Must a fire captain or battalion chief assessment center use fire managerial personnel as the assessors? In the government sector, the most common application of the assessment center method is for police and fire; thus, this paper will focus on police and fire assessment centers in dealing with the issue.

Summary

There are no studies that have found any support for requiring that assessors be drawn from the same occupational area as candidates. Two studies, each over 30 years old, found support for using occupational managers as assessors because they did not perform any worse than professional assessors (Greenwood & McNamara, 1969; and Thomson, 1970).¹

These two studies formed the basis for an evolving idea about the use of occupational assessors among assessment consultants. Early on, assessment consultants used these studies to defend the use of occupational managers rather than professional assessors. As time passed, it was common for assessment consultants to use occupational managers as assessors without being challenged. As more time passed, it became common for assessment consultants to claim that the assessors had to be occupational managers in the same area as the candidates. In 2004, we encountered a major jurisdiction that had been advised by a consultant to adopt rules stating that police assessment centers could use only police managerial personnel as the assessors.

The two aforementioned studies are not supported by other studies on this issue. A 1987 meta-analysis study reviewed all studies on this issue, and found that professional assessors produced better, more valid ratings. The authors were very emphatic in stating their findings. There is nothing in the professional literature or in professional guidelines or standards that requires the use of police or fire managers as assessors in police or fire assessment centers. Consultants who claim otherwise are not correct, and are apparently clinging to an outmoded assessment model.

Background

Over the past 30 years or so, many assessment consultants have promoted the idea that assessment centers should use assessors drawn from the same occupational area as the candidates being assessed. Some of these consultants have elevated their claim to say that the assessors must be drawn from the same occupational area. In the professional literature, such

assessors are referred to as occupational managers, e.g., a police captain serving as an assessor for a police lieutenant assessment center would be an "occupational manager."

In the early days of assessment, all assessors were professional assessors -- either psychologists or other personnel extensively trained in assessment. As the assessment center method expanded in usage in the late 1960's, there weren't enough professional assessors readily available to conduct all of the desired assessments. If an organization could use its own managerial personnel as assessors, there would be no impediment to the use of assessment centers for selection or development. Thus, the question of whether an organization's managerial personnel could be used was important to assessment center practitioners.

When the two aforementioned studies found that assessment ratings made by occupational managers were not significantly different than those made by professional assessors, the move to train and use an organization's managers as assessors was on. These two studies fueled the growth of assessment centers by validating a model in which a consultant would train managers as assessors, then oversee their performance in evaluating candidates in the assessment center.

In the 1970's and 1980's, it was common practice for consultants, including myself, to assure clients that occupational managers could serve as the assessors. We sometimes added that the assessors needed to be at the same level as the position being filled, and preferably would be one or even two levels higher. This recommendation was an outgrowth of a suggestion found in the 1970 study by Thomson.

From 1977 to 1987, I conducted numerous assessor training programs for occupational managers drawn from all segments of state and local governments. By 1984 or 1985, it had become apparent to me that occupational managers had great difficulty in scoring more complex assessment exercises, especially the in-basket. After training about 50 school administrators in a four day assessor training program and overseeing them evaluate candidates in a series of 11 managerial assessment centers, I concluded that four days of training just wasn't enough. In-baskets scored by the assessors in the morning hours would appear to me to have been scored differently than those scored in the afternoon. I found that the assessors would get "tougher" as the day wore on. When scoring group discussion exercises, I found that the assessors would give higher marks to the candidates who talked the most, even if I didn't think those candidates were particularly effective.

In 1987, a major meta-analysis study was completed by leaders in the assessment industry. A meta-analysis study is one that accumulates results across all available studies. These studies are designed to overcome the limitations of sample size and error that plague any single study. By using the meta-analysis technique, researchers can resolve contradictory findings that may have been reported in different studies; and in this way, find the truth.

The 1987 Gaugler, Rosenthal, Thornton and Bentson meta-analysis study looked at a number of hotly debated assessment center issues. At the time, the results of their study were anxiously awaited by those in the assessment profession. One of the researchers, Dr. George Thornton, was subsequently named one of the co-chairs of the task force that developed the 1989 Guidelines and Ethical Considerations for Assessment Center Operations.

The article, published as a Journal of Applied Psychology monograph, was entitled Meta-analysis of assessment center validity. The issue of occupational managers vs. professional assessors was dealt with as follows:

"In contrast to other researchers (Greenwood & McNamara, 1969; Thomson, 1970) who have found no difference in the assessment center ratings of professional (i.e., psychologists) and nonprofessional (i.e., in-house managers) assessors, we found evidence that assessment centers that use psychologists as assessors are significantly more valid than those that use managers as assessors. Many people in the field believe that managers are better able to interpret the meaning of different behaviors for a particular job than are psychologists, because they are more familiar with the requirements of the job. However, the results of this meta-analysis suggest that psychologists provide more valid assessment center ratings than do managers. In fact, this moderator is particularly robust given that it is negatively related to many other moderators (page 505)."

This strongly stated finding dismisses the two studies that found no difference between occupational managers and professional assessors. The implications of this meta-analysis study were clear: professional assessors not only could be used, but should be used because they produce more valid assessment center ratings.

The results of this study should have put an end to the contention of consultants that police and fire managers should (must) be used as the assessors in police and fire assessment centers. However, the assessment industry is apparently so entrenched in its model of using occupational managers (which are typically available free of charge) that assessment consultants continue telling prospective clients that occupational managers either should or must be used.

A Brief Look at the Study that Started the Controversy

Greenwood and McNamara (1969)¹ gave impetus to using occupational managers when their study found no difference in the reliability of ratings by occupational managers and professional assessors. Oddly enough, many of the reliability coefficients reported by the authors were quite low for both professional and non-professional assessors. The finding of low reliability for the ratings made by the experts included in the study is inconsistent with the implicit assumption that true experts in assessment will demonstrate high reliability (i.e., agreement). However, the authors do not address this issue.

The authors stated the following:

"In some instances the interrater reliability is alarmingly low (Table 5), but when one compares in a given situation the heterogeneous group (professional and nonprofessional) reliabilities, they generally appear equivalent. Also, neither homogeneous group consistently appears to have greater interrater reliability than the other homogeneous or heterogeneous groups."

In other words, the authors concluded that all groups were equally deficient; that the professional

assessors were just as bad as the occupational managers, and that therefore, it didn't matter if professional assessors or occupational managers were used.

In retrospect, this study is flawed in that there was nothing to establish that the professional assessors were highly trained professionals in assessment. If the study had started by establishing that professional assessors could achieve a respectable degree of reliability in their ratings, it would have been meaningful to determine if occupational managers with minimal training could achieve the same level of reliability in their ratings, but the study didn't do this.

An easier example to understand would be as follows: Suppose five medical doctors examine a patient and no two doctors agree on their diagnosis. A researcher comes along and randomly picks auto mechanics to examine the patient, and no two mechanics agree on their diagnosis. What do we think when the researcher concludes that since the level of agreement is poor in both groups, the hospital should feel free to hire the lower-paid blue-collar employees to diagnose patient illnesses (i.e., both groups have alarmingly low reliability, so it doesn't matter which group is used)?

Here's the catch: Experts in any scientific discipline are expected to demonstrate high agreement among themselves in their findings and conclusions. Where there is little or no agreement among a group of experts, it's time to check their credentials. In the Greenwood & McNamara study, the professional assessors didn't demonstrate that they were experts; thus, the occupational managers had a very low hurdle to cross and meeting this low standard should never have caused the authors to conclude that it didn't matter whether professional assessors or occupational managers were used. Instead, they should have said that the failure on the part of the professional assessors to demonstrate highly reliable results made a comparison with occupational managers meaningless. It may very well be that the professional assessors held the title of psychologist but were inexperienced in assessment center rating processes. The article does not provide detailed background information on the professional assessors so this issue can't be reviewed.

Professional Guidelines and/or Standards

In 1989 the Guidelines and Ethical Considerations for Assessment Center Operations were updated, and the standards for training assessors were increased (over the objections of many consultants) to suggest two days of training per one day of assessment (page 7):

"Precise standards for the minimum number of hours or days required for assessor training are difficult to specify. However, extensive experience has shown that, for the initial training of assessors who have no experience in an assessment center that conforms to the guidelines for this document, it is desirable to have at least two days of assessor training for each day of the administration of assessment center exercises."

These Guidelines do not state a requirement for a particular type of assessor, i.e., they do not suggest that occupational managers are required or even preferred. The only statement on the issue of assessor selection is: "A participant's supervisor should not assess him or her in an assessment center (page 3)."

A typical one day assessment center would consist of three or four exercises. Thus, the Guidelines were essentially approving a two day training program for an assessment process that included an in-basket, leaderless group discussion, and employee performance counseling. In our view, two days of training for three or four exercises wasn't nearly enough. We argued for more training prior to the 1989 meeting that adopted the new Guidelines, and we argued for more training during the meeting, but there wasn't enough hard research to persuade the conference participants to adopt a higher standard.

Indeed, many consultants objected to the language that was eventually adopted, arguing that since there was little research on the subject, there was no proof that two days of training was better than one-half day of training. These consultants were routinely conducting one day of training or less for their assessment centers, with the remainder of the week devoted to processing the candidates through the assessment center. Adding training time might have extended the process to the next week, requiring a weekend stay-over. It was more convenient and profitable to wrap up in five days, then move on to the next project or location.

At MPS, we concluded that it would take at least two days of training just to train occupational managers to evaluate candidates in a leaderless group discussion exercise. The leaderless group discussion is a moderately complex exercise. To accurately evaluate candidates in a leaderless group discussion, assessors need to understand group dynamics and be very well trained in observing the whole group. When consultants have only one day to train occupational managers in all of the exercises in the assessment center, they don't have time to train the assessors in group dynamics. Moreover, they typically limit the assessor's responsibilities by training them to observe only one or two of the six candidates in the group. As a result, the assessors fail to understand the group as a whole, and their ratings suffer. This leads to the syndrome of giving the highest ratings to those who talk the most because they are the ones noticed by all assessors.

At MPS, our experience made it perfectly clear that occupational managers could not be adequately trained to score a one day assessment center in two days, which was the suggested standard in the Guidelines, much less one day as practiced by many consultants. We recognized that it was difficult to obtain borrowed assessors for more than two days of training, but this didn't change our conclusion that two days was insufficient.

In 2000, the 1989 Guidelines and Ethical Considerations for Assessment Center Operations were revised. However, the new guidelines failed to address the issue of inadequate assessor training programs for occupational managers. They didn't increase the suggested amount of training contained in the 1989 Guidelines (two days of training for a one day assessment center). At MPS, we viewed this as unfortunate.

There is nothing in the 1989 or 2000 Guidelines that requires assessors to be drawn from the same occupational area as candidates, i.e., nothing in either set of Guidelines can be construed to require police managers to be the assessors in a police assessment center, or fire managers to be the assessors in a fire assessment center.

In fact, the 2000 Guidelines make it clear that assessment exercises need not be evaluated by human beings, that if a computer can do the job as well, then the computer may be used:

"Computer technology may be used to assess in those situations in which it can be shown that a computer program evaluates behaviors at least as well as a human assessor (page 4)."

The Guidelines recognize that the real issues in any testing program are reliability and validity. No group of professionals can alter professional testing standards, and these are the two key elements of professional standards. The assessment professionals who adopted the guidelines showed deference to these concepts:

"Whatever the approach to assessor training, the objective is to obtain reliable and accurate assessor judgments (page 6)."

If assessment consultants and practitioners would follow this guidance, they would pay attention to the research showing the problems occupational managers have in providing reliable and accurate ratings, especially for complex exercises that are typically included in assessment centers. The research has shown that professional assessors provide more reliable and valid ratings, yet the assessment industry chooses to look the other way.

The MPS Position/Approach

Shortly after the 1987 article by Gaugler, Rosenthal, Thornton and Bentson, we at MPS began practicing what the research literature showed to be the best approach. That is, we began rating all exercises using our own professional assessors. Our professional assessors are given extensive training. An assessor scoring our in-basket has completed 80 hours of training because this is what it really takes to achieve a high level of proficiency. Even with this amount of training, however, we find that our assessors still need to be able to turn to a more experienced assessor when new or unexpected situations arise.

Even though the research clearly supports the use of professional assessors, we continue getting reports of consultants telling prospective clients that the assessors **MUST** be occupational managers who are one or two levels above the position to be assessed. Apparently, most consultants do not want to do the actual ratings themselves, and they don't want to train and oversee a group of professional assessors. After all, assessors supplied by the client organization represent free labor. On the other hand, we have to pay our assessors, and because they are true professionals, we have to pay them well.

One argument that makes sense on the surface is that it takes police personnel to evaluate performance in a police tactical exercise, that it takes fire personnel to evaluate performance in a fire tactical exercise, etc. After all, how can non-police or non-fire personnel claim to be able to evaluate candidate performance in such exercises? At MPS, we have a simple answer: develop a detailed scoring plan beforehand, getting input from police or fire experts and establishing a fair system for scoring all candidates.

We originally researched this approach for the City of St. Louis Fire Department in the late 1980's and proved that two Human Resources staff could rate a fireground tactical exercise for Fire Captain just as reliably and validly as four fire personnel, provided there was agreement on

an objective scoring key. During the key development process, we discovered that supervisors of the position (i.e., Battalion Chiefs) often did not agree on the actions that should be taken in a specific situation. We had to resolve these disagreements, but in order to do so, we had to go as high as Assistant Chief and Deputy Chief. Only then could we get a final, "approved" scoring key.

When we reflected on this project, we realized that if we had used the traditional approach of relying on Fire Battalion Chief assessors from other jurisdictions, permitting them to rate the candidates without an objective key, they would have used different standards in rating the candidates. In other words, we reasoned, "Why would they agree on all issues when the Battalion Chiefs in St. Louis did not agree?"

Our approach for all technical/tactical exercises is to first devise a detailed, objective scoring key by getting input from top experts. When we use the exercise or any form of it in a new jurisdiction, we get detailed input from top management. When we are finished, we have a valid scoring key. Our professional assessors are trained in the application of these scoring keys. We understand them, and we apply them with high reliability. We have empirical studies proving that we can accurately rate these exercises. Consultants who wish to continue using assessors supplied by the jurisdiction (at no cost to the consultant) have no research or anything else to back up their arguments. All they can possibly rely on are two research studies conducted over 30 years ago which have since been rejected.

Conclusion

In summary, the common notion that occupational managers are required when conducting an assessment center is, pure and simple, **in error**. The published research does not support the idea that occupational managers are required in order to obtain reliable and valid ratings. There is no requirement in any professional guidelines or testing standards to use occupational managers as the assessors. Put simply, this is a myth that seems to have legs. Any consultant claiming that occupational managers are required, or that MPS assessors can't do the job, is selling, not reporting professional guidelines or the published research. We will gladly accept any offer to publicly debate the issue.

As a final word, we have served as assessors for approximately 45,000 managerial assessment exercises, of which approximately 20,000 are police and fire personnel, and no candidate has ever challenged our use of professional assessors. Further, no candidate has ever successfully challenged us or had any appeal upheld for any reason. We think our track record speaks for itself. If there were a problem with our approach of using professional assessors, we would have been successfully challenged a long time ago.

Notes

1. The Greenwood & McNamara study is routinely cited with a 1969 date even though the study was published in 1967. To avoid confusion and apparent inconsistencies in the present paper, we have also used the 1969 date. The incorrect date citation is evident as early as 1973 in Huck, James. R. Assessment centers: A review of the external and internal validities. *Personnel Psychology*, 1973, 26, 191-212. On page 206, the author cites Greenwood and McNamara (1969), but the reference section correctly cites the article as being published in 1967. This error has apparently perpetuated itself because Greenwood & McNamara wrote an article on a different subject that appeared in the same journal in 1969 (*Journal of Applied Psychology*).

References

Gaugler, Barbara B., Rosenthal, Douglas B., Thornton III, George C., and Bentson, Cynthia (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology Monograph*, 72, 493-511.

Greenwood, J.M., & McNamara, W.J. (1967). Interrater reliability in situational tests. *Journal of Applied Psychology*, 31, 101-106.

Guidelines and Ethical Considerations for Assessment Center Operations, Endorsed by the Seventeenth International Congress on the Assessment Center Method May 17, 1989, Pittsburgh, Pennsylvania.

Guidelines and Ethical Considerations for Assessment Center Operations, Endorsed by the 28th International Congress on the Assessment Center Method May 4, 2000, San Francisco, California.

Thomson, H.A. (1970). Comparison of predictor and criterion judgments of managerial performance using the multitrait-multimethod approach. *Journal of Applied Psychology*, 54, 496-502.