

Validity of a Standardized Reference Checklist

DeNell Daniel
Central Intelligence Agency

In this study, the reference checklist was investigated as a means for predicting the performance of college work study students. Though most research shows that typical recommendation letters correlate poorly with performance scores, this study shows that a standardized reference checklist can be an accurate predictor of performance ($r = .33$).

Within the personnel selection process, there are many ways to predict future performance of employees. Among these methods are psychological tests, assessment centers, and letters of recommendation or reference checks. Of these methods, psychological tests are considered to be the most valid predictors of job performance (Hunter & Hunter, 1984). Unfortunately, little is known about how well recommendation letters or reference checks will predict employee performance.

Use of Recommendations

Letters of recommendation and reference checks rely on the opinions of relevant others (past supervisors or employers) to help evaluate what an individual did and how well she/he performed in the past (Cascio, 1987). Despite their potential deficiencies, references and letters of recommendation are still extensively used by most organizations (Maier & Verser, 1982). There are basically two types of references: 1) the character reference, which is obtained from friends and/or relatives and 2) the previous employer reference, which is used to check previous performance.

According to LoPresto *et al.* (1985), there are two major reasons why an organization should use reference checking: 1) the reference itself is an indicator of future performance, and 2) the characteristics and integrity of

the individual (recommendee) may be seen. Muchinsky (1979) states that of the companies not using an employment recommendation questionnaire, 32 percent report using either letters of recommendation or telephone checks. Beason and Belt (1976) also report, from their survey of employers, that 82 percent of employers verified some information found on the application form through reference checks and 18 percent verified all information. Although there are many problems with and shortcomings of recommendation letters and reference checks, few companies are willing to abandon altogether the practice of using these forms of screening (Cascio, 1987).

Muchinsky (1986) agrees that letters of reference or recommendation should substantiate the information given by the recommendee. In addition, Muchinsky believes that the sole purpose of reference checks of any kind should only be to identify a small proportion of people who should not be given further consideration for a job due to past employment problems. Yoder (1962) agrees that a reference from a former employer should be discounted unless the work done on the old job is similar to the work done on the new job.

The comment by Muchinsky in some ways suggests that references should only be used when the reference method chosen minimizes any distortion of the individual (recommendee) and when the recommendee knows the individual writing the reference has confidence in the recommendee's integrity (McCormick & Ilgen, 1985). Browning (1968) stated that recommendations are used because they reflect how well an individual performed in the past and that past performance will be a factor in predicting future performance.

Problems with Reference Checks

There are many problems one must consider when recommendations are used to predict future performance. Therefore, one must be made aware of these problems and the research that is available when interpreting the recommendation.

Different Types of Raters

Those individuals asked to write recommendation letters know the recommendee from a particular setting (i.e. church or school); therefore the reference given is limited to that specific environment and is not

always an accurate predictor of how the recommendee may perform in an altogether different environment. In other words, the recommender is often without true knowledge of the recommendee's work habits and/or how s/he relates to others.

Ceci and Peters (1984) state that the failure of references to describe known shortcomings of the individual for whom they are writing the reference often results in people being hired or promoted who, more than likely, will be unsuccessful. The unsuccessfulness of these individuals could have been predicted earlier if the reference writers were more honest.

There is research both for and against recommendation letters being used as valid predictors of performance. Most of the negative material seems to be directed toward the fact that people who complete reference forms or recommendation letters are not always the right people to give a true picture of the recommendee. Either the recommenders are not qualified because they are friends or well wishers, or because the recommendee only asks those s/he thinks will give a favorable recommendation letter.

Low Validity

One problem that researchers have to confront is the low validity of the recommendation letter itself. Most of the efforts to validate recommendations involve correlations with either supervisory ratings or tenure (Siegel & Lane, 1987). Research shows that recommendation letters and reference ratings typically have validity coefficients of about .13 (Browning, 1968; Mosel & Goheen, 1959; Muchinsky, 1979; Reily & Chao, 1982). Although references may have some predictive value, the average level of prediction is quite low (Browning, 1968; Mosel & Goheen, 1958). Paunonen *et al.* (1987) contend that one reason for this low validity and predictive failure is that the reference reports are usually quite homogeneous with respect to the evaluation of individuals' attributes as well as their qualifications.

Leniency

A second problem with references stems from the leniency of recommendations. That is, recommendation writers tend not to make negative comments about the individuals about whom they are referencing. When

unfavorable or negative information is not included in letters of recommendation, there is no room to discriminate among candidates (Cascio, 1987).

Tucker and Rowe (1979) suggest one reason we only find favorable information in recommendation letters as opposed to unfavorable information is that there is a tendency to attribute the cause of past success/failure in performance to the individual rather than external influences that may be present. This tendency to internally attribute past success/failure causes many reference writers not to include unfavorable information because of the harm it may cause them or the individual for whom they are writing the letter. Mosel and Goheen (1959) found a tendency for friends and acquaintances to give the most favorable evaluations and that their recommendations are frequently more biased as well as more lenient than recommendations from previous employers, co-workers and supervisors. Reily and Chao (1982) found recommendations rarely contain negative statements about the recommendee, perhaps providing a reason why recommendations have been dismissed by many researchers. Even though the potential validity has been dismissed, the idea of past performance as being a good predictor of future performance is a good one and previous employers' evaluations (the recommendation letter) should have some merit.

Lack of Opportunity for Observation

When an individual is asked to write a reference for another individual, more than likely the individual doing the writing is a former supervisor, employer, or peer. Because this is the case, one has to take into consideration that the reference writer does not always know how the recommendee has performed on a certain job. This is where the lack of opportunity to observe comes in; the recommender does not have specifics on how well an individual performs her/his duties. Even the supervisor is sometimes without true knowledge of how the recommendee performs.

Another point that needs to be considered is that letter writers are often rarely given much of a description of the job for which they are making the recommendation (Siegel and Lane, 1987). This makes room for the recommender to generalize about the recommendee instead of using specifics.

One last factor that relates to the lack of observation is that different types of people view different types of behavior. This makes the rating

process itself an even more difficult task.

Extraneous Factors

Extraneous factors are those factors that letter writers often use in writing or evaluating references that are unrelated to occupational skill or ability and do not play a role in the evaluation process (Golightly et al; 1972). According to Baxter et al. (1981), writers of recommendation letters do not describe distinctive qualities of the recommendee and an individuals' qualities are not always mentioned. These omissions are because the letter writers are often not given a good description of the position for which they are making the recommendation. Due to this lack of knowledge, the letter writers tend to construct their own ideas about the position and duties needed.

Knouse (1983) states that people who write recommendations should use specific examples in order to clarify statements that are too general about a recommendees' performance. If specific examples are used, then the position description would no longer be needed to determine what to say about a recommendee. This brings up the fact that many recommendations are in some ways useless. Even though great value is placed on the information in these letters, they are useless because they tend to be vague, general, and give only a favorable account of the recommendees (Sleight & Bell, 1954). The letters are also useless because there is a lack of understanding between the person who writes the letter of reference and the person who requests the letter of reference as to what type of information is needed and desired.

There is a lot of skepticism about recommendations and there needs to be an improvement in the way references are handled by an organization. Not only do the aforementioned problems exist, but there are legal ramifications that influence how individuals write a letter of recommendation as well as what they say to another individual about the person s/he is referencing. An example of these legal ramifications is the Family Educational and Privacy Act that permits students to waive or not waive their right to see their letters of recommendation when applying to institutions of higher learning (Muchinsky, 1986).

McCormick and Ilgen (1985) suggest that for references of any kind to be useful at least four conditions must be fulfilled: 1) the recommender should have enough time and be able to observe the individual she/he is recommending; 2) the recommender must be competent; 3) the recom-

mender must be able to write clear and concise opinions so that the opinions are interpreted correctly; and 4) the recommender must give those opinions about the recommendee.

Methods for Improvement

In 1962, Peres and Garcia decided that reference letters were too difficult to use as an effective guide in the selection process for identifying desirable engineers. The letters they received for the engineering positions described applicants in terms of adjectives or generalized traits as opposed to observable behavior. Peres and Garcia developed a new and better way of predicting future performance by using a checklist. Peres and Garcia (1962) conducted a factor analysis of 170 adjectives chosen from letters of reference resulting in five factors or categories that each consisted of a list of adjectives. The checklist was used not only as a means of identifying undesirables, but also as a method to predict performance evaluations. Other researchers such as Aamodt, Bryan, and Whitcomb (1989) report validities of .32 and .38 when recommendation letters based on Peres and Garcia are correlated with performance evaluations.

Carroll and Nash (1972) suggested that another way to correct or eliminate most of the problems that exist with recommendations would be to use a reference checklist as opposed to having someone write a recommendation letter. They suggest that leniency on part of those writing the letter of recommendation could be reduced if a reference checklist were used along with a forced-choice format. If the forced-choice format is used along with the reference checklist, not only will the negative conclusions about the validity of the reference be reversed but the leniency as well. Nash (1971) states that the forced choice format has been used in tests for over 25 years but the success as an appraisal instrument has been limited.

Reily and Chao (1982) also cite studies that attempt to reduce leniency and improve validity by using the checklist. Rhea (1966) tried to predict the success of junior naval officers in officer candidate school by using the checklist and Carroll and Nash (1972) attempted to predict supervisory ratings for secretarial/clerical hires and found that information from references provided very little to other predictors.

With the use of the reference checklist and the forced-choice format, the inability of a given predictor to predict could be resolved and the

reluctance of an employer to provide negative information about a former employee to other employers would be eliminated (Carroll and Nash, 1972). However, if reference givers are forced to choose among negative sounding alternatives when they only want to say something positive, then there will more than likely be negative reactions to the reference checklist. Therefore, the reference checklist should be of forced-choice format that will always have positive responses. Berkshire and Highland (1953) have supporting research that states that the problem of negative responses can be dealt with by using only positive sounding items. The checklist also makes past employment easier to verify. By using the reference checklist as opposed to recommendation letters, many of the problems presented earlier may not be prevalent.

Under the forced choice format, a reference giver would be forced to choose the items of behavioral descriptions among a group of items which best describe a former employee. Because all of the items used in the forced - choice reference form appear to be favorable or positive to the reference giver, this allows or enables reference givers to satisfy their needs for giving a good reference as well as helping to alleviate the problem of leniency (Carroll & Nash, 1972).

There is an increase in the use of structured, standardized questionnaires and a decrease in the use of open-ended letters of recommendation (Scott et al, 1954). This increase has brought about the reemergence of the reference checklist. One such checklist is the Employment Recommendation Questionnaire (ERQ). The ERQ obtains judgments and information by mail from persons familiar with the applicants and it also attempts to utilize the judgments and information of individuals familiar with applicants as an aid in personnel selection. The ERQ's, which are reference checklists, may in fact, tell how well an applicant performed on a previous job but it is used in order to obtain solid references that will not produce inadequate responses that recommendation letters may produce (Mosel & Goheen, 1958b). ERQ's are used in industry for two reasons: 1) to verify claimed work history and 2) to secure further information not otherwise obtainable.

The reference checklist has two major advantages: 1) brevity and 2) completion time and a high probability of percentage returns (Sleight & Bell, 1954). Since the quality of performance is very difficult to verify when using a recommendation letter, the need for a reference checklist is very important and almost a must (Ceci & Peters, 1984).

Statement of Purpose and Hypotheses

The purpose of this study was to investigate the relationship between the reference checklist and performance evaluations. It was hypothesized that when the performance evaluations and the reference checklists were correlated, the following would occur: 1) the correlation between the total category score of the reference form and the overall performance evaluation score would be significant, 2) the correlation of the five categories with the individual categories on the performance evaluation would predict performance ratings for categories under which they fall: a) dependability would predict dependability and responsibility, b) vigor would predict initiative, c) mental agility would predict judgment, d) cooperation-consideration would predict cooperation and e) urbanity would predict personal appearance and courtesy and friendliness, f) the correlation between the variables would prove to be a good predictor of performance.

Method

Subjects

A total of 92 students were recruited to participate in this study: 42 students from undergraduate psychology courses, 30 dining hall scholarship students, and 22 students from various other job sites. Jobs held by the participating students were switchboard operator, linen and photocopy worker, reading clinic assistant, information office assistant, file clerk, office clerk, library assistant, laboratory assistant and R.U. dispatcher. All of the participants, both male and female, were currently enrolled in the College Work Study program at Radford University. Because complete data sets were needed for this study, a total of only 39 students were used in the actual analysis.

Procedure

A brief explanation of the study was first given to the participants before they were asked to complete a consent form. The completion of this consent form allowed the researcher to obtain and use both supervisor ratings of performance as well as a reference checklist completed by a reference. Once the consent forms were signed by the students, the

students were then asked to identify a reference. A reference checklist was then attached to a cover letter and sent to the reference specified by the subject. An addressed, stamped return envelope was also provided with the reference checklist for the convenience of the students' reference. The cover letter and reference checklist were mailed to the students' reference by the researcher. Once the reference checklist was sent to the reference, the performance evaluation of each participating student was then sent to the student's College Work Study Supervisor.

The Reference Checklist

The reference checklist used in this research consisted of a list of words taken from categories developed by Peres and Garcia (1962). These categories are Vigor, Urbanity, Mental Agility, Dependability-Reliability, and Cooperation- Consideration.

To develop this checklist, students were given a list of words from the five categories and asked to rate a friend according to the scale provided. After several trials of administering a list of potential words for the reference checklist, 15 words were chosen. The words chosen for the reference checklist had more reliability and discriminability than those not chosen.

The checklist in this study consisted of a list of 15 words, three each from the Peres and Garcia categories. This list was not only developed to help determine the predictability of performance but also as a means to help the Financial Aid Office with the placement of College Work Study Students.

To assure participation, the references were told that this research would assess a new reference form and that before it could be implemented the validity of the form must be investigated. Each reference was asked to score or rate the individual on all items using the scale provided and return the form to the researcher using the enclosed self addressed, stamped envelope.

When the reference forms were returned, they were scored as follows: an individual receiving an "N" (not a good description of this person) for a specific description received zero points; an individual receiving a "O" (occasionally a good descriptor of this person) for a specific description received 1 point; and an individual receiving an "E" (an excellent description of this person) for a specific description received two points. The overall total for the categories combined were calculated as well as the

overall total for each category and the overall total for each word. An individual could receive a maximum of 30 points. For those words not marked, the average of the words within that category was taken and applied to that word. For example: if generous was not marked, for the category cooperation-consideration, and considerate was given a two and cooperative was given a two then the average of these two words were taken and applied to generous.

The Performance Evaluation

The performance evaluation used by the supervisors also consisted of a list of words, along with their definitions taken from the five categories derived by Peres and Garcia. The supervisor was asked to complete an updated version of the current performance evaluation used by the Financial Aid Office. A cover letter explaining the study and the purpose of completing an updated version of the performance of the College Work Study student accompanied the performance evaluation form. The supervisor was asked to rate the individual College Work Study student according to the rating scale given on the performance evaluation form. The supervisor was then asked to rate the student on a rehiring scale. Both the supervisors and the reference givers were given deadlines as to when they should return the information.

Once the performance evaluations were returned, performance scores were calculated. If an individual received an "A" for a specific category s/he received four points, if a "B" was marked on the performance evaluation the individual received three points, if a "C" was marked on the performance evaluation the individual received two points, a mark of "O" received one point, and an "N" received zero points. As with the reference checklist the overall score of the performance evaluation was calculated (total PA), the average score of the performance evaluation (AVEPA), the rate of rehire, and the individual score received on each category. An individual could receive a maximum of 52 points.

Results and Discussion

The dependent variable for this study was the students' performance evaluation and the reference checklist served as the predictor. To determine the validity of the reference checklist, each performance evaluation category was correlated with each of the reference checklist categories.

As shown in Table 1, the correlational analysis indicated that many categories from the reference checklist predicted ratings of performance in several categories. The cooperation-consideration category predicted ratings of performance for responsibility, quality, initiative, attitude, attendance and rehire. Mental agility predicted ratings of performance for quality, attitude and rehire as well as the average performance score. Vigor predicted ratings of performance for dependability and responsibility and Dependability predicted ratings of performance for cooperation, attitude and courtesy and friendliness and urbanity predicted ratings of performance for responsibility. Table 2 shows the intercorrelations between the reference categories

The hypotheses stated earlier in this study have been partially supported. Hypothesis one has been supported as the correlation between the overall performance evaluation score and the overall reference score resulted in significant correlations. The cooperation-consideration category predicted more ratings from the performance evaluation categories than any of the other categories from the reference checklist. However, when the reference categories are combined, the data suggest that mental agility is the single best predictor of performance.

The results did not support hypothesis two. The performance appraisal categories did not correlate significantly with any of the five reference categories under which they fall (i.e. the dependability-reliability category did not predict dependability, cooperation-consideration did not predict cooperation, and urbanity did not predict courtesy and friendliness and personal appearance). Significance was found only between performance appraisal categories and reference categories that were not logically related (i.e. cooperation-consideration predicted responsibility and vigor predicted dependability).

The correlational analysis also reveals that 23 of the 90 correlations are significant. Five of the 90 correlations would be expected by chance, but with 23 significant, it is doubtful that the results could have occurred by chance alone. It is obvious that the reference checklist has its limitations and one of those limitations deals with the fact that it can not be used for applicants without any previous employment history (Reily & Chao, 1982). However, the reference checklist used in this study is based on traits and may be usable anywhere.

Table 1

Correlations Between Reference Ratings and Performance Ratings

Performance Category	Reference Category					Overall
	CC	MA	V	D-R	U	
Dependability	.21	.28	.33*	.22	.05	.28
Responsibility	.45**	.57**	.38*	.42**	.32*	.56**
Initiative	.39*	.41**	.20	.27	.11	.35*
Judgment	.18	.28	.20	.11	.06	.21
Cooperation	.10	.18	.01	.38*	.02	.13
Leadership	.18	.19	.26	.12	.25	.30
Quality	.45**	.39*	.18	.14	.18	.37*
Attitude	.35*	.54**	.20	.35*	.20	.41*
Courtesy	.20	.27	.09	.35*	.02	.20
Personal Appearance	.14	.25	.06	.09	.16	.05
Potential	.18	.17	.04	.18	.08	.06
Attendance	.33*	.28	.27	.26	.04	.30
Tardiness	.05	.08	.02	.17	.05	.04
Rehire	.38*	.45**	.11	.16	.14	.33*
Avg. Performance Score	.30	.36*	.24	.30	.11	.33*

* P < .05

** p < .01

Table 2**Correlations Among Reference Categories**

Reference Category	CC	MA	V	DR	U
Cooperation/Consideration	1.00	.61**	.61**	.17	.26
Mental Agility		1.00	.55**	.24	.45**
Vigor			1.00	.11	.42**
Dependability/Reliability				1.00	.12
Urbanity					1.00

Another obvious limitation of this type of reference check is that the return percentage rate through the mail is very low. Mosel and Goheen (1958b) report a 56 percent return rate for 4,000 reference questionnaires. However, this high percentage for the rate of return is only due to the numerous times references were sent out for certain individuals. If a participant's reference did not return the reference questionnaire, another reference was sent to yet another reference. This was done until a reference for that particular individual was received. Although the sample size for this study is tremendously lower than the aforementioned study, the return rate was 14 percent below that of Mosel and Goheen. The return rate for this study was 42 percent for 92 participants.

Although this study proves that the reference checklist is a good predictor of performance, it is obvious that more research needs to be done on the validity of the reference checklist as a predictor of performance.

References

- Aamodt, M. G., Bryan, D. A., & Whitcomb, A. J. (1989). Predicting performance with letters of recommendation. *Proceedings of the Annual Meeting of the International Personnel Management Association Assessment Council*, Orlando, Florida.
- Baxter, J. C., Brock, B., & Hill, P. C. (1981). Letters of recommendation: A question of values. *Journal of Applied Psychology*, *66*, 296-301.
- Beason, G., & Belt, J. A. (1976). Verifying applicant backgrounds. *Personnel Journal*, *55*, 345-348.
- Berkshire, J., & Highland, R. (1953). Forced-choice performance ratings: A methodological study. *Personnel Psychology*, *6*, 355-378.
- Browning, R. C. (1968). Validity of reference ratings from previous employers. *Personnel Psychology*, *21*, 389-393.
- Carroll, S. J., & Nash, A. N. (1972). Effectiveness of a forced-choice reference check. *Personnel Administration*, 42-46.
- Cascio, W. F. (1987). *Applied Psychology in Personnel Management*. Englewood Cliffs, NJ: Prentice-Hall.
- Ceci, S. J., & Peters, D. (1984). Letters of reference: A naturalistic study of the effects of confidentiality. *American Psychologist*, *39*, 29-31.
- Cowan, G., & Kasen, J. H. (1984). Form of reference: Sex differences in letters of recommendations. *Journal of Personality and Social Psychology*, *46*, 636-645.
- Golightly, C., Huffman, P. M., & Byrne, D. (1972). Liking and loaning. *Journal of Applied Psychology*, *56*, 521-523.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*, 72-98.
- Knouse, S. B. (1983). The letter of recommendation: Specificity and favorability of information. *Personnel Psychology*, *36*, 331-342.
- LoPresto, R., Mitcham, D. E., & Ripley, D. E. (1985). *Reference Checking Handbook*. Alexandria, VA: ASPA.
- Maier, N. R. F., & Verser, G. C. (1982). *Psychology in industrial organization*. NJ: Houghton, Mifflin.

- McCormick, E. G., & Ilgen, D. (1985). *Industrial and Organizational Psychology*. Englewood Cliffs, NJ: Prentice-Hall.
- Mosel, J. N., & Goheen, H. W. (1958a). Use of the "ERQ" in hiring. *Personnel Journal*, 36, 338-340.
- Mosel, J. N., & Goheen, H. W. (1958b). The validity of the employment recommendation questionnaire in personnel selection: Skilled traders. *Personnel Psychology*, 11, 481-490.
- Mosel, J.N., & Goheen, H. W. (1959). The employment questionnaire III: Validity of different types of reference. *Personnel Psychology*, 12, 469-477.
- Muchinsky, P. M. (1979). The use of reference reports in personnel selection: A review and evaluation. *Journal of Occupational Psychology*, 52, 287-297.
- Muchinsky, P. M. (1986). Personnel selection methods. In C. L. Cooper & I. T. Robertson (Eds.), *International Review of Industrial and Organizational Psychology* (pp. 37-70). NY: John Wiley and Sons.
- Nash, A. N. (1971). Modification of forced-choice format for use in personnel selection and appraisal. *Psychological Reports*, 71, 108-110.
- Paunonen, S. V., Jackson, D. N., & Oberman, S. M. (1987). Personnel selection decisions: Effects of applicant personality and the letter of reference. *Organizational Behavior and Human Decision Processes*, 40, 96-114.
- Peres, S. H., & Garcia, R. (1962). Validity and dimensions of descriptive adjectives used in reference letters for engineering applicants. *Personnel Psychology*, 15, 279-286.
- Pursell, E. D., Campion, M. A., & Gaylord, S. A. (1983). Structured interviewing: Avoiding selection problems. In K. Pearlman, F. L. Schmidt, & W. C. Hamner (Eds.), *Contemporary Problems in Personnel* (pp. 112-121). NY: John Wiley & Sons.
- Reily, R. R., & Chao, G. T. (1982). Validity and fairness of some alternative employee selection procedures. *Personnel Psychology*, 35, 1-62.
- Rhea, B. D., Rimland, B., & Githens, W. H. (1965). *The development and Evaluation of a Forced Choice Letter of Reference Form for Selecting Officer Candidates*. (Technical Bulletin STB 66-10). San Diego, CA: U. S. Naval Personnel Research Activity.
- Scott, W. D., Clothier, R. C., & Spriegel, W. R. (1954). *Personnel Management*. New York: McGraw-Hill.

- Siegel, L., & Lane, I. M. (1987). *Personnel and Organizational Psychology*. Homewood, IL: Irwin.
- Sleight, R. B., & Bell, G. D. (1954). Desirable content of letters of recommendation: A question of values. *Journal of Applied Psychology*, 32, 421-422.
- Tucker, D. H., & Rowe, P. M. (1979). Relationship between expectancy, causal attributions and final hiring decisions in the employment interview. *Journal of Applied Psychology*, 64, 27-34.
- Yoder, D. (1962). *Personnel Management and Industrial Relations*. Englewood Cliffs, NJ: Prentice-Hall.