Predicting Performance with Letters of Recommendation

Previous research has indicated that letters of recommendation are poor predictors of future performance, in part because characteristics of the letter writer and letter reader interfere with the objective analysis of the content of the letter. To help correct this problem, Peres and Garcia (1962) developed a technique for analyzing the content of letters of recommendation by identifying traits mentioned in each letter and placing the traits into one of five categories. It was the purpose of this paper to determine if the Peres and Garcia technique would be a valid method of predicting performance of psychology instructors and graduate students. The results of the two studies indicate that traits from letters of recommendation can be reliably classified into the five Peres and Garcia categories and that these traits are valid predictors of future performance (r's = .32 and .38).

By
Michael G. Aamodt
Devon A. Bryan
Alan J. Whitcomb

In psychology, it is commonly believed that the best predictor of future performance is to look at past performance. Thus, if an organization wanted to hire a fire fighter, the best applicant might be one who has not only previously been a fire fighter, but one who was also a successful fire fighter in his/her previous job.

While it is not very difficult to verify the previous employment of an applicant, it can be rather difficult to verify the quality of his/her previous performance. The authors of this article recently watched the National Football League (NFL) draft on television and were envious of the fact that professional football teams could assess a college player's previous performance by watching game films. That is, the football scores did not have to rely on opinions of the other coaches. Instead, they could watch every minute that the player had played while in college.

Unfortunately, very few applicants bring "game films" of their previous employment performance. Instead, an employer must obtain information about the quality of an applicant's previous performance by asking an applicant either to supply names of references that the employer can call or to provide letters of recommendation from previous employers.

Michael G. Aamodt, Ph.D. is an associate professor of industrial/organizational psychology at Radford University.

Devon A. Bryan received his Master's degree in I/O psychology at Radford University and is now a compensation analyst for J.I. Case.

Alan J. Whitcomb received his Master's degree in I/O psychology at Radford University and is currently a manager for TW Recreation Services.

Even though references are commonly used to screen and select employees, they have not been successful in predicting future employee success (Muchinsky, 1979). In fact, the average validity coefficient for references is only .13 (Browning, 1968; Mosel & Goheen, 1959). This low validity is due mostly to four main problems found with references and letters of recommendation: Leniency, knowledge of the applicant, low reliability, and extraneous factors involved in the writing and reading of letters of recommendation.

Leniency

Choice of References - Research is very clear on the fact that most letters of recommendation are positive (Carroll & Nash, 1972; Yoder, 1962; Myers & Errett, 1959). Because at some point we have all worked with terrible employees, it would at first seem surprising that references are so positive. However, it should not be surprising when it is kept in mind that applicants choose their own references. If given the chance to choose their own references, even undesirables such as Nazi leader Adolph Hitler, serial killer Ted Bundy, and terrorist Abu Nidal would be able to find three people who would provide them with favorable references.

Confidentiality of the Reference - A second factor that will increase the leniency of references is the confidentiality of the reference. By law, people have a right to see their reference letters, but by signing a waiver, they can give-up that right. Research by Ceci and Peters (1984) and by Shaffer and Tomarelli (1981) has indicated that references tend to be less lenient when the applicant waives his/her right to see the letter of reference. That is, if a person writing a reference letter knows that the applicant will be allowed to see the letter, the letter writer will be more inclined to say nicer things.

Sex and Race - Two other minor factors which affect the leniency of references are the sex of the letter writer and the race of the letter reader. Carroll and Nash (1972) found that female letter writers are more lenient when referring female applicants while Bryan (1989) found that black professionals are more lenient in evaluating the contents of letters than are white professionals. The combination of applicants choosing their own references, retaining the right to see a reference letter, being referred by a female reference, and being evaluated by a black professional will make the reference letter much more positive than it should be based on the applicant's actual performance.

Knowledge of the Applicant

A second problem with letters of recommendation is that often, the person writing the letter either does not know the applicant well or has not observed all aspects of the applicant's behavior. Professors are always being asked to supply recommendations for students who they only know from one class. Such a recommendation is not likely to be as accurate and complete as one provided by a professor who has taught the student in several classes as well as worked with the student outside the classroom.

Even in a work setting where the supervisor is supplying the recommendation, he/she does not see all aspects of the employee's behavior.

Employees often act very differently around their supervisors than they do around co-workers and customers. Furthermore, the behaviors that are actually recalled when making a recommendation are only a fraction of the behaviors that actually occurred in the presence of the person writing the recommendation.

Reliability

The third problem with references and letters of recommendation involves the lack of agreement between two people who provide references for the same person. Research has revealed that the reliability of references is only about .40 (Baxter, Brock, Hill, & Rozelle, 1981; Mosel & Goheen, 1959; Mosel & Goheen, 1952). The reliability problem is so severe that Baxter and his colleagues (1981) found that there is more agreement between two recommendations written by the same person for two different applicants than there is between two people writing recommendations for the same person. Thus, letters of recommendation may say more about the person writing the letter than about the person being written about.

This low reliability is probably due to the idea alluded to earlier that a person writing a reference has not seen all aspects of the applicant's behavior. Thus, a reference provided by a professor who has seen the applicant in the classroom would not necessarily be expected to agree with a reference provided by a supervisor who has seen the applicant in a work setting. However, even though there might be a good reason for the low reliability, it does limit the potential validity of references and raises an interesting question that research has yet to answer; if references don't agree, which one should be taken most seriously?

Extraneous Factors

The fourth problem with letters of recommendation involves extraneous factors which affect the writing of the letters. Research has indicated that the method used by the letter writer is often more important than the actual content. For example;

1) Knouse (1983) found that letters which contained specific examples were rated higher than letters that contained generalities,

2) Cowan and Kasen (1984) found that male and female letter of recommendation writers use different forms of reference when referring to applicants in their letters - females refer to applicants by titles such as Mr. or Mrs. while males refer to applicants by first names, and

3) Mehrabian (1965) and Weins, Jackson, Manaugh, and Matarazzo (1969) found that even though most letters of recommendation are positive, letters written by references who actually like the applicant are longer than letters written by references who are not as positive toward the applicant.

Improving the Validity of References

To improve on the low validity of references, Peres and Garcia (1962) developed a unique way to make letters more useful by focusing attention on the relevant content of the reference letter rather than on the positiveness of the letter. As an example, look at the following two hypothetical letters of recommendation.

(A)

Dear Personnel Director:

Mr. John Anderson asked that I write this letter in support of his application as manager and I am pleased to do so. I have known John for six years as he was my assistant in the accounting department.

John always had his work completed accurately and promptly. In his six years here, he never missed a deadline. He is very detail oriented, alert in finding errors, and methodical in his problem solving approach. Interpersonally, John is a very friendly and helpful person.

(B)

Dear Personnel Director:

Mr. John Anderson asked that I write this letter in support of his application as manager and I am pleased to do so. I have known John for six years as he was my assistant in the accounting department.

John was one of the most popular employees in our agency as he is a friendly, outgoing, sociable individual. He has a great sense of humor, is poised, and is very helpful. In completing his work, he is independent, energetic, and industrious.

Both of the letters describe the applicant in very favorable terms. However, the two letters differ greatly in the words that are used to describe the applicant rather than in the favorability of the words.

After examining thousands of letters of recommendation, Peres and Garcia (1962) found that all of the adjectives that were contained in letters of recommendation could be placed into one of five categories: dependability-reliability, consideration-cooperation, mental agility, urbanity, and vigor. A complete list of the trait words in each category can be found in Peres and Garcia (1962).

To use letters of recommendation to predict performance, an employer would:

1) Determine the importance of each of these five categories to the performance of a particular job,

2) Read each letter of recommendation and underline the traits in each letter used to describe the applicant,

3) Use the list of words composed by Peres and Garcia (1962) to place each trait into one of the five categories, and

4) Total the number of words for each of the five categories.

To demonstrate this process, Figure 1 (see appendix) shows the traits from the previous hypothetical letters that have been underlined and summed, providing a score on each of the five categories for the two letters.

As promising as this technique sounds, Peres and Garcia (1962) unfortunately did not attempt to validate their technique. Thus, it is the purpose of this study to investigate the reliability and validity of the technique using two separate samples.

Method

Subjects

The subjects for the first sample consisted of 78 (39 male, 39 female) former graduate students who had completed the graduate program in psychology at Radford University. The subjects for the second sample consisted of 26 (11 male, 15 female) Graduate Teaching Fellows (GTFs) at Radford University. Each GTF was working toward a Master's Degree in psychology and had complete responsibility for teaching two sections of introductory psychology each semester.

Procedure

Graduate Student Sample - Because Radford University uses a reference rating form and actual letters of recommendation are not required, the files of over 200 former graduate students were first examined to locate those students who had at least one letter of recommendation. This process resulted in 78 students for whom at least one letter of recommendation as well as an overall graduate G.P.A. were available. Two of the authors then independently:

1) Read each letter

2) Highlighted the traits in each letter

3) Used the list composed by Peres and Garcia (1964) to place each highlighted trait into one of the five categories

4) When two letters were available, the number of traits in each of the five categories were averaged across the two letters

5) To control for effects of letter length and number of traits used, the number of traits in each category was divided by the total number of traits across the five categories.

   Graduate Teaching Fellow Sample - The letters of recommendation for each of the 26 GTFs were analyzed according to the procedure listed above. However, student ratings of the GTF's teaching served as the criterion rather than the GPAs used with the graduate student sample. The rating used was the final question on the rating form already used by the university which asked for an overall rating of the instructor's performance based on a five point rating scale with a "1" indicating poor performance and a "5" indicating excellent performance.

Results and Discussion

Reliability of Letter Writers

   To determine the extent to which letter writers referred to applicants with similar traits, the number of traits in each of the five categories sued by each letter writer were correlated. As shown in Table 1, the coefficients across the two samples were fairly low, and in some cases the coefficients were negative. Thus, it would appear that two people writing letters for the same individual will not say the same things.

Reliability of the Letter Readers

   To determine the extent to which personnel professionals reading each letter agree about the traits that are present as well as the category in which each trait belongs, the number of traits placed by the two raters in the five categories for each letter were correlated. As shown in Table 2, the coefficients were reasonably high for the first sample with the exception of the urbanity category. One of the problems encountered in the first sample was that many of the traits listed in the letters were not contained in Peres and Garcia's lists. Prior to collecting data for the second sample, these new traits were added to the lists and as can be seen from Table 2, the agreement levels increased for four of the five categories.

Table 1   Agreement of Recommendation Letter Writers

	Sample	
Trait Category	Graduate Students	Teaching Fellows
Mental Agility	.12	.18
Vigor	-.03	-.58
Dependability-Reliability	.04	.48
Urbanity	.08	.31
Cooperation-Consideration	.15	.17

Table 2   Agreement of Recommendation Letter Readers

	Sample	
Trait Category	Graduate Students	Teaching Fellows
Mental Agility	.77	.91
Vigor	.86	.64
Dependability-Reliability	.70	.86
Urbanity	.53	.86
Cooperation-Consideration	.87	.96

Validity of the Trait Categories

To determine the validity of the trait categories, the number of traits in each of the five categories for the first sample was /correlated with the student's graduate GPA and the number of traits in each of the five categories for the second sample was correlated with the GTF's overall student teaching ratings. As shown in Table 3, the number of traits in the mental agility category significantly correlated with graduate GPAs while the number of traits in the urbanity category positively correlated with teaching ratings and the number of traits in the mental agility category negatively correlated with teaching ratings.

Table 3   Validity of Trait Categories

	Sample	
Trait Category	Graduate Students	Teaching Fellows
Mental Agility	.32	-.44
Vigor	-.08	.27
Dependability-Reliability	-.13	-.22
Urbanity	.03	.38
Cooperation-Consideration	.04	.23

While the negative correlation between mental agility and teaching ratings may seem surprising, it is consistent with data kept by our department which indicate that GRE scores are also negatively correlated with

teaching success. Apparently our general psychology students have difficulty following instructors who are overly bright.

In addition to being valid single predictors of graduate student G.P.A. and teaching performance, the letters also showed incremental validity over other predictors. Graduate Record Exam (GRE) scores were the single best predictor of graduate student grades ($r = .41$), yet because GRE scores and traits in the mental agility category are only slightly correlated ($r = .04$), the addition of the mental agility category to GRE scores yields a very significant multiple R of .71.

In an attempt to measure the incremental validity of the trait categories in predicting teaching performance, we entered GRE scores, undergraduate G.P.A. and graduate G.P.A. into a regression equation along with the trait categories. The trait categories were the only significant predictors of teaching performance.

Overall, our findings indicate that the technique developed by Peres and Garcia (1964) shows promise as a predictor of performance. With both samples, the significant validity coefficients were more than twice the magnitude of the .13 previously found with references, and accounted for additional variance when compared with currently used predictors of performance.

While the Peres and Garcia categories were successful in predicting both criteria, it is possible that traits might be better classified into a different system. For example, Peres and Garcia's dependability-reliability category appears to be two separate categories; one involving dependability and includes items such as "responsible" and "dependable" and another involving assertiveness with traits such as "tenacious", "confident", and "determined". This would be an excellent topic for future research.

An interesting finding in our study was that the traits used by used letter writers to describe the same person were not highly correlated. This finding is consistent with the low reliability found by Mosel and Goheen (1952) and by Baxter, Brock, Hill, and Rozelle (1981) and certainly makes sense if one assumes that each letter writer probably observes different aspects of the applicant's behavior as would be the case if one of the letter writers were a professor and the other an employer.

However, even though the low agreement of letter writers is understandable, it does pose potential problems for the validity of the Peres and Garcia technique. Thus, it is important in the future to investigate issues such as the sources from which letters should be obtained as well as the optimal number of letters that should be written for each applicant. Because highlighting and categorizing traits can be a time consuming process,

future research might also focus on the development and validation of a trait based checklist such as the one created by Carroll and Nash (1972).

References

Baxter, J.C., Brock, B., Hill, P.C., & Rozelle, R.M. (1981). Letters of recommendation: A question of value. *Journal of Applied Psychology, 66,* 296-301.

Browning, R.C. (1968). Validity of reference ratings from previous employers. *Personnel Psychology, 21,* 389-393.

Bryan, D.A. (1989). Differences in trait interpretations between black and white professionals when evaluating letters of recommendation. *Proceeding of the Annual Graduate Conference in I/O Psychology and Organizational Behavior,* New Orleans, LA.

Carroll, S.J., & Nash, A.N. (1972). Effectiveness of a forced-choice reference check. *Personnel Administration, 35,* 42-146.

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 recommendation. *Journal of Personality and Social Psychology, 46, 636-645.*

Knouse, S.B. (1983). The letter of recommendation: Specificity and favorability of information. *Personnel Psychology, 36,* 331-341.

Mehrabian, A. (1965). Communication length as an index of communicator attitude. *Psychological Reports, 17,* 519-522.

Mosel, J.N., & Goheen, H.W. (1952). Agreement among replies to an employment recommendation questionnaire. *American Psychologist, 7,* 365-366.

Mosel, J.N., & Goheen, H.W. (1958). The validity of the Employment Recommendation Questionnaire in personnel selection: Skilled traders. *Personnel Psychology, 11,* 481-490.

Muchinsky, P.M. (1979). The use of reference reports in personnel selection: A review and evaluation. *Journal of Occupational Psychology, 52,* 287-297.

Myers, J.H., & Errett, W. (1959). The problem of pre-selection in weighted application blank studies. *Journal of Applied Psychology, 43,* 94-95.

Peres, S.H., & Garcia, J.R. (1962). Validity and dimensions of descriptive adjectives used in reference letters for engineering applicants. *Personnel Psychology, 15,* 279-296.

Shaffer, D.R., & Tomarelli, M. (1981). Bias in the ivory tower: An unintended consequence of the Buckley Amendment for graduate admissions. *Journal of Applied Psychology, 66,* 7-11.

Weins, A.N., Jackson, R.H., Manaugh, T.S., & Matarazzo, J.D. (1969). Communication length as an index of communicator attitude: a replication. *Journal of Applied Psychology, 53,* 264-266.

Yoder, D. (1962). *Personnel management and industrial relations.* Englewood Cliffs, NJ: Prentice-Hall.

Appendix

Figure 1
Example of the Peres and Garcia (1962) method for scoring letters of recommendation.

Dear Personnel Director:

Mr. John Anderson asked that I write this letter in support of his application as assistant manager and I am pleased to do so. I have known John for six years as he was my assistant in the accounting department.

John always had his worked completed <u>accurately</u> and <u>promptly</u>. In his six years here, he <u>never missed a deadline</u>. He is very <u>detailed</u> oriented, <u>alert</u> in finding errors, and <u>methodical</u> in his problem solving approach. Interpersonally, John is a very <u>friendly</u> and <u>helpful</u> person.

I have great confidence in John's ability. If you desire more information, please let me know.

MA <u>0</u>   CC <u>2</u> DR <u>6</u> U <u>0</u>   V <u>0</u>

Dear Personnel Director:

Mr. John Anderson asked that I write this letter in support of his application as assistant manager and I am pleased to do so. I have known John for six years as he was my assistant in the accounting department.

John was one of the most <u>popular</u> employees in our agency as he is a <u>friendly, outgoing, sociable</u> individual. He has a great sense of <u>humor</u>, is <u>poised</u>, and is very <u>helpful</u>. In completing his work, he is <u>independent, energetic</u>, and <u>industrious</u>.

I have great confidence in John's ability. If you desire more information, please let me know.

MA <u>0</u>   CC <u>2</u> DR <u>0</u> U <u>5</u>   V <u>3</u>

Key   *MA=mental agility*
           CC=consideration-cooperation
           DR=dependability-reliability
           U=urbanity
           V=vigor