

## Effects of Paper Color and Print Type on Resume Desirability

Claude Henson  
Dominion Bankshares

*In this study, the effects of paper color (white, off-white) and print type (typeset, high quality typewriter, dot matrix printer) on resume desirability were investigated. The results indicated that resumes should be written on white paper and that typeset resumes are evaluated the same as resumes printed with a high quality typewriter. Dot matrix printing was rated as least desirable*

---

When looked at closely, there are a number of variables relating to a resume that may influence its overall perceived desirability as a selection tool. Some of these variables include paper length, paper width, paper quality, content of the resume, format, resume length, type of print, and color of paper.

This paper investigates two of these variables: type of print and color of paper. Using these two variables, the question posed is, "Does paper color and/or type of print on a resume influence its desirability as in initial selection tool?"

A second question in relation to this research deals with how college students perceive these two variables. That is, "Do college juniors and seniors place the same emphasis on these variables as do personnel professionals?" Literature regarding these aspects of resumes is basically that of anecdotal evidence. Empirical research regarding these variables is almost nonexistent.

Advice varies from using only white bond paper (Horton, 1979) to using paper that is off-white, light blue, or gray in color (Kern-Foxworth,

1983). Empirical data regarding the color of resume paper is limited in scope. No data were found regarding resumes, though the literature review did yield some research with paper color relating to the use of questionnaires.

Pressley and Tuller (1977) found no significant difference in the return rate of mailed questionnaires when paper color varied from blue, yellow, green, or white. In a similar study, Jobber and Sanders (1983) mailed out blue and white questionnaire. No affect was found regarding the color of paper used in the questionnaires.

Similarly the advice regarding type of print was anecdotal in nature. The advice generally was that the resume should be typeset if financially feasible, though typesetting should not make a difference (Kern-Foxworth, 1983). Micolo (1984) recommends typesetting as a means for individuals to differentiate their resumes from those of other people.

The focus of this research was to look at these three variables: paper color, type of print, and occupational status (HRM professionals compared to students).

## **Method**

### **Subjects**

HRM professionals were defined as anyone in a position to review resumes fo ruse as an initial selection tool (n=32). The students were junior and senior level college students enrolled in introductory level industrial psychology classes at a midsized southern university (n=32).

### **Materials**

The materials consisted of nine resumes in which content, format, and paper bond were held constant. This was accomplished by first selecting and then randomly assigning nine commonly used names. All names were men's names to eliminate gender type as a variable. All applicants were students at Radford University with only the mailing address being different. The applicants were also given the same degree as well as the same graduation date. Work experience was held constant by randomly assigning jobs to this category. The jobs were matched for equivalency prior to assignment through use of a pretest.

The category of college activities was held constant by randomly assigning one actual organization and one fictitious organization to each resume. One fictitious fraternity was also randomly assigned to each resume from a list of nine such fraternities. The resumes were then printed on plain white paper using the same format for each. The resumes were then subjected to a pilot study to test for equivalency prior to random assignment to one of nine cells formed by a three (paper color: white, off-white, blue) by three (print type: type set, typed, dot matrix) matrix.

After being assigned to one of the nine cells, the resumes were then printed onto the appropriate color paper dependent upon the cell variable assignment. Paper quality was held constant by printing all of the resumes onto "Classic Laid" paper of 25% rag bond.

The method for rating the resumes consisted of a slip of paper attached to each resume with a scale of one to seven which was anchored good to bad.

The nine resumes with an attached rating card were then placed in random order to form a packet of nine resumes for each rater to rate. A standardized instruction sheet stating that the resumes were to be rated independently from each other was included as a cover sheet to each of the resume packets.

### **Procedure**

The resumes were pre-tested prior to assignment to the nine cells for variable assignment. The subjects for the pretest were eight university students enrolled in a graduate level employee selection and placement class.

After appropriate preparation of the resumes and development of the packets as previously mentioned, the resumes were then distributed to the professionals for rating by a basic door to door canvassing of businesses in the southwest Virginia area. They were collected in the same manner.

Students from three undergraduate industrial psychology classes at Radford University served as raters for the second aspect of the research regarding occupational status.

## Results

### Pretest

A one-way within subjects ANOVA was used to analyze the results of the test for equivalency. No significant difference was found between the nine resumes.

### Data Analysis

Three main effects and one interaction were significant. These were: occupational status, color of paper, type of print, and the interaction between status and type of print. The effect for occupational status indicated that students rated the resumes as being lower in quality than did the professionals,  $F(1, 558) = 10.95, p < .001$ . In regard to color, white was rated better overall when compared to blue,  $F(2, 558) = 3.35, p < .05$ . Type of print also was a significant variable. Both typeset and typed resumes were rated as being higher in quality than resumes printed with a dot matrix printer,  $F(2, 558) = 10.16, p < .001$ .

The interaction of occupational status by print type was significant,  $F(2, 558) = 4.02, p < .05$ . Students rated dot-matrix printed resumes as being of higher quality than did the HRM professionals.

## Discussion

The data generally support the advice given by professionals in the business community. That is, high quality type on clean white paper is preferred by persons in the position of selecting applicants for jobs.

Students rated resumes lower overall than did H.R.M. professionals. This could be due to a tendency of students to rate peers more critically. This tendency could be due to the fact that college is a competitive environment.

The implications regarding dot matrix as a print type should be noted here. In this era where many students either own or have access to a word processor, the tendency to use such for generation of a resume is great. As the data suggest, if one does generate a resume on a word processor, it should be done using a letter quality printer.

Further research needs to be conducted regarding other resume variables. These variables include paper quality, resume format, resume

content, and resume length. This is not an all-inclusive list. The empirical findings of such research may have substantial implications for a successful job search.

### References

- Horton, T. A. (1979). Resumes: tooting your own horn. *Essence*, 40-44.
- Jobber, D., & Sanderson, D. (1977). The effects of a prior letter and coloured questionnaire paper on mail survey response rates. *Journal of the Market Research Society*, 25, 339-349.
- Kern-Foxworth, M. (1983). A basic guide to resumes. *The Black Collegian*, 84-88.
- Micolo, A. (1984). How to write an effective resume. *The Computers and Electronics Graduate*, 62-63.
- Pressley, R., & Tuller, V. (1977). *Journal of Marketing Research*, 41, 108-112.

## References

have long been the most used and relied upon employee selection method, and it is estimated that over 90% of all businesses in the United States use the interview as their major selection tool (Beach, 1985). It is unfortunate then that research has consistently documented low levels of reliability and validity in interview situations (Arvey & Campion, 1982; Mayfield, 1964; Schmitt, 1976; Ulrich & Trumbo, 1965; Wagner, 1949).

Conceptually, it is not difficult to understand this lack of reliability and validity. Instead of questioning the applicant about job related knowledge and skills, the interviewer poses such questions as, "Where do you expect to be 10 years from now?" Even a good answer means little, for it not only is unrelated to performance, but chances are that interviewers will not agree on what constitutes a good answer.

Realizing that managers remain committed to the use of the interview despite such negative evidence, research in the past 10 years has sought to increase interview reliability through standardizing, or structuring the process. According to Weekly and Gier (1987), the most successful areas of research in terms of application of this structuring lay in the behaviorally based approaches. The most successful of these is the situational interview. Developed by Latham in 1980, situational interviews are based on the premise that a person's expressed behavioral intentions are related to subsequent behavior (Weekly & Gier, 1987).

In a situational interview, applicants are presented with a number of job related situations and asked what they would do in each case. Designed to identify behaviors critical to effective performance on the job (Latham & Saari, 1984), critical incidents are obtained from a job analysis and transformed into questions. Because the applicants' responses are given a score between 1 and 5 (1=poor, 5=outstanding), interviewers are supplied

with examples of responses (called “benchmark answers”) that would warrant a 1, 3, or 5 rating in order to guide them in the scoring process.

Compared to an unstructured interview’s interrater reliability coefficient of .35 (Landy, 1985), interrater reliabilities of situational interviews have been shown to be between .76 and .84 (Latham et al., 1980, Weekly & Gier, 1987). It is assumed that the utilization of benchmark answers partly explains this vast difference (Latham et al., 1980).

Because the purpose of providing these anchors is to reduce the subjectivity involved in scoring, it is logical to assume that agreement between raters would increase as a function of such a reduction.

As was previously discussed, the current format of a situation interview utilizes a 5-point rating scale with benchmark answers only at the 1, 3, and 5 levels. If one accepts the assumption that benchmark answers increase interrater reliability, it follows that adding one or more benchmarks to *every* level would further increase interrater agreement.

The purpose of this study was to test such an assumption. More specifically, if it was hypothesized that by assigning benchmark answers to each point level (1, 2, 3, 4, 5), the interrater reliability of a situational interview would be greater than an interview using only one benchmark at levels 1, 3, and 5.

## **Method**

### **Subjects**

All 39 participants (19 male, 20 female) were students of psychology at a medium sized university. Of these, 14 graduate and seven undergraduate students served as raters, while 18 lower level students served as interviewees.

### **Procedure**

This study involved three steps: (1) development of a situational interview for the position of teller at a credit union, (2) examination of the interrater reliability of that interview using 5 or more benchmarks as opposed to the same version using only three, and (3) data analysis.

**Step 1:** Using the method developed by Latham et al. (1980), over 185 critical incidents were obtained from a previously conducted job

analysis. In addition, four credit union employees (ranging from tell to branch manager) were interviewed, and another 27 critical incidents were generated. All 212 incidents were reviewed and categorized into 6 job dimensions. Two to three incidents from each dimension were converted into questions.

Once the questions were developed, 3 graduate students generated as many answers as possible for each questions and rated them on a scale of 1 to 5. It was decided that 2 questions were essentially equal to other ones so they were removed, resulting in a final interview of 14 questions.

**Step 2:** A total of 18 interviews were conducted. While the author actually administered questions to the interviewees, answers were scores by 4 raters who were present during the interview. To assist them in scoring, each rater had a copy of the interview questions, along with corresponding benchmark answers. All four had identical questions, although the number of benchmark answers varied. Two raters had 3 anchors weighted at the 1, 3, and 5 levels, and the other two had the same benchmarks, plus anchors at levels 2 and 4. After each question was answered by the interviewee, raters recorded their score on a separate answer sheet.

**Step 3:** For each subject, scores given by the two raters using three benchmarks were correlated with each other and the same was done with the raters using five or more benchmarks.

### **Results and Discussion**

The interrater reliabilities of the scoring methods using either three or five benchmark answers were calculated to be .44 and .66, respectively. These findings seem to support the hypothesis that adding more benchmark answers to a situational interview results in increased interrater reliability.

One possible explanation for these results is that the addition of more benchmarks reduces the amount of subjectivity in assigning point values to an interviewee's responses. Considering that many answers are not clearly bad (1), average (3), or outstanding (5), providing examples of "in between" answer would serve to reduce the amount of judgment that might otherwise be necessary with only three benchmarks.

Another reason for the higher reliability of interviews with five or more benchmarks may be a matter of probability. Simply put, the more benchmarks included, the higher the probability that an interviewee's response would exactly match one of the benchmarks. No judgment is



needed in the determination of an appropriate score. Therefore, any variability among scores would be severely reduced (if not eliminated) because the correct score is literally “given” to the interviewer. It would seem then, that in a situation such as this, it makes very little difference in interrater reliability whether or not 5 or 3 benchmarks were used. Since their purpose is to provide guidance for the rater, the benchmarks are useless under conditions when answers exactly match one of the benchmarks listed.

According to data in this study, however, the number of benchmarks had a definite impact on interrater reliability even when responses to questions generally matched the benchmarks. It is interesting to note that, in theory, if every possible answer to a question were listed as a benchmark, interrater reliability would be near perfect. The next question is: How many benchmarks would be needed to cover the entire range of answers? Also, is it necessary to cover every possible answer, or is there a limit to the reliability obtained? At what point does adding benchmarks cease to increase reliability?

Implications for further research may include the examination of practice effects on the reliability of the situational interview. Logic tells us that the provision of benchmarks assists raters in determining an appropriate point value, but after a person has conducted several interviews and has heard the entire range of possible answers, benchmarks would seem to be unnecessary. It would be interesting to measure at what point this may occur.

The results of this study are promising for the future of the situational interview. Although reliability and validity is already much higher than that of an unstructured interview, the evidence shows that adding a few more benchmarks will increase the interrater reliability even more, ultimately resulting in a more effective selection device.

### References

- Arvey, R. D., & Campion, J. E. (1982). The employment interview: A summary and review of recent research. *Personnel Psychology, 35*, 281-322.
- Beach, D. S. (1985). *Personnel: The management of people at work*. New York, Macmillan.
- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51*, 327-358.

- Landy, F. J. (1985). *Psychology of work behavior*. Homewood, IL: Dorsey Press.
- Latham, G. P., Saari, L. M., Pursell, E. D., & Campion, M. A. (1980). The situational interview. *Journal of Applied Psychology, 65*, 422-427.
- Latham, G. P., & Saari, L. M. (1984). Do people do what they say? *Journal of Applied Psychology, 69*, 569-573.
- Mayfield, E. C. (1964). The selection interview: Reevaluation of published research. *Personnel Psychology, 17*, 239-260.
- Schmitt, N. (1976). Social and situational determinants of interview decisions: Implications for the employment interview. *Personnel Psychology, 29*, 79-101.
- Ulrich, L., & Trumbo, D. (1965). The selection interview since 1949. *Psychological Bulletin, 63*, 100-116.
- Wagner, R. (1949). The employment interview: A critical review. *Personnel Psychology, 2*, 17-46.
- Weekly, J. A., & Gier, J. A. (1987). Reliability and validity of the situational interview for a sales position. *Journal of Applied Psychology, 72*, 484-487.