

Polly A. Phipps, Kenneth W. Robertson, Keith G. Keel
Bureau of Labor Statistics, 441 G St. N.W., Washington, D.C. 20212

Keywords: mail survey; t-test; anova.

1. Introduction

The Occupational Employment Statistics (OES) survey is a Federal-State establishment survey of wage and salary workers designed to produce data on occupational employment by industry for the Nation, each State, and selected areas within States. This survey is conducted by cooperating State employment security agencies with the assistance of the Bureau of Labor Statistics' regional and national offices. The OES survey provides employment data for approximately 700 detailed occupations by surveyed industries. It does not, however, provide any wage information for the occupational employment data. In fact, the current occupational wage data are limited in the level of occupational, industrial and geographic coverage.

In order to address this critical void in the Federal statistical effort, the OES program conducted pilot studies in 1989 and 1990 to test the feasibility of incorporating wage questions into the OES survey. With the addition of wage data, the OES survey will be able to provide occupational wage data at the necessary levels of coverage in a manner which will be cost effective and responsive to data quality concerns. Data produced by this survey will have several uses. For example, OES survey wage data can be used as components in the development of employment and training programs, for educational and career guidance, and as inputs into national and state industry-occupation employment matrices and projections.

In 1989, the OES wage pilot survey (OESWPS) yielded high response rates. In fact, initial mailing response rates were higher for units receiving the OESWPS than for units receiving the standard OES survey. This was also true for most states' subsequent mailing responses. One factor that was thought to contribute to the high response rate was the use of color questionnaires, a practice not in place for the OES survey. In assessing the response rates, States felt that color made the questionnaire easier for respondents to identify and remember.

In the 1990 OESWPS, an attempt was made to determine if questionnaire color contributed to high response rates. A controlled experiment was designed to measure the effect of color on response rates. In this paper we discuss the experiment and the test results. In section 2 we review the experimental literature on the effect of questionnaire color on survey response rates, and in section 3 we present our hypotheses. In section 4 we describe the OESWPS experiment and the techniques used in our analysis. We document the test results in section 5. Finally, in section 6 we offer conclusions on our experiment and make suggestions for further research.

2. Published Findings

In this section we review a number of studies which have tested the effect of questionnaire color on mail survey response rates. For each study, we include any theoretical reasons for use of a particular color, the type of respondents and the experimental findings. Many of the studies conduct multiple experiments, however, the only experiments that are described here are those primarily relating to color. Unless otherwise noted, the studies have carried out only one mailing of a questionnaire. In all of the experiments respondents are randomly assigned color and white questionnaires.

Color effects have been of interest to marketing and academic researchers for many years. In one of the earliest studies, Gullahorn and Gullahorn (1963) tested green and white questionnaires in a survey of former Fulbright and Smith-Mundt grantees (N=7,370). Green questionnaires were found to increase response rate in a positive direction, 51 versus 49 percent, but the difference was not statistically significant. Pucel and colleagues (1971) also tested the difference in response rates for green and white questionnaires, along with several other "incentives." They surveyed students who graduated from vocational programs (N=1,100). No incentive by itself, including color, was found to significantly increase response rates, although color did increase response in a positive direction. The number of incentives, however, did increase response rates significantly.

Matteson (1974) mailed pink and white survey questionnaires to members of a professional organization (N=2,040). Pink was chosen because it was thought to have the highest intensity (brightness). Matteson also varied the solicitation letter using a semi-personal and a form letter. Color did not significantly affect the response rate of those receiving the semi-personal letter (semi-personal letter recipients had significantly higher response rates than form letter recipients). But for those receiving the form letter and pink questionnaire, there was a significant increase in the response rate at the five percent level (from 20 to 24 percent). However, color alone did not significantly increase response rate.

In a survey of university alumni of a student business honorary association (N=168), Crittenden, Crittenden and Hawes (1985) tested the effect of yellow versus white questionnaires. They tested yellow versus white because yellow was a psychological primary color, the brightness contrast between print on white and yellow paper was minimal, and both white and yellow had traditionally yielded high returns. The yellow questionnaire obtained a higher response rate than the white questionnaire (49 versus 37 percent), but the difference was not statistically significant.

Pressley and Tullar (1977) surveyed marketing research directors from firms in four different industries (N=272). They tested white, green, blue and yellow questionnaires and conducted two mailings (the data were not presented by mailing). All of the color questionnaires were found to have lower response rates than white questionnaires. However, the differences were not statistically significant. In a later paper,

Pressley (1980) argued against the use of color paper with industrial organizations because of its association with junk mail.

Jobber (1986) hypothesized that a color questionnaire would be more conspicuous on a business person's desk, compared to white, which could be misplaced among other papers. He suggested that this would cause a color questionnaire to be completed some time after the respondent initially received it. Jobber and Sanderson (1983) tested blue and white questionnaires in a study of textile managers and directors (N=800). Two mailings were conducted. The blue questionnaire had higher response rates than the white for both mailings, but the difference was not statistically significant. Jobber and Sanderson suggest that inducement techniques which increase response rates for the general public may not do the same for commercial populations.

Fox, Crask and Kim (1988) conducted a metaanalysis of experimental studies testing factors that might influence mail survey response rates. As part of this metaanalysis, Fox and colleagues analyzed the effect of green versus white questionnaires. Ten experiments on color were included in the analysis (from Gullahorn and Gullahorn, Pressley and Tullar, and Pucel et al., above). Fox and colleagues found that green as opposed to white questionnaires increased mail survey response rates. The effect was significant at the five-percent level.

Two early experimental studies of direct-mail advertisements also tested the effects of color. Dunlap (1950) tested different colors of a renewal notice card sent to 572 members of an alumni association. Those receiving yellow cards had the highest return rate for the renewal, then blue, white and cherry cards, in that order. However, no significant differences were found due to the use of color. Bender (1957) hypothesized that colors appeal to the respondent's sensory and emotional faculties. He tested combinations of colors of stationery and envelopes to see if they increased the response to life insurance advertisements (N=8,600 officers and 3,200 enlisted men in the Navy and Air Force). The colors included: blue, green, canary, pink, goldenrod and white. Bender found that color combinations increased response rate overall, but the differences were not found to be significant, and several of the color combinations yielded fewer returns than white.

In summary, most of the studies show that color questionnaires have a positive effect on mail survey response rates. Yet only the metaanalysis (Fox et al., 1988) found color **significantly** increases response rate. Several reasons are offered for why color might affect the response rate. First, color is hypothesized to have a psychological affect, i.e., it could have emotional or sensory appeal to potential respondents, causing the color questionnaire to be returned more often. Second, in an office environment, color might be more noticed or misplaced less often than white papers.

3. Hypotheses

There are several hypotheses that can be generated for the OESWPS. First, if color is psychologically appealing to respondents, the response rates for an initial mailing of a

questionnaire should be higher; there should be a fairly immediate response. Second, if color questionnaires are more noticeable and less likely to be misplaced among office papers, response rates from the follow-up mailings may be higher. Third, if the response rate for color questionnaires is higher after a telephone call prompt, it may also support the less likely to be misplaced hypothesis.

4. Experimental Procedures

An experiment was conducted in fifteen states in conjunction with the 1990 OESWPS to test if the use of color questionnaires had an effect on survey response rates. The fifteen states included: Alabama, Colorado, Connecticut, Delaware, Florida, Georgia, Kentucky, Maine, Mississippi, Nebraska, North Carolina, Ohio, Pennsylvania, Texas, and Utah. Four industries were chosen for the OESWPS. They included: Advertising (SIC 731); Mailing, Reproduction, Commercial Art and Photography, and Stenographic Services (SIC 733); Services to Dwellings and Other Buildings (SIC 734); and Computer Programming, Data Processing, and Other Computer Related Services (SIC 737). Industries were selected from the 1990 OES sample with the conditions they would not overlap with other BLS wage surveys.

In the OESWPS, those units whose employment is 50 or more receive a "long" questionnaire and units whose employment is 49 or less receive a "short" questionnaire; the same practice is followed in the regular OES Survey. The long questionnaire contains more occupational categories than the short version.

The sample of firms receiving the long questionnaire was divided further in the OESWPS as an experimental test. Wage data for residual occupations--those occupations not falling into any predefined occupational category--were requested on half of the long questionnaires, while the other half did not request such information. We refer to these differing questionnaires as the "wage residual" (WR) and "non-wage residual" (NWR). This test was conducted to ascertain if requesting data on wages of residual occupations affected the survey response rate.

State samples were allocated by industry using a stratified sampling design. The sample establishments were divided into two groups: those who would receive a color questionnaire and those who would receive the white questionnaire. Green and yellow were selected as the colors to test as we felt they had the most consistent effects in the experimental literature. For administrative ease, green was used for the short and yellow for the long questionnaire.

Each type of questionnaire was given a code letter for identification. Questionnaires A-D were distributed across the four industries included in the survey. Questionnaires E through L were short questionnaires, paired by industry. The distribution of OESWPS questionnaires by type (long vs. short), "wage residual," color, and letter code is depicted in Table 1. While each pair of groups (i.e., A and B, C and D...K and L) were intended to be equal in size, some deviation was caused by multiple-unit establishments which were mailed the standard OES questionnaire to reduce confusion.

TABLE 1. Distribution of OESWPS Questionnaires

SIC	FORM	RESIDUAL	COLOR	LETTER
ALL	LONG	YES	WHITE	A
ALL	LONG	YES	YELLOW	B
ALL	LONG	NO	WHITE	C
ALL	LONG	NO	YELLOW	D
731	SHORT		WHITE	E
731	SHORT		GREEN	F
733	SHORT		WHITE	G
733	SHORT		GREEN	H
734	SHORT		WHITE	I
734	SHORT		GREEN	J
737	SHORT		WHITE	K
737	SHORT		GREEN	L

The participating states used a "mailout" code, designated as 1 through 4 on the data base to indicate the time period during which the questionnaire was received. Nonrespondents were assigned a code of zero. An initial and two follow-up mailings were conducted; then establishments were followed up by telephone. The periods were defined as follows:

- (1) Responses received before the first follow-up mailing - mainly from the initial mailing.
- (2) Responses received before the second follow-up mailing.
- (3) Responses received before telephone follow-up began.
- (4) Responses received after telephone follow-up.

The final data bases forwarded to BLS Washington were subjected to extensive editing to ensure that proper selection and coding procedures had been followed. Due to this editing, several States' data were eliminated from our data base, while several other States' data were restricted to inclusion in the testing for the final response rates. In the end, we used complete data from nine States, and partial data from three States. The response rate for the OESWPS color experiment was 78 percent, with a sample size of 3,213 (November 30, 1990 was the cut-off date for units to be included in the experiment; States continued to collect data after this date).

5. Tests and Results

In order to test the hypotheses, the response rates were computed separately for the initial mailing (mailout code 1), the follow-up mailings (mailout code 2 and 3), and the telephone follow-up (mailout code 4). Additionally, final response rates were also calculated. These data are shown in Table 2. Once we had computed the response rates, the data were analyzed by performing analysis of variance (ANOVA) and t-tests to determine if there were any significant differences. We present our results by questionnaire type because there are fundamental differences in the way that large establishments report as compared to small

establishments. An indicator of this is the short questionnaire response rate, which, in general, is much higher than long questionnaires. Additionally, we felt that our results might be affected by the fact that there are many more small establishments than large ones. A significant effect for long questionnaires could be negated by a majority of the short questionnaires.

Long Questionnaires

For the initial mailing, the white questionnaire received a better response in all industries as compared to the yellow questionnaires (see Table 2). This was also true for follow-up mailings in all industries except computer programming, etc. (SIC 737). This could support Pressley's contention that establishments associate color questionnaires with junk mail, however, the results of the ANOVA for the initial and follow-up mailings showed that neither color, wage residuals nor SIC had a significant effect on the response rate for this collection period.

For the telephone follow-up, the white questionnaire again had a slightly higher (3.3%) response rate than yellow questionnaires. When data were combined for all industries, the WR questionnaire, combined with the yellow questionnaire, did surprisingly better during this collection period (+19.9%; see Table 2) than the NWR questionnaire.

The results of the ANOVA for telephone follow-up show that there is a significant difference (P-Val = 0.0001) in response rates for this period. The primary contributor to this difference is the SIC. However, the interaction between color and SIC is also significant (P-Val = 0.0245). The t-tests conducted showed that in advertising (SIC 731) yellow questionnaires had a significantly higher response than white questionnaires. This lends some support to the hypothesis that color questionnaires are more easily identifiable in an office setting.

The final response rates for the long questionnaires showed an extremely small difference between yellow and white colors, with white questionnaires obtaining the higher response rate (+1.7%; see Table 2). At the industry level results were mixed, with yellow questionnaires having the higher response in industries 731 and 737--advertising, and computer programming, data processing and other computer-related services, and white questionnaires receiving a better response in industry 733--mailing, reproduction, commercial art and photography, and stenographic services and 734--services to dwellings and other buildings.

In all industries the long yellow questionnaire with the wage residual (WR) category had a higher response rate than the questionnaire without it (NWR). For the white questionnaires, three out of four industries had a higher or the same response rate for the WR compared to the NWR questionnaire. Combining these gives a slight edge to the total WR questionnaire, although the differences are not statistically significant. The yellow questionnaires had a significantly higher response rate for the WR (+9.6%; P-Val 0.0358) compared to the NWR questionnaire.

The ANOVA results show that there is a significant difference (P-Val = 0.0263) in response rates for the long questionnaires for the final response rate. However, this is

not due to color or wage residuals, but the difference in response among the different industries.

Short Questionnaires

The initial response for the short questionnaires was mixed. Advertising (SIC 731) and mailing, reproduction, etc. (SIC 733) showed a higher response to white questionnaires, while the other two industries showed a higher response to green questionnaires. Overall for this period the white questionnaires performed slightly better. Once again, this may indicate that color has no psychological effect on respondents and support Pressley's observation that establishments associate color questionnaires with junk mail. ANOVA results for this collection period indicate that the only factor significantly affecting response rates was SIC ($P\text{-Val} = .0001$).

Our expectation for the follow-up mail collection period (as in all others) was that color would raise the level of response, but that the effect would not be statistically significant. The response rate for the green questionnaires during this follow-up period was higher (from 0.6% to 9.9%, depending on industry) than the response to the white questionnaires for all industries tested. However, for industry 733--mailing, reproduction, commercial art and photography, and stenographic services--and industry overall, t-tests showed the difference between green and white questionnaires was statistically significant at the five percent level, respectively ($P\text{-Val} = .0366$, $.0214$). The results of the ANOVA for this period designates color as the only component significantly affecting response rates ($P\text{-Val} = .0214$). Thus, color, as opposed to white questionnaires may be more memorable to respondents, causing them to be more likely to complete them.

Over the telephone follow-up period, green questionnaires once again had a higher response than white questionnaires. This was true for three of the four industries tested, with SIC 733 getting a better response to white questionnaires (as opposed to the follow-up mailing). ANOVA results for this period show SIC ($P\text{-Val} = 0.0001$) as the only factor contributing to the difference in response rates.

The final response rates of all four industries combined for the short questionnaires showed that the green questionnaires had a 3.5 percent higher response than the white questionnaires (see Table 2). This difference was statistically significant. All industries showed a higher level of final response for the green questionnaire. The ANOVA results indicated that color played a significant ($P\text{-Val} = 0.0287$) role in the difference between response rates, as did SIC ($P\text{-Val} = 0.0005$).

6. Conclusions

The initial response, which was not favorable to color questionnaires, tends to cast doubt on the hypothesis that respondents find color psychologically appealing. The response to the follow-up mailings of the short questionnaires supports the hypothesis that color questionnaires are more noticeable and less likely to be misplaced among office papers. However, the response to the long questionnaires for

this period is contrary to that hypothesis. It may be that large establishments, which are sampled more often, have familiarity with white survey questionnaires and associate color questionnaires with junk mail. Also, States have existing relationships with large companies, which may outweigh the effects of any other factors. A third explanation is that yellow, as opposed to green questionnaires, may not be as noticeable or memorable to respondents.

The response to the telephone follow-up tends to support the same conclusions reached for the follow-up mailings. Even though the results were not significant, the level of response was in the same direction. The only additional conclusion for this stage of collection is that it was beneficial to include the wage residual category on the long questionnaires.

We conclude that at least for short questionnaires in both the follow-up mailings and telephone follow-up, green questionnaires did produce a higher response rate than white questionnaires. This, in turn, resulted in a final response rate that was higher than would have been realized had color questionnaires not been utilized. Yellow questionnaires did not produce a higher response rate among long questionnaire recipients, although at the time of telephone follow-up they had a significant effect in one industry. It may be that the yellow, as opposed to the green questionnaire was less visible on a desk, however there are too many confounding factors associated with large establishments for us to make that conclusion.

Also, contrary to expectation, obtaining wage data for residual occupation categories (wage residual) resulted in an increase in response rates (not significant). Since this category is critical to estimating industry wide distribution of wages, and since it does add to response, we can recommend that this category continue to be included on the long questionnaire.

We would like to conduct a further test for the smaller establishments or a survey which uses a single questionnaire for all establishments. At the time of initial and follow-up mailings, one group would receive all white (1) and another all green (2) questionnaires. Two other experimental groups would receive green for the initial and white for the follow-up mailings (3), and vice versa (4). We would expect groups 3 and 4 to have the lowest response rate, as receiving varying colors should make the first survey questionnaire less easy to recall. Group 2 should have the highest response rate as green questionnaires would have the highest probability of being noticed or remembered. The group receiving all white questionnaires should have the second highest response rate, for reasons similar to the green, except that we believe white questionnaires are not quite as noticeable or memorable as green questionnaires.

Bibliography

Bender, Donald H. (1957) "Colored Stationery in Direct-Mail Advertising." *Journal of Applied Psychology* 41 (3): 161-164.

Crittenden, William F., Vicky L. Crittenden and Jon M. Hawes. (1985). "Examining the Effects of Questionnaire Color and Print Font on Mail Survey Response Rates." *Akron Business and Economic Review* 16: 51-56.

Dunlap, J. William. (1950) "The Effect of Color in Direct Mail Advertising." *Journal of Applied Psychology* 34: 280-281.

Fox, Richard J., Melvin R. Crask and Jonghoon Kim. (1988). "Mail Survey Response Rate: A Meta-Analysis of Selected Techniques for Inducing Response." *Public Opinion Quarterly* 52: 467-491.

Gullahorn, Jeanne E. and John T. Gullahorn. (1963) "An investigation of the effects of three factors on response to mail questionnaires." *Public Opinion Quarterly* 27: 294-296.

Jobber, David. (1986). "Improving Response Rates in Industrial Mail Surveys." *Industrial Marketing Management* 15: 183-195.

Jobber, David and S.M. Sanderson. (1983) "The Effects of a Prior Letter and Coloured Questionnaires on Mail Survey Response Rates." *Journal of Market Research Society* 25: 339-349.

Matteson, Michael T. (1974) "Type of Transmittal Letter and Questionnaire Color as Two Variables Influencing Response Rates in a Mail Survey." *Journal of Applied Psychology* 59: 535-536.

Pressley, Milton M. (1980). "Improving Mail Survey Responses from Industrial Organizations." *Industrial Marketing Management* 9: 231-235.

Pressley, Milton M. and William L. Tullar. (1977). "A Factor Interactive Investigation of Mail Survey Response Rates from a Commercial Population." *Journal of Marketing Research* 14: 108-111.

Pucel, David J., Howard F. Nelson and David N. Wheeler. (1971) "Questionnaire Follow-Up Returns as a Function of Incentives and Responder Characteristics." *Vocational Guidance Quarterly* 19: 188-93.