Potential versus actual faking of a biodata form: An analysis along several dimensions of item type

Article *in* Personnel Psychology · December 2006

DOI: 10.1111/j.1744-6570.1992.tb00855.x

CITATIONS
85

READS
488

2 authors:

Thomas E. Becker
University of South Florida
52 PUBLICATIONS **11,076** CITATIONS

SEE PROFILE

Alan Colquitt
Eli Lilly
4 PUBLICATIONS **392** CITATIONS

SEE PROFILE

PERSONNEL PSYCHOLOGY
1992, 45

POTENTIAL VERSUS ACTUAL FAKING OF A BIODATA FORM: AN ANALYSIS ALONG SEVERAL DIMENSIONS OF ITEM TYPE

THOMAS E. BECKER
Washington State University

ALAN L. COLQUITT
Eli Lilly and Company

Two studies examined faking of a 25-item biodata questionnaire. The first study investigated potential and actual faking of the form using three groups: a group told to make themselves look as good as possible, a group told to complete the form honestly, and a group completing the instrument in a real selection situation. Subjects were 58 current employees and 231 job applicants. Results indicated that subjects could fake the instrument when instructed to do so. Also, some faking appeared to be occurring in practice, although results depended upon the composition of the comparison group. Only eight items appeared to be fakable, and only three of these seemed to be faked in practice. In Study 2, 26 business majors rated the biodata items on eight dimensions of item type. Results showed that the three items faked in practice were *less* historical, objective, discrete, verifiable, and external than other items, and were *more* job relevant.

Much of the past research on the faking and distortion of tests has been directed toward personality and interest inventories. Some of these studies instruct one group of subjects to answer the test items in such a way as to make themselves look as good as possible, and instruct another group to respond to the items honestly. Such investigations address the issue of whether or not the test *can be faked* (that is, the fakability of the test). For example, it has been found that, when told to do so, people can improve their standing on the Gordon Personal Preference Profile (Braun, 1965), the Edwards Personal Preference Schedule (Borislow, 1958), and the Strong Vocational Interest Blank (Abrahams, Neuman, & Githens, 1968). Other studies make use of real or simulated testing conditions, with the comparison of interest being between subjects in these conditions and subjects told to be honest. These studies address the issue of whether or not the test *is faked in practice*. For example, it

We would like to thank John Callender, Adrienne Colella, Avraham Kluger, Garnett Stokes Shaffer, and this journal's reviewers for comments on earlier versions of this article.

Correspondence and requests for reprints should be addressed to Thomas E. Becker, Department of Management and Systems, College of Business and Economics, Washington State University, Pullman, WA 99164-4726.

has been demonstrated that some degree of actual faking occurs on the Gordon Personal Preference Profile (Bass, 1957) and the Strong Vocational Interest Blank (Kirchner, 1961). Recent work has continued this trend of investigating faking of personality inventories (Hough, Eaton, Dunnette, Kamp, & McCloy, 1990).

The research on the faking of biodata instruments is scanty. Early work on the accuracy of work histories (as assessed by interviews or application blanks) suggested that job applicants do not substantially distort their reports (Keating, Paterson, & Stone, 1950; Mozel & Cozan, 1952). However, this work examined only a very narrow range of applicant histories (i.e., past wages, length of previous employment, and prior job duties). Items included on contemporary biodata forms encompass a far broader domain. More recent evidence suggests that subjects can improve their biodata scores when instructed to do so (Klein & Owens, 1965; Schrader & Osburn, 1977).

The evidence for actual faking of biodata forms, however, is mixed. Some work has reported considerable disagreement between reported and verified information (Goldstein, 1971; Weiss & Dawis, 1960), while other work has found that biodata responses are reasonably accurate (Cascio, 1975; Shaffer, Saunders, & Owens, 1986). This apparent discrepancy in research findings is probably due to differences in the types of biodata items examined. As several typologies of biodata have suggested, biodata items vary along continua of verifiability, objectivity, job relevance, and so on (Asher, 1972; Cascio, 1982; Mael, 1991). Thus, for example, items which are more objective and verifiable (e.g., age when first married, number of jobs held prior to present employment) might be expected to be less amenable to faking than those which are less objective and verifiable (e.g., questions dealing with attitudes, interests, preferences, or opinions). Unfortunately, research to date has not adequately addressed either the issue of which types of items *can* be faked, nor the more important issue of which types *are* faked in practice.

Assuming that at least some types of biodata are fakable, faked in practice, or both, contemporary researchers have studied ways to reduce socially desirable responding. Empirical keying of scoring keys appears to reduce the influence of socially desirable responding on applicant scores (Haymaker, 1986); option keying (i.e., separately evaluating the validity of each item alternative) seems to reduce such influences even further (Kluger, Reilly, & Russell, 1991). Warning respondents that answers will be verified also appears to reduce faking (Hough et al., 1990; Schrader & Osburn, 1977).

Apart from the paucity of research on biodata faking, there are other limitations to the information currently available. One limitation is that much of the prior work has used college students as subjects. This group

is unlikely to adequately represent the population of individuals in the work force or labor market. A more serious limitation is that no single study has examined both the extent to which a given biodata instrument can be faked, and the extent to which the instrument is faked. This is important because the former probably sets the limits for the latter. In general, all else being equal, the lower the fakability of the instrument the lower will be faking in practice. A final limitation of available research is that the implications of actual faking for selection decisions have not been studied. Given that, theoretically, any amount of faking introduces measurement error, it remains an open question as to whether or not the extent of such error results in false positives and false negatives.

The contributions of the current investigation to the extant literature are (a) real employees and job applicants serve as participants, (b) the concern for potential versus actual faking of biodata is addressed, (c) the issue of the implications of actual faking for selection decisions is considered, and (d) the susceptibility of different types of biodata to faking is examined.

Based on the research discussed above we formulated the following hypotheses:

Hypothesis 1: Our biodata form will be fakable.

Hypothesis 2: Some degree of faking will occur in practice.

Further, since we assumed that at least some job applicants are more honest than they have to be, we hypothesized that

Hypothesis 3: Faking in practice will be less than potential faking.

Given the lack of relevant theory and research, we treated the issue of the effect of faking on selection decisions as an exploratory concern.

Finally, we used Mael's (1991) typology and discussion to make predictions about which types of biodata are likely to be faked in practice. Mael (1991) specified ten dimensions for classifying biodata items, seven of which are relevant to this study. These are: history (i.e., the extent to which items refer to events that have taken place or continue to take place), objectivity (i.e., the extent to which items require recall but not interpretation), first-handedness (i.e., the extent to which items ask about personal, direct observation rather than estimation of others' perceptions), discreteness (i.e., the extent to which items ask about a single, unique event or count of unique events rather than a summary of responses), verifiability (i.e., the extent to which items ask about an event that can be corroborated from an independent source), job relevance (i.e., the extent to which items ask about events which are clearly linked

to the job), and externality (i.e., the extent to which items ask about observable events). Further description and discussion of these dimensions are presented in Mael (1991). Mael's other dimensions (controllability, accessibility, and invasiveness), while important to ethical and legal issues surrounding biodata, were not directly relevant to the faking issue, so hypotheses regarding these dimensions were not made.

Following the reasoning presented in Mael (1991), we predicted that

> *Hypothesis 4*: Items faked in practice will be *less* historical, objective, firsthand, discrete, verifiable, and external than items not faked in practice. Further, items that are faked in practice will be *more* job relevant than other items.

The rationale is that actual faking (intentional or unintentional) will occur only on those items for which the truth is less knowable. As has been pointed out by others, items that are visibly job relevant are more likely to be transparent and fakable than other items (Mael, 1991; Mumford & Owens, 1987). We would add that applicants would also be more motivated to fake such face-valid items than to fake items which appear less pertinent to the job.

Our last prediction was that

> *Hypothesis 5*: Items faked in practice will be more susceptible to socially desirable responding than items not faked in practice.

Socially desirable responding has been defined as presenting oneself favorably regarding current social norms and standards (Zerbe & Paulhus, 1987). An obvious motive for job applicants to fake their responses to biodata questions is to make themselves look better than they are so that they will have an increased chance of getting the job. One way applicants might accomplish this is to respond to biodata items in a socially desirable manner. The logic for our last hypothesis, then, is that items which appear to be more susceptible to socially desirable responding will be more amenable to faking because respondents recognize the greater opportunity to make themselves look good on such items.

We conducted two studies to explore the above issues. The first study addressed the general fakability and actual faking of a biodata form and the effect of faking on selection decisions. The second investigations examined the relationship between Mael's (1991) dimensions of biodata and the extent of faking in practice.

Study 1

Method

Subjects

Two sets of subjects were involved in this study. One set included current employees from one division of a large international goods-producing company. Employees from this particular division were chosen because the division used the biodata form extensively and because earlier studies had validated the form for this division. The pool of potential subjects was defined as those employees who worked in this division but were hired before the biodata form was used as part of the selection process. From this pool, 141 employees in the division were solicited, and 58 (41%) agreed to participate. The average tenure for these people was 31 months. During pilot study discussions, several employees from the division suggested that, due to the potential for identification of participants, gathering additional demographic data (besides tenure) might greatly lower response rates or affect the quality of responses. Given these concerns, and given prior work within the company demonstrating that scores on the form are not related to variables such as gender and ethnic group, we opted to forego collection of demographic information.

All 58 employees had previously been screened under the company's complete selection process, *excluding* the biodata form used in the present study. Restricting this set of participants to those who had not completed the biodata form was done in order to obviate practice effects. The selection system involved five steps: (1) screening of resumes, (2) (for those passing Step 1) a 45-minute screening interview, (3) (for those passing the first interview) completion of an ability test, followed by a second interview, (4) (for those passing Step 3) a series of four to seven interviews with select personnel and line management employees working in the division for which the participants had applied, and (5) (for those passing Step 4) completion of a physical exam. The second set of subjects contained 231 job applicants, 50 of whom were later hired. These people were all applying for the same job within the division from which the first set of subjects were selected. They had been screened through the company's selection system, including the biodata form (administered during Step 3 above). Note that the only difference between applicants later hired and current employees in terms of screening involved completion of the biodata form by applicants. Note also that *all* applicants in this study survived at least the first two steps of the selection process described above.

The Biodata Form

The biodata form used in this study was developed using a sample of 430 exempt employees from four divisions (including the one included in this study) of the company; employees were matched on length of service, race, and gender. Initial items were mailed to these employees via the company mail system; instructions informed the participants that their responses would be used for research purposes only. Responses were keyed against an overall performance criterion. An empirical keying strategy was used whereby items were retained that discriminated significantly between high and low performers. The original key was then cross-validated with a second sample of employees; the average validity coefficient for this sample was .39, with a range from .29 to .49 for various divisions included in the sample. Test-retest reliability for the instrument was .79 (time intervals between the administrations ranged from 2 to 6 months). Correlations of biodata scores with scores on the ability test and ratings in the interviews ranged from .15 to .20.

The final version of the biodata form contained 25 multiple choice questions. For descriptive purposes, these items may be rationally grouped into three categories: school activities, ability and motivation, and personal interests. These content domains are similar to those widely used in biodata investigations (e.g., Childs & Klimoski, 1986; Mumford & Owens, 1987), and fall within the general requirement that biodata items reflect a current or past part of the person's life history. Our items were not intended to measure personality. As Mael (1991) pointed out, personality measures attempt to assess a respondent's disposition, often by including behavioral intent items or responses to hypothetical situations. In contrast, the biodata form used in the current research attempted to assess specific abilities, motivations, and interests using items focused on behaviors and events that actually take place or have taken place. The following examples illustrate the types of items and kinds of response options included on the form.

1. School Activities: What kinds of activities were you involved in during high school? Athletics? Social clubs? Honor societies?

2. Ability and Motivation: How much energy do you typically have? Not much? Some? A lot?

3. Interests: What do you like to do in your free time? Go to a movie? Work on a hobby?

Possible scores for the entire instrument ranged from −6 to 36. Possible scores on individual items ranged from −1 to 3. Each item was scored assuming a meaningful underlying continuum (i.e., no items were scored ipsatively). Based on concerns regarding the proprietary nature of the biodata form (it is now copyrighted), the company has reserved the right

to maintain confidentiality regarding additional content and format of the form.

Procedure

Potential and actual faking. Assessing potential and actual faking involved comparisons of subjects participating in three conditions, as follows. In the *straight-take condition*, subjects were given the following instructions:

> Attached to this sheet is a biographical information questionnaire. We would like your straightforward, honest answers to these 25 questions. Your responses will be strictly anonymous, and will not be used to evaluate you in any way, so please provide as accurate answers as you can.

In the *fake-take condition*, subjects were given these instructions:

> Attached to this sheet is a biographical information questionnaire. We would like you to answer these 25 questions in such a way as to make yourself look as good as possible. Imagine that you are an applicant for a desirable position within this company, and that you really want the job. Complete this questionnaire in the way that you think would most likely get you the position. Do not worry about providing accurate responses.

Finally, subjects in the *real-take condition* were told:

> The purpose of this form is to help us learn more about your job-related experiences, work interests, personal achievements, and other data to supplement what you have been able to tell us so far. It is to your advantage to answer every question candidly and accurately. Information you provide in this form may be verified with other data collected during the employment process.

Current employees were informed that the study was being conducted to examine the usefulness of the biodata form for selecting future employees. Employees were also told that they were to complete the instrument as if they were applying for the entry-level job for which they had been hired (all employees had been hired for the same entry-level job).

Employees were randomly assigned to either the fake- or straight-take conditions. All job applicants were in the real-take condition. The real-take condition served as a basis for forming two comparison groups. The first group contained only those 50 applicants who were subsequently hired. The second group contained all 231 applicants, including those who were later hired. The first group contained subjects who were

most comparable to the employees in the straight- and fake-take conditions. Recall that this group, like the current employees, had survived the extensive, multi-staged selection system, with the only difference being that the members of this comparison group had completed the biodata form as applicants while current employees had not. Also, recall that scores on the biodata instrument are not related to gender or ethnic group, and only very weakly related to scores on the cognitive ability test. Since this group contained people who were judged of sufficient quality to be hired, the subjects in this group were likely to be similar to other employees in terms of job-related ability and aptitude. While this is certainly desirable from a design standpoint, the trade-off is that this group could possibly have a restricted range on the biodata scores and, therefore, be less likely to be representative of the applicant pool. The advantage of the second comparison group is that the restriction of range is not an issue. Given that there were pros and cons of each operationalization, both groups were included in the analyses.

Impact of faking on selection decisions. To explore the effect of actual faking on selection decisions, it was assumed that people scoring at or above the median would be accepted (hired) while people scoring below the median would be rejected. The impact of faking on decisions was then evaluated by comparing accept and reject decisions based upon (a) straight-take and real-take scores, and (b) real-take scores and scores adjusted for faking. The ways in which scores were adjusted will be described shortly.

Results

Table 1 presents the sample sizes, means, standard deviations, and range of biodata scores for the four groups. Table 2 presents the results of a one-way ANOVA and orthogonal planned comparisons carried out to test our first two hypotheses. Note that the real-take group in this analysis was the control group including hires only. As can be seen, the omnibus F test indicates significant differences among the three conditions. Also, the findings in the table indicate that the mean of the fake-take group is significantly higher than the combined mean of the straight-take and real-take groups. This supports our first hypothesis that the form is fakable.

With respect to the second hypothesis, results reported in Table 2 indicate that the mean of the real-take group (hires only) was higher than the mean of the straight-take group. However, when the second comparison group was used in the comparison, the mean of the real-take group (applicant pool) was not higher than that of the straight-take group; in fact, as shown in Table 1, the mean of the applicant pool scores

TABLE 1
Descriptive Statistics

Group	*n*	*M*	*SD*	Range
Straight-take	29	17.71	5.42	9–26
Fake-take	29	23.41	5.62	11–31
Real-take (hires)	50	20.84	4.29	13–31
Real-take (applicants)	231	17.25	4.59	3–31

TABLE 2
Results of ANOVA and Planned Comparisons

Source	*df*	*SS*	*MS*	*F*
Group	2	471.37	235.68	9.50**
Error	105	2,605.55	24.81	
Contrasts:				
Fake vs. straight & real[a]	1	354.88	354.88	204.49**
Straight vs. real	1	178.20	178.20	51.55*

[a]Real-take subjects were applicants who were later hired.
*$p<.01$; **$p<.001$

was slightly less than that of the straight-take group. Thus, the findings are mixed regarding this hypothesis.

Since the way in which the biodata form is actually used to select employees is proprietary information, and since different divisions within the company combine the biodata information with other data gathered in the selection process differently, we could not directly examine or report the affect of biodata faking on actual selection decisions. Instead, to estimate the expected proportions of accepts and rejects for the applicant pool, the proportions of accepts and rejects in the straight-take group were used. Using the straight-take median as a cutoff (recall that those at or above the median were defined as acceptable), about 55% (16/29) of the straight-take group would have been accepted and 45% would have been rejected. Therefore, it was assumed, the "true" biodata scores of applicants would have resulted in the same proportion for this group; i.e., using the straight-take median as a cutoff for the applicant pool scores, about 55% (127/231) of the applicants were expected to be accepted. In fact, the observed proportion of applicants accepted was 52.8% (122/231), while 47.2% were rejected. The difference between the observed and expected frequencies was not significant, χ^2 (1) = .44. This indicates that the observed proportions of accepts and rejects among applicants did not differ significantly from the proportions expected under honest (i.e., straight-take) conditions.

To explore the effect of faking on selection decisions in a worst case scenario, it was assumed that every applicant distorted his or her score. Specifically, an adjusted distribution of scores was produced by subtracting three points from everyone's original score; recall that three points was the mean difference (rounded) between the straight- and real-take scores. This adjusted distribution was then compared with the observed scores of the applicants and the accept/reject decisions based on the straight-take median cutoff. Adjusted scores were designated "hits" when the decision category did not change following the adjustment. A "miss" was designated when the adjusted score indicated a reject decision while the original score indicated an accept decision.

Of course, 100% of the accept decisions based on adjusted scores (n = 70) were also accepts based on the original scores. More relevantly, 75.8% (n = 122) of those rejected based on the adjusted scores were also rejected based on their original scores. Overall, 83.1% (n = 192) of the observations were hits. Next, we turn to the issue of biodata faking as it related to item type.

Study 2

Method

Subjects

Twenty-six business majors (juniors and seniors) enrolled in an upper-level human resource management course at a northwestern university participated in this study. Participation was offered as an alternative to a short paper assignment, and class credit was given for participation.

Rating Form and Procedure

A rating form for assessing the 25 items along seven of Mael's (1991) dimensions was developed. One page per dimension was provided, with the subjects being instructed to rate all 25 items on the dimension before going on to rate each item on the following dimension. At the top of each page a definition of the relevant dimension was given in question form; for example, on page one "historical" was defined as "Does the item refer to events that have taken place or continue to take place?" Subjects were then given a stimulus question for the dimension which encouraged the subjects to think of the dimension along a continuum (e.g., "To what extent does the item refer to one or more historical events?" "To what extent does the item refer to objective events?"), and were then asked to rate the items using a 5-point Likert scale where 1 = not at all, 2 = very

small extent, 3 = moderate extent, 4 = great extent, and 5 = completely. The dimensions included on the form included history, objectivity, first-handedness, discreteness, verifiability, job relevance, and externality. In addition, we included a dimension called social desirability, which was defined as "Can respondents [i.e., job applicants] respond to the item in such a way as to make themselves look better than, in fact, they are?"

Subjects made their ratings outside of class. They were given both oral and written instructions as to how to complete their ratings. These instructions emphasized finding a quiet place to work, completing the ratings at one sitting, working alone, and taking great care in considering the ratings. Subjects were also told that their responses would be checked for carelessness and that no credit would be given to students who had not made their ratings carefully. Students reported later that the task took them about one and a half hours to complete.

Results

Data from Study 1 were used to identify which items were fakable and which appeared to be faked in practice. Twenty-five ANOVA's were run to compare the fake-take, straight-take, and real-take (hires only) scores for each item. To ensure that familywise error did not exceed .20, the *p*-value for each ANOVA (i.e., alpha per comparison) was set at .01, and paired-comparisons were conducted only if the omnibus test was significant. Due to the large number of comparisons being conducted, the alpha for each *t* test was also set at .01.

Table 3 shows the results of these analyses for the 8 (out of 25) items having significant ANOVAs. The table includes mean scores and standard deviations for each item for each condition, and the *F*-values from the relevant ANOVAs. The table also contains the results of the paired comparisons, showing which means are significantly different from means in other conditions. As can be seen, the fake-take mean is significantly greater than the straight-take mean for all but one of the items, indicating that these items are fakable. However, the real-take mean is significantly higher than the straight-take mean for only three of the items (i.e., items six, seven, and eight); these three items are the only ones which appear to be significantly faked in practice.

Table 4 contains the intercorrelations among the eight dimensions of item type. This table also reports the intraclass correlation for each dimension as a measure of interrater reliability. In examining the differences between the three items that appear to be faked in practice and the 22 items that do not, we conducted *t* tests for dependent samples for each of the dimensions. This involved (a) computing the mean across the three faked items for each dimension,(b) computing the mean across the

TABLE 3

Results of Significant ANOVAs and t Tests

Item	*M*			*SD*			*F* values
	Straight	Fake	Real	Straight	Fake	Real	
1.	$.69^{a}$	$.97^{b}$	$.78^{ab}$	.22	.03	.18	3.88
2.	$.17^{a}$	$.66^{b}$	$.32^{a}$	.15	.23	.22	8.80
3.	1.38^{ab}	2.03^{a}	1.34^{b}	1.10	1.03	1.17	4.41
4.	$.79^{a}$	1.41^{b}	$.90^{a}$	.74	.54	.46	6.03
5.	$-.21^{a}$	0.00^{b}	$-.08^{ab}$	.17	0.00	.08	3.95
6.	$.62^{a}$	1.00^{ab}	1.22^{b}	.53	.93	.50	5.29
7.	$.24^{a}$	$.72^{b}$	$.60^{b}$	.19	.21	.24	8.52
8.	$.55^{a}$	$.97^{b}$	$.90^{b}$	.26	.03	.09	12.54

Note: Means having the same superscript are not significantly different at $p<.01$. All F values are significant at $p<.01$.

TABLE 4

Intercorrelations Among Dimensions of Item Type

Dimension	1	2	3	4	5	6	7	8
Historical (1)	(.95)							
Objective (2)	−.18	(.95)						
First-hand (3)	.07	.38	(.83)					
Discrete (4)	.03	.27	.40	(.92)				
Verifiable (5)	−.02	−.02	−.01	.39	(.98)			
Controllable (6)	−.33	.20	.00	.20	.20	(.94)		
External (7)	−.03	.42	−.09	.20	.35	.04	(.96)	
Social desirability (8)	−.01	.30	.14	.35	.33	.06	.56	(.51)

Note: Sample size is 26, correlations greater than .40 are significant at $p<.05$, and numbers in parentheses are intraclass correlations.

22 nonfaked items for each dimension, and (c) testing the significance of mean differences. To control for familywise error, we set alpha per comparison at .01. The results are shown in Table 5. As can be seen, the faked items differed from the nonfaked items in that the faked items were *less* historical, objective, discrete, verifiable, and external. The faked items were also *more* job relevant than the nonfaked items. These analyses support our fourth hypothesis. Contrary to this hypothesis, however, faked items were not significantly less first-hand than nonfaked items. Interestingly, the faked items did not differ from the nonfaked items regarding perceived susceptibility to socially desirable responding.[1]

[1]The above analyses were re-run using all eight fakable items rather than just the three items faked in practice. The results were nearly identical.

TABLE 5

Mean Differences Between Items Faked in Practice and Items Not Faked in Practice

Dimension	M_F	M_{NF}	M_{diff}	*SE*	*t*
Historical	2.29	3.32	1.03	.17	6.15*
Objective	2.60	3.41	.81	.18	4.54*
First-hand	4.29	4.13	.17	.14	−1.16
Discrete	2.37	3.14	.77	.16	4.70*
Verifiable	2.54	3.32	.77	.09	8.33*
Job relevance	3.68	2.95	.73	.12	−6.23*
External	2.65	3.15	.50	.13	3.73*
Social desirability	3.95	3.94	−.01	−.08	.94

Note: M_F is the mean rating on the dimension across the three items faked in practice. M_{NF} is the mean rating on the dimension across the 22 items not faked in practice. M_{diff} is the average difference across these means, and *SE* is the standard error of the difference.
*$p < .001$

Discussion

Consistent with past research and our first hypothesis, this study indicates that biodata forms can be faked. Subjects in the fake-take condition scored an average of 5.7 points higher than those in the straight-take group and 2.6 points higher than applicants who were later hired. This information, however, is not very enlightening per se but, rather, is useful as a reference point for assessing the extent of faking in practice.

The findings related to the second hypothesis were mixed. We considered it prudent to include both the entire applicant pool and only those applicants who were later hired in our analyses. When only applicants later hired were used as a comparison, it appeared that some faking was occurring in practice; this supports the second hypothesis. Also, congruent with the third hypothesis, the amount of actual faking appeared to be less than the amount of potential faking; in other words, the difference between the real- and straight-take scores was less than the difference between the fake- and straight-take scores. However, when the entire applicant pool was used as a comparison, there appeared to be no difference in real- and straight-take scores. This contravenes the second hypothesis but, like the former analysis, is consistent with the notion that less faking occurs in practice than is theoretically possible (the third hypothesis).

This research corroborates past research that more objective and verifiable items are less amenable to faking. However, it also goes beyond prior studies in demonstrating that items which are faked in practice also tend to be (a) less historical, discrete, and external, and (b) more job relevant. Further, the current investigation suggests that one can not tell if

an item is likely to be faked by looking at it; faked items do not necessarily differ from nonfaked items in terms of their perceived susceptibility to socially desirable responding. It should be noted, however, that there was relatively low interrater agreement on the social desirability dimension and that this could explain why the predicted effect was not found.

Future work needs to investigate further exactly why some of the dimensions are related to faking; this would advance further our knowledge of the faking process. We note here, however, that our analysis does not suggest that people are necessarily consciously lying when they provide more inflated responses in a real-take condition than they may have under straight-take conditions. Rather, when provided with a motive to make themselves look good (as in an applicant setting), people may unintentionally alter their perceptions of themselves in order to avoid cognitive dissonance between their perceptions and their responses to the biodata items. This argument is consistent with Zerbe and Paulus's (1987) distinction between impression management and self-deception. Thus, we would suggest, it is easier for applicants to alter their perceptions (deceive themselves) when the items are less historical, discrete, and external because the greater ambiguity inherent in such items allows greater latitude in interpretation. Further, respondents may be more motivated toward self-deception when items are visibly job relevant because the implications of responses to such items are more immediately apparent and direct than responses to less visibly job relevant items.

It is interesting to note that neither the fakable items nor the items faked in practice fell neatly into a single content category. That is, these items did not appear to measure affect or temperament to a greater extent than did other items, nor did they seem to be disproportionately associated with one of our content classes (school activities, ability and motivation, or interests). It is, of course, possible that items attempting to assess attitudes, feelings, and value judgments are more susceptible to faking than items measuring more concrete constructs. This may be because temperament items are typically less objective, verifiable, historical, and so on, than more behavioral items. Since the instrument used in our work did not include items explicitly intended to measure temperament or personality, we are unable to draw firm conclusions on this topic. Later work should explore this issue in depth.

The item-level information should prove useful to practitioners involved in the development and assessment of biodata items. Specifically, the current evidence is that (if the goal is to reduce faking) items should be written which fall relatively high on continua of history, objectivity, discreteness, verifiability, and externality. Pilot work of new biodata items could identify items high on these dimensions; all else being equal, these items would be preferable to items scoring lower on

the dimensions. A potentially difficult issue is raised by our finding that items faked in practice tend to be more visibly job relevant than items not faked. Given our other results, we would expect that visibly job relevant items which are high on the other dimensions (objectivity, verifiability, and so on) would be less likely to be faked than visibly job relevant items low on these dimensions; in fact, perhaps the effect of visible job relevance may disappear under these conditions. Future work should examine this issue carefully.

Taken together, the exploratory analyses suggests that there was only a minimal effect of actual faking on selection decisions. The proportion of applicants accepted and rejected was not significantly different than the proportion that would be expected if everyone were answering honestly (as they presumably were in the straight-take condition). Further, even assuming that all applicants did distort their scores in an upward direction (by an amount equal to the average difference between the average straight-take and applicant pool scores), over 80% of the selection decisions would be unaffected.

It would be premature, however, to conclude that faking is irrelevant to selection decisions. In our worst-case scenario, we assumed that all applicants were faking to the same extent. While this was a useful heuristic for exploratory analyses, future work should investigate the variability in faking of job applicants. This is important because to the extent that applicants with lower true biodata scores fake more than those with higher scores, our findings underestimate the effect of faking on false accepts and rejects. Future work must also carefully examine whether or not there is an effect of actual faking on the criterion validity of biodata instruments; if an effect is found, its magnitude and direction should be determined. Given our results, we would predict that attenuation of validity, if it occurs at all, will be minimal.

This study, like all others, has its limitations. First, the between-subjects design and issues of sampling raise several concerns. We opted for a between-subjects approach because our concern was more with the nature of the instrument than the nature of the participants. Still, it is probably true that there are individual differences in faking of biodata items and that these differences may have ramifications for the use of biodata forms. Future research could address such issues by using a within-subjects design. Ideally, this would involve applicants (including both those hired and those not hired) taking the instrument under real-, fake-, and straight-take conditions. Such a design would also probably avoid the problems we had regarding the second hypothesis. One possible problem of the within-subject approach, however, is that contrast effects may result from subjects serving in multiple conditions. Another potential problem is that complete data may be hard to obtain

(i.e., getting applicants who are not hired to complete a biodata form under straight, fake, and real conditions would be difficult). The between-subjects design does not suffer from these problems.

Second, we were unable to make comparisons between those employees who volunteered to participate in the study and those who did not. As mentioned in a previous section, this was because of a strong concern for anonymity and confidentiality. We believe pushing employees for demographic data would have reduced the response rate, resulted in less forthright responses, or both. Further, given that scores on the biodata form used in the study have proven to be unrelated to gender and ethnic group, concerns that volunteers and nonvolunteers may have differed on such variables should be reduced.

A more serious sampling issue is that current employees served in two conditions (straight- and fake-take groups) while job applicants served in another (real-take group). Thus, the comparability of the employee and applicant groups may be questioned. While we cannot rule out this concern altogether, several considerations lead us to believe that initial differences among the groups are not primarily responsible for our findings. First, as noted above, scores on the instrument have been shown to be unrelated to several demographic variables; thus, differences among employees and applicants on these variables should not be problematic. Second, only applicants later hired were used in the bulk of our analyses. These people, like current employees, had survived the company's extensive, multi-staged selection process; the only difference was that the applicants had completed the biodata form during selection while current employees had not. Thus, we believe it is unlikely that these applicants and current employees differ greatly on job-related ability and aptitude. Finally, scores on this biodata form have been shown to be only weakly related to cognitive ability and interview ratings. This further suggests that applicants later hired and current employees are unlikely to differ substantially on general ability and potential. Despite our conclusion that the groups were essentially comparable to begin with, we suggest that future research replicate our study using more complete random sampling and assignment for all three conditions. This is especially important given that the motivation of job applicants while completing the biodata form could have differed from the motivation of the employees.

A final potential limitation is the sample sizes in several of the conditions. While it is not unusual for experimental studies to have less than 30 observations per condition, larger sample sizes generally allow greater confidence in the generalizability of empirical findings. Replicating our work in different settings with larger samples would help to determine the extent to which our results are generalizable.

In conclusion, from a practitioner's perspective the results of this study are encouraging. While this research should be replicated and extended in other settings, the current evidence is that actual faking of professionally developed and keyed biodata forms is unlikely to substantially increase errors in selection decisions.

REFERENCES

Abrahams NM, Neuman I, Githens WH. (1971). Faking vocational interests: Simulated versus real life motivation. PERSONNEL PSYCHOLOGY, *24*, 5–12.

Asher JJ. (1972). The biographical item: Can it be improved? PERSONNEL PSYCHOLOGY, *25*, 251–269.

Bass BM. (1957). Faking by sales applicants of a forced-choice personality inventory. *Journal of Applied Psychology*, *41*, 403–404.

Borislow B. (1958). The Edwards Personal Preference Schedule and fakeability. *Journal of Applied Psychology*, *42*, 22–27.

Braun JR. (1965). Effects of specific instruction to fake on Gordon Personal Profile scores. *Psychological Reports*, *17*, 847–850.

Cascio WF. (1975). Accuracy of verifiable biographical information blank responses. *Journal of Applied Psychology*, *60*, 767–776.

Cascio WF. (1982). *Applied psychology in personnel management* (2nd ed.). Reston, VA: Reston Publishing Co.

Childs A, Klimoski RJ. (1986). Successfully predicting career success: An application of the biographical inventory. *Journal of Applied Psychology*, *71*, 3–8.

Goldstein IL. (1971). The application blank: How honest are the responses? *Journal of Applied Psychology*, *55*, 491–492.

Haymaker JC. (1986, August). *Biodata as a predictor of employee integrity and turnover.* Paper presented at the 94th Annual Convention of the American Psychological Association, Washington, DC.

Hough LM, Eaton NK, Dunnette MD, Kamp JD, McCloy RA. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology*, *75*, 581–595.

Keating E, Paterson DG, Stone CH. (1950). Validity of work histories obtained by interview. *Journal of Applied Psychology*, *34*, 6–11.

Kirchner WK. (1961). "Real-life" faking on the Strong Vocational Interest Blank by sales applicants. *Journal of Applied Psychology*, *46*, 128–130.

Klein SP, Owens, WA. (1965). Faking of a scored life history blank as a function of criterion objectivity. *Journal of Applied Psychology*, *49*, 452–454.

Kluger AN, Reilly RR, Russell CJ. (1991). Faking biodata tests: Are option-keyed instruments more resistant? *Journal of Applied Psychology*, *76*, 889–896.

Mael FA. (1991). A conceptual rationale for the domain and attributes of biodata items. PERSONNEL PSYCHOLOGY, *44*, 763–792.

Mozel JN, Cozan LW. (1952). The accuracy of application blank work histories. *Journal of Applied Psychology*, *36*, 365–369.

Mumford MD, Owens WA. (1987). Methodology review: Principles, procedures, and findings in the application of background data measures. *Applied Psychological Measurement*, *11*, 1–31.

Schrader AD, Osburn, HG. (1977). Biodata faking: Effects of induced subtlety and position specificity. PERSONNEL PSYCHOLOGY, *30*, 395–404.

Shaffer GS, Saunders V, Owens WA. (1986). Additional evidence for the accuracy of biographical data: Long-term retest and observer ratings. PERSONNEL PSYCHOLOGY, *39*, 791–809.

Weiss DJ, Dawis RV. (1960). An objective validation of factual interview data. *Journal of Applied Psychology*, *44*, 381–385.

Zerbe WJ, Paulhus DL. (1987). Socially desirable responding in organizational behavior: A reconception. *Academy of Management Review*, *12*, 250–264.