The Value of Biodata for Selecting Employees: Comparable Results for Job Incumbent and Job Applicant Samples?

James Breaugh
University of Missouri-St. Louis

Jeffrey Labrador
Kenexa

Kathleen Frye
Independent Consultant

Deborah Lee
University of Missouri-St. Louis

Vanessa Lammers
University of Missouri-St. Louis

Jenna Cox
Kenexa

The study examined whether three biodata scales (28 items, 16 items, and 6 items), that were developed based on results for job incumbents working in customer service jobs would provide scores for job applicants who were subsequently hired for these jobs that predicted their performance. We found that each of the three biodata scales was a strong predictor of job performance, was practical to use (e.g., did not require the use of a lengthy scale or complex scoring system), and did not have adverse impact for gender or race/ethnicity.

INTRODUCTION

The results of several studies suggest biodata is one of the better predictors of employee performance (Schmitt & Golubovich, 2013). Therefore, it is surprising that it is not used, at least in a formal way, by most organizations for making employee selection decisions (Gatewood, Feild, & Barrick, 2011). In addition to being used infrequently by organizations, biodata has received relatively little attention from researchers over the last decade (Cortina & Luchman, 2013). As noted by Hom (2011), it appears biodata "has lost favor among researchers" (p. 351).

In this paper, we provide a selective review of biodata research that highlights a number of factors (e.g., concerns about practicality) that likely explain why biodata is not used by more organizations.

Based on the results of this review, we offered several hypotheses that were tested in the study reported. More specifically, in an attempt to stimulate greater use of biodata by employers and more biodata research, the study we conducted examined: (a) the equivalence of validity coefficients for samples of job incumbents and job applicants in predicting employee performance, (b) the comparability of the average biodata scores for these two groups which sheds light on the possibility of applicant faking, (c) the ability of two shorter biodata scales to provide equivalent predictive accuracy in comparison to a longer scale, and (d) adverse impact linked to gender, race/ethnicity, and age.

A Selective Review of Biodata Research

A computerized search of the term biodata located more than 300 studies. Given the breadth of this research, it is not possible to summarize all of the findings. Instead, we provide a selective review of research that is most germane to the study we conducted (readers interested in a thorough review of biodata research are referred to Mumford, Barrett, & Hester, 2012).

In terms of validity, biodata has been shown to predict employee performance with greater accuracy than many commonly used selection devices (e.g., Schmidt & Hunter, 1998) and to have incremental validity when used in combination with measures of cognitive ability or personality (e.g., Mount, Witt, & Barrick, 2000). With regard to applicant faking, research (e.g., McFarland & Ryan, 2000) has shown that individuals can intentionally raise their scores on a biodata scale. In terms of reducing response distortion, studies (e.g., Lautenschlager, 1994) suggest that biodata items that are historical and verifiable are less prone to faking.

To date, biodata researchers have given relatively little attention to the issue of adverse impact (Mumford et al., 2012). The data that do exist suggests biodata has minimal adverse impact in terms of gender (e.g., Becton, Matthews, Hartley, Whitaker, 2009). With regard to race, research (e.g., Becton et al.) suggests minorities are likely to receive lower scores (for an exception, see Van Iddenkinge, Eidson, Kudisch, & Goldblatt, 2003). In terms of adverse impact linked to age, insufficient data are available to draw any firm conclusions.

Having reviewed the biodata literature, Mumford et al. (2012) noted that biodata scales often involve 150-300 items and can take 20-40 minutes to administer which can be impediments to their use by employers. In order to examine whether lengthy scales are necessary, Van Iddekinge et al. (2003) compared the validity coefficients in predicting job performance of a 42-item scale and a 6-item scale which was comprised of items from the 42-item scale. The items selected for the 6-item scale were seen as tapping skills, knowledge, and personal experience that were most important for performing well. They reported validity coefficients of .10 and .16 (both $p < .01$) for the 42-item and 6-item scales respectively.

The final area of biodata research, which is relevant for the study presented in this paper, involves comparisons of the results for job incumbents and job applicants when the same biodata scale was used for the same job in a single organization. Such studies are important because most studies have involved current employees (Mumford et al., 2012) and it is generally thought that results based on job incumbents will overestimate the validity coefficient when a biodata scale is used for selection purposes with applicants (Breaugh, 2009). We located two studies (Harold, McFarland, & Weekly, 2006; Stokes, Hogan, & Searcy, 1993) that provide insight into the legitimacy of generalizing results based on current employees to applicants.

Stokes et al. (1993) administered 168 biodata items to a sample of employees who worked in sales positions for an equipment company. Based on the results of statistical analyses (i.e., correlation and regression analysis), they selected 13 items for their final biodata scale. Stokes et al. used the same scale development strategy for a sample of applicants for sales positions with the equipment company. Statistical analyses on the 168 items resulted in a final 10-item scale. With regard to the comparability of results for the job incumbent and applicant samples, two findings are noteworthy. First, when the 13-item scale developed on current employees was applied to applicants, the validity coefficient for predicting employee turnover was .08 ($p > .05$). Second, the 13-item and 10-item scales had no items in common. Stokes et al. concluded that their "results call into question the common practice of

developing biodata keys on incumbents and assuming their generalizability to applicant samples" (p. 756).

Harold et al. (2006) had individuals who worked as call center operators and applicants for this position respond to the same 20 biodata items. These items, which had "demonstrated strong item-criterion relationships" (p. 339) in pilot testing with job incumbents, represented a subset of the biodata items administered to the incumbents. Harold et al. provided several comparisons of their two samples. However, the comparability of the validity coefficients for job incumbents and applicants is most relevant to our study. In predicting employee performance, the validity coefficients for incumbents and applicants were .27 and .17 (both $p < .01$).

Although they focused on different criteria (i.e., turnover and performance), the studies by Stokes et al. (1993) and Harold et al. (2006) suggest a validity coefficient based on a sample of current employees may overestimate what is found for an applicant sample. In the case of Stokes et al., the nonsignificant validity coefficient of .08 for applicants based on the scale developed on job incumbents is particularly concerning.

Practitioner Concerns about Biodata

Given biodata has been shown to be one of the better predictors of job performance, albeit with data generally gathered from job incumbents, it is curious that it is not widely used for selecting employees (Zibarras & Woods, 2010). A study by Furnham (2008) helps explain this lack of use. He surveyed 255 human resource (HR) professionals concerning their views on 12 selection methods (e.g., interviews, references, personal hunch). They ranked biodata 10th in terms of its validity, 9th in terms of its practicality, and 10th in terms of its legality.

One possible explanation for the HR practitioners' concern about the validity of biodata (Furnham, 2008, did not ask the participants in his study to explain their rankings) is that they do not believe results based on current employees will generalize to their organization in selecting job applicants (e.g., the results of the study by Stokes et al., 1993, are consistent with this belief). Conceivably, the practitioners surveyed by Furnham felt that, although current employees may respond honestly to biodata items, applicants are likely to distort their responses which can result in a lack of validity (Graham, McDaniel, Douglas, & Snell, 2002). With regard to the low ranking reported for biodata practicality, a number of factors may have had an influence. For example, as noted earlier, biodata scales can be quite lengthy and therefore quite time-consuming for an applicant to complete (Mumford et al., 2012). Perceptions of practicality also may have been influenced by sample size requirements. For example, in order to adequately analyze data from a scale having 100+ items, a sample of at least several hundred individuals is recommended (Cucina, Caputo, Thibodeaux, & MacLane, 2012). The practitioners in Furnham's study also ranked the use of biodata poorly in terms of legality. This may be due to some of the items (e.g., age, home ownership, marital status) that have been used by researchers. Practitioner-oriented articles (e.g., Kethley & Terpstra, 2005) and law reviews aimed at corporate attorneys (e.g., Aryani, 2009) suggesting the use of biodata can result in adverse impact also may have generated legal concerns.

A final issue that may have resulted in practitioner wariness about the use of biodata is uncertainty about what biodata actually encompasses. In this regard, although some researchers have defined biodata in terms of an individual's personal history (e.g., educational background, work experience), other researchers have "taken a much broader perspective to personal history information and include personality, preferences, and interests; future expectations, values, and self-assessed skills" (Gatewood et al., 2011, p. 338). Given biodata has been operationalized in different ways, it may be difficult for a practitioner to comprehend exactly why biodata "works" and how likely published results are to generalize to a particular organization's selection context.

Considering the practitioner concerns discussed, it is understandable why biodata has not been frequently used by employers. It seems that additional research involving job applicants that addresses biodata validity, practicality, and legality is needed to change practitioner perceptions. In response to this need, the present study utilizes applicant data and focuses on each of these topics. Furthermore, it adopts

a narrow definition of biodata (i.e., self-report personal history information), as suggested by Gatewood et al. (2011), so as not to blur the boundaries between biodata and other variables such as personality.

The Present Study

To date, only two studies (i.e., Harold et al., 2006; Stokes et al., 1993) have directly compared the results when the same biodata scale was administered to job incumbent and job applicant samples drawn from the same organization. This lack of research is unfortunate given such studies allow an accurate assessment of how confidently one can generalize from the results of a study that used current employees, which is characteristic of most biodata studies (Gatewood et al., 2011). As discussed earlier, the results of the studies by Harold et al. and Stokes et al suggest that, when a biodata scale developed based on the results for current employees is used with applicants, the validity coefficient for the applicants is likely to be lower than that for the job incumbents. This is to be expected given the way the biodata scales used in the studies by Harold et al. and Stokes et al. were developed (i.e., items were selected for the final biodata scale from a larger pool of items based on their relationship with the criterion in the incumbent sample). However, in our view, the key question is not whether a validity coefficient for applicants is lower than the one for incumbents. Rather, it is whether a biodata scale developed on incumbents will predict the criterion of interest for applicants. In answering this question, the studies by Stokes et al. and Harold et al. have limitations. For example, Stokes et al. equated their incumbent and applicant samples for work experience. However, in many selection situations, applicants often do not possess the same level of experience as current employees. Given this fact, removing the influence of experience may distort the results one would find if it were left free to vary naturally. A second aspect of the study by Stokes et al. that may limit the generalizability of their results is their use of biodata items that reflected preferences and interests. Although broadening the definition of biodata beyond that of personal history information is not uncommon, doing so blurs the distinction between the concept of biodata and concepts such as personality. We similarly view Harold et al.'s use of biodata items that assessed self-rated ability as a limitation of their study, given such items do not reflect personal history information per se.

In summary, although previous research suggest results based on job incumbents may not generalize to applicants (e.g., validity coefficients will be lower), more research is needed given the importance of this issue to practitioners and the limitations of the two studies relevant to this issue. Based on the assumptions that, in comparison to current employees, applicants are more likely to distort their responses to biodata items (to increase their likelihood of being hired) and such distortion will add error variance to their biodata scores, such faking should result in scores for applicants that are less valid. Furthermore, given the biodata items administered to applicants are commonly selected based on their ability to predict the criterion of interest for job incumbents (Stokes et al., 1993), this too suggests a lower validity coefficient for applicants.

> *Hypothesis 1: For each of the three biodata scales examined, the validity coefficient for job applicants will be lower than that for job incumbents.*

Although Hypothesis 1 predicts lower validity coefficients for job applicants, this does not mean that a biodata scale developed on job incumbents should not predict the performance of applicants. Rather, based on the findings of Harold et al. (2006) and the fact that indicators of past behavior generally have been shown to predict future behavior (Gatewood et al., 2011), we expect that biodata scores of job applicants will predict their job performance.

> *Hypothesis 2: For each of the three biodata scales examined, the scores of job applicants will predict their job performance.*

Researchers often have assumed that, if a validity coefficient for a selection device (e.g., a personality test) for job applicants is lower than that for job incumbents, this is due to the impact of

applicant faking. A method for examining the possibility of faking is to compare the mean biodata scores received by these two groups. In this regard, it is noteworthy that, although Harold et al. (2006) reported a lower validity coefficient for applicants than incumbents (.18 vs. .27), the incumbents in their study had a higher average biodata score (M = 3.34 vs. M = 3.00, $t_{(833)} = 9.27$, $p < .01$). This latter finding was not predicted given applicants were expected to raise their scores through faking. Harold et al. discussed how this mean difference could result from the job incumbents having greater job experience (i.e., many biodata items focus on work experience). Although offering a hypothesis concerning mean differences between applicants and incumbents is difficult due to countervailing forces (e.g., faking vs. work experience), given the biodata items used in our study were historical and verifiable (which has been shown to reduce faking) and given the findings reported by Harold et al. for mean differences, we expected higher biodata scores for job incumbents.

> *Hypothesis 3: Biodata scores of job incumbents will be higher than those of job applicants.*

Biodata scales frequently have been quite lengthy (Mumford et al., 2012). For example, McManus and Masztal (1999) administered a 183-item scale. The length of such scales is likely one reason that Furnham (2008) found HR practitioners viewed biodata as lacking in terms of practicality. However, as demonstrated by Harold et al. (2006) who used a 20-item biodata scale, a scale containing relatively few items can be a valid predictor of job performance. A potential limitation of using a small number of items to assess biodata is that a shorter scale could result in lower validity because the use of fewer items resulted in lower scale reliability. However such a reduction in validity could be avoided if a shorter scale consisted of items that were more predictive than those in a longer scale. This appears to be what occurred in the study by Van Iddekinge et al. (2003) who reported a higher validity coefficient for their 6-item scale (.16) than for their 42-item scale (.10) despite the fact that the reliability of their 6-item scale (.23) was lower than their 42-item scale (.44). It is recalled that Van Iddekinge et al. selected items for their 6-item scale from their 42-item scale which they saw as tapping the most critical personal attributes for job performance. In order to provide additional data concerning the potential value (i.e., practicality) of using a short biodata scale, we compared the validities of a 28-item biodata scale and two shorter scales (one having 16 items and one having six items) which used items drawn from the 28-item scale. Given the results of prior research and the procedure we used to select items for the shorter scales (paralleling the approach of Harold et al., only items predicting job performance at a statistically reliable level for job incumbents on the 28-item scale were selected for inclusion in the shorter scales), we expected scores on the shorter scales to be equally valid predictors of job performance as scores on the 28-item scale.

> *Hypothesis 4*: *Applicant scores on the 16-item and the 6-item biodata scales will account for equivalent variance in job performance as scores on the 28-item biodata scale.*

Although concerns have been raised about adverse impact resulting from the use of biodata given the different life experiences of members of different demographic groups, relatively little research has addressed this issue (Breaugh, 2009). The research that has been conducted (e.g., Bliesener, 1996) has found little evidence of differences linked to gender. Concerning race and ethnicity, although there are exceptions (e.g., Van Iddenkinge et al., 2003), the majority of studies (e.g., Bobko, Roth, & Potosky, 1999) suggest minorities receive lower biodata scores than non-minorities. Concerning applicant age, insufficient research exists to draw conclusions about adverse impact. However, as discussed by Mumford et al. (2012), it seems likely that older applicants should have a greater wealth of job experiences which should result in their receiving higher biodata scores. Given past research, we expected the following.

Hypothesis 5a: For job applicants, adverse impact linked to gender is not expected.

Hypothesis 5b: Minority applicants are expected to have lower biodata scores than non-minority applicants.

Hypothesis 5c: Older job applicants are expected to have higher biodata scores than younger applicants.

In summary, by comparing the results for job incumbents against those for job applicants who were applying for the same job held by the incumbents, the study we conducted should provide important information with regard to the concerns that have been raised about biodata validity, practicality, and legality.

METHOD

Participants and Procedure

The research reported is based on two samples. The first sample involved 131 current employees who worked as customer service representatives for a large retail organization. In terms of demographic variables (data were self-reported), there were 88 females and 31 males (no data were provided by the other 12 persons). Their average age was 36 (information on age was only provided by 44 individuals). Concerning race/ethnicity, the sample was comprised of 36 Whites, 32 Blacks, eight Latinos, with 45 individuals not reporting information (10 individuals checked the "other" category). The job incumbents responded to 37 biodata items that focused primarily on work experience (e.g., experience working with customers, tenure in previous position) and, to a lesser extent, educational background. Responses were made on a five-point scale. The supervisors of these incumbents provided ratings of their customer service, teamwork, and task performance. These ratings, which were made solely for research purposes, were averaged to create an overall measure of employee performance. All performance items were rated on a five-point continuum.

The second sample consisted of 87 job applicants who ultimately were hired for customer service representative positions with the organization for whom the incumbents worked. In terms of demographic variables, there were 41 females and 35 males (no data were provided for 11 persons). Their average age was 29 (information on age was available for 84 individuals). Concerning race/ethnicity, the sample was comprised of 42 Whites, 25 Blacks, 12 Latinos, with four individuals not reporting information (four individuals checked the "other" category). A 28-item biodata scale (these 28 items were included in the 37-item scale given to the job incumbents) was administered as part of a selection process that also involved an interview and a personality test. These applicants were rated by their supervisors using the same measure used for the job incumbents. The performance ratings, which were solely for research purposes, were made within the first 3-4 months of employment.

The Biodata Scales

Based on a number of factors (e.g., lack of item variance, employer feedback), nine of the 37 items administered to the job incumbents were not used for selection purposes (four of these items correlated at a $p < .05$ level with the overall performance rating). The resulting 28-item scale focused heavily on work experience. In order to investigate the potential value of using fewer items, we created a 16-item and a 6-item biodata scale. The 16-item scale consisted of those items in the 28-item biodata scale that correlated at a $p < .10$ level with the performance rating for the sample of job incumbents. The 6-item scale consisted of those items in the 28-item scale that correlated at a $p < .01$ level with the performance rating for the incumbents. For each scale, items were unit weighted to arrive at an overall score.

RESULTS

Descriptive Statistics and Correlations

Tables 1 and 2 present information concerning the variables examined in our study. Given much of the information in these tables is referred to in discussing the tests of our hypotheses, we only highlight a few findings here. In terms of demographic characteristics, both the sample of job incumbents and the sample of job applicants had a reasonable mix in terms of gender and race/ethnicity (Blacks and Latinos were combined in our analyses due to the relatively small number of Latinos; the results for comparisons of Whites and Blacks are similar to those we report comparing Whites vs. Blacks and Latinos combined). In terms of age, on average incumbents were older than applicants (36 vs. 29). With regard to reliability, for both job incumbents and applicants, the scale alphas for the 28-item (.77 and .86) and the 16-item (.71 and .80) biodata scales were reasonably high; the alphas for these two groups for the 6-item scale were more modest (.44 and .43) which is to be expected when fewer items are used. Concerning the evaluation of job performance, coefficient alpha for each sample was quite high (.95 and .94), and the average job incumbent rating was higher (4.01) than the applicant rating (3.65). A final issue to note is that for the two samples the variability of scores for the three biodata scales and the performance rating is similar. For example, the standard deviation for performance is .74 for incumbents and .73 for applicants.

TABLE 1
MEANS, STANDARD DEVIATIONS, AND CORRELATIONS FOR JOB INCUMBENTS

Variable	Mean	SD	1	2	3	4	5	6
1) 28-item biodata scale	3.36	.40	(.77)					
2) 16-item biodata scale	3.72	.45	.85**	(.71)				
3) 6-item biodata scale	4.15	.51	.74**	.87**	(.44)			
4) Overall performance	4.01	.74	.28**	.40**	.45**	(.95)		
5) Gender	1.26	.44	-.01	-.10	-.07	-.03	--	
6) Race/ethnicity	1.53	.50	-.31**	-.21*	-.20*	-.25*	-.02	--
7) Age	35.89	11.60	.26*	.19	.06	.10	-.08	-.10

Note: Unless otherwise noted, N = 131. For gender, N = 119 (88 females, 31 males). For race/ethnicity, N = 76 (36 Whites, 32 Blacks, and 8 Latinos). Correlations reported for race/ethnicity are based on combining those in the latter two groups. For age, N = 44. Coefficient alphas are in parentheses and appear on the diagonal. *p < .05, **p < .01.

TABLE 2
MEANS, STANDARD DEVIATIONS, AND CORRELATIONS FOR JOB APPLICANTS

Variable	Mean	SD	1	2	3	4	5	6
1) 28-item biodata scale	3.25	.56	(.86)					
2) 16-item biodata scale	3.50	.58	.95**	(.80)				
3) 6-item biodata scale	3.86	.60	.88**	.89**	(.43)			
4) overall performance	3.65	.73	.27**	.28**	.35**	(.94)		
5) gender	1.46	.50	-.15	-.19	-.12	-.07	--	
6) race/ethnicity	1.47	.50	.09	.10	.14	.00	-.12	--
7) age	28.62	9.02	.42**	.49**	.28**	.08	-.16	-.04

Note: Unless otherwise noted, N = 87. For gender, N = 76 (41 females, 35 males). For race/ethnicity, N = 79 (42 Whites, 25 Blacks, and 12 Latinos). Correlations reported are based on combining those in the latter two groups. N = 84 for age. Coefficient alphas are in parentheses and appear on the diagonal. *p < .05, **p < .01.

Hypothesis Tests

Hypothesis 1 predicted that the validity coefficients for the 28-item, the 16-item, and the 6-item biodata scales would be lower for job applicants than for job incumbents. In terms of absolute magnitude, Hypothesis 1 received support for the 28-item (.27 vs. .28), 16-item (.28. vs. .40), and the 6-item (.35 vs. .45) scales. However, none of these differences approached statistical significance. Thus, Hypothesis 1 was not supported. For example, for the 16-item scale, which had the largest difference between correlations, $z = .97$ ($p > .17$). Concerning Hypothesis 2, as predicted, job applicant scores on the 28-item ($r = .27$, $p < .01$), 16-item ($r = .28$, $p < .01$), and 6-item ($r = .35$, $p < .01$) biodata scales predicted performance ratings made by supervisors.

Based on a comparison of the means reported in Tables 1 and 2 for job incumbents and applicants, Hypothesis 3 was supported. More specifically, incumbents had higher biodata scores for the 28-item scale (3.36 vs. 3.25, $t_{(216)} = 1.69$, $p < .09$, $d = .23$), the 16-item scale (3.72 vs. 3.50 , $t_{(216)} = 3.15$, $p < .01$, $d = .43$), and the 6-item scale (4.15 vs.3.86, $t_{(216)} = 3.82$, $p < .01$, $d = .52$). Given we did not have access to information on work experience, we were not able to assess whether group differences in experience explained these findings. However, results using age as a proxy for work experience are suggestive. For example, job incumbents were older than applicants (35.89 vs. 28.62, $t_{(126)} = 3.92$, $p < .01$, $d = .70$). A second approach for drawing indirect inferences about work experience as a variable that at least partially explains the mean difference between incumbents and applicants is to examine the correlation between age and biodata score for each of the samples. In terms of job incumbents, the correlations of scores on the 28-item, 16-item, and 6-item scales with age were respectively: .26, .19, and .06 (only .26 reached a $p < .05$ level). However, given only 44 of the incumbents reported data on age, the results for applicants are more instructive (84 applicants reported information). For applicants, the correlations between age and biodata score for the three scales (see Table 2) were: 28-item scale ($r = .42$, $p < .01$), 16-item scale ($r = .49$, $p < .01$), and 6-item scale ($r = .28$, $p < .01$).

Although the 16-item and 6-item biodata scales had fewer items than the 28-item scale, based on their being comprised of items that had a higher average relationship with employee performance for the job incumbent sample and assuming this relationship would generalize to the sample of applicants, Hypothesis 4 predicted equivalent validity coefficients for the three scales. This hypothesis was supported. No difference was found even at a $p < .10$ level in comparing the 28-item and 16-item scales (.27 vs. .28, $z = .30$) or the 28-item and 6-item scales (.27 vs. .35, $z = 1.58$).

In considering the use of any selection device, an important consideration is adverse impact. Based on the findings of past research and theory concerning prior life experiences, no gender differences were predicted (Hypotheses 5a). However, differences were expected for race/ethnicity (Hypothesis 5b) and age (Hypothesis 5c). We examined adverse impact in two ways. First, we compared the average scores of members of the different demographic groups. Second, we looked for group differences if a biodata score cutoff one standard deviation above the mean were used. For the most part, the results of these two strategies were equivalent.

Concerning gender differences, although female applicants had higher scores on the 28-item, 16-item, and 6-item biodata scales (as reflected by the negative correlations of -.15, -19, and -.12 in Table 2), none of these correlations was significant at a $p = .10$ level. With regard to a chi square analysis that compared the percentage of females and males who had a biodata score one standard deviation above the mean, no evidence of a gender difference was found. For example, with regard to the 28-item scale, 7 of 41 females (17.07%) and 6 of 35 males (17.14%) had a score greater than one standard deviation above the mean ($\chi^2 = .01$, $p = .99$). Similar results were found for the 16-item scale ($\chi^2 = 1.21$, p = .27) and the 6-item scale ($\chi^2 = .49$, $p = .48$). In summary, Hypothesis 5a was supported.

In terms of adverse impact linked to race/ethnicity, it was predicted that White applicants would have higher biodata scores. From the correlations reported in Table 2, (.09, .10, and .14), it is apparent that no support was found for Hypothesis 5b for any of the three biodata scales. With regard to differences when a cutoff one standard deviation above the mean was used, no difference was found for the 28-item ($\chi^2 = .07$, p = .79) or the 16-item ($\chi^2 = .06$, p = .79) scale. In terms of the 6-item scale, a

marginal race/ethnicity effect was found (χ^2 = 3.68, p = .06). However, this difference reflected a higher selection rate for minorities.

Hypothesis 5c predicted higher biodata scores for older applicants. The correlations reported in Table 2 (.42, .49, and .28; all $p < .01$) support this hypothesis. Similar support is found if one focuses on biodata scores one standard deviation above the mean. Specifically, the chi square values for the 28-item (7.07, $p < .01$), 16-item (12.01, $p < .01$), and 6-item (8.04, $p < .01$) scales all show that applicants age 40 or above received higher scores. In interpreting these differences, it is important to note that only 8 of the 84 applicants for whom information on age was available were 40 or older. That having been said, the age effects were sizable. For example, of the eight individuals forty or older, four (50%) had a score on the 28-item biodata scale that was greater than one standard deviation above the mean. In contrast, of the 76 applicants less than 40, only four (5%) had a score greater than one standard deviation above the mean.

Ancillary Analyses

Given one of the objectives of this paper was to assess the practical importance of biodata as a selection device, we placed heavy emphasis on the results for job applicants. In this regard, for completeness, we conducted two additional types of analyses. The first type involved tests of differential validity linked to age. That is, given we found evidence of adverse impact for age, we examined whether, having included main effects for scores on the biodata scale and age, the interaction of these two variables accounted for additional variance in job performance. We found no evidence of differential validity for any of the three biodata scales. For example, for the 28-item scale, the unstandardized regression coefficient for the age x biodata score interaction was -.01 (t = -.95, p = .35).

In order to express the potential value of using biodata, we created expectancy tables in which we trichotomized the job applicant sample in terms of the biodata score received and in terms of terms of job performance. As portrayed in Table 3, for the 28-item scale, a biodata score in the top 1/3 gave an applicant a 48% chance of being in the top 1/3 in performance compared to a 33% expectation with random selection (e.g., the likelihood of the applicants being in the top third in terms of performance increased almost 50%). Similar results were found when scores on the 16-item and 6-item scores were trichotomized.

TABLE 3
EXPECTANCY TABLE FOR USE OF THE 28-ITEM BIODATA SCALE WITH THE APPLICANT SAMPLE

	Overall Performance Rating		
Biodata Score	Bottom 1/3	Middle 1/3	Top 1/3
Bottom 1/3	13 (45%)	9 (31%)	7 (24%)
Middle 1/3	13 (48%	6 (22%)	8 (30%)
Top 1/3	4 (13%)	12 (39%)	15 (48%)

Note. N = 87; parenthesized values are row percentages (e.g., 15 of 29 individuals, 48% of those receiving a biodata score in the top third received a performance score in the top third)

DISCUSSION

As noted by Gatewood et al. (2011) in discussing the use of biodata, "common practice has been to develop scoring keys on incumbents and assume they are generalizable to applicants" (p. 343) even though "the actual validity coefficients based on current employees may overestimate the validity of

applying the procedure to applicants" (p. 344). In our opinion, Gatewood et al. are correct in concluding that "until further research has shown otherwise, it seems that scoring procedures of biodata forms should be based on applicants rather than incumbents" (p. 343). The study we conducted addresses this call for "further research" by examining the findings for applicants when biodata scales developed on job incumbents were used with them. More broadly speaking, our study provides evidence that addresses serious concerns that have been raised about biodata validity, practicality, and legality.

Major Findings

The major findings of our study are easily summarized. No difference was found in terms of the magnitude of the validity coefficients for job incumbents and applicants for the three biodata scales we examined (Hypothesis 1 was not supported). The 28-item, 16-item, and 6-item scales were found to be valid predictors of job performance (Hypothesis 2 was supported). Scores on the three scales were higher on average for incumbents than for applicants (Hypothesis 3 was supported). Equivalent validity coefficients were found for the three scales (Hypothesis 4 was supported). Finally, in terms of adverse impact, no difference was found for job applicants for gender (Hypothesis 5a was supported) or race/ethnicity (Hypothesis 5b was not supported) for the three scales, but older applicants were found to have higher scores than younger applicants (Hypothesis 5c was supported).

The results reported, including those for the hypotheses that did not receive support (e.g., a lack of adverse impact linked to race/ethnicity), are quite positive in terms of the concerns that have been raised with regard to biodata validity (e.g., although items included in the 28-item, 16-item, and 6-item biodata scales were selected based on results for job incumbents, statistically significant validity coefficients were found for all three scales when used with applicants), practicality (e.g., in comparison to the lengthy biodata scales used by many researchers, scores on shorter scales – including one consisting of only six items – were shown to be valid predictors of job performance), and legality (e.g., no gender or race/ethnicity adverse impact was found nor was there evidence of differential validity for age).

Limitations and Suggestions for Future Research

We would note three limitations of our research, all of which point to the importance of our findings being replicated. The first limitation concerns our criterion measure. We focused solely on an overall evaluation of an employee's performance based on supervisory ratings made during the first few months of employment. Although our intention was to differentiate three facets of performance (i.e., customer service, team work, and task performance), the large correlations we found among scales based on these dimensions (all greater than .85) did not support attention being given to facets of job performance. In the future, it would be beneficial for researchers to measure employee performance in a more complex manner than we did. For example, coworkers or customers could evaluate job performance. It also would be important to assess whether the relationship we found between biodata score and job performance rating during the first few months of employment would be found over a longer time frame (e.g., the first year of employment). Other research directions might include looking at performance indicators that are not directly based on evaluations (e.g., attendance). A second limitation of our study was the relatively small size of the applicant sample (i.e., 87 individuals). With a larger sample size, it is likely that we could have made finer distinctions in some of our analysis. For example, a larger sample might have allowed us to differentiate between Black and Latino applicants. A third limitation of our study was that we did not have information concerning more psychologically-oriented variables (e.g., level of education) that may underlie some of the results reported. For example, it has been suggested that older individuals may receive higher scores on biodata scales because they have both a greater quantity and a greater breadth of work experience. Given we did not have data on work experience, future research that can incorporate such information is important in terms of better understanding why biodata may be predictive.

CONCLUSION

The study reported examined whether three biodata scales (28 items, 16 items, and 6 items), that were developed based on results for employees who worked in customer service jobs for a large retail organization, would provide scores for job applicants that predicted their job performance. Our results suggest that concerns that have been raised about biodata validity, practicality, and legality may be overstated. More specifically, we found that each of the three biodata scales was a strong predictor of job performance, was practical to use (e.g., did not require the use of a lengthy scale or complex scoring system), and did not have adverse impact for gender or race/ethnicity. As previously discussed, biodata has not been widely used by employers for making hiring decisions and has received less attention from researchers in recent years. Our hope is that the positive findings of the study reported increase both employer usage and researcher interest.

REFERENCES

Aryani, H. (2009). The legal implications of biodata use as an employment selection practice. *University of Pennsylvania Journal of Business Law*. Vol 11, L. 1051, Summer 2009.

Becton, J. B., Matthews, M. C., Hartley, D. L., & Whitaker, D. H. (2009). Using biodata to predict turnover, organizational commitment, and job performance in healthcare. *International Journal of Selection and Assessment,* 17, 189-202.

Bliesener, T. (1996). Methodological moderators in validating biographical data in personnel selection. *Journal of Occupational and Organizational Psychology,* 69, 107-120.

Bobko, P., Roth, P. l., & Potosky, D. (1999). Derivation and implications of a meta-analytic matrix incorporating cognitive ability, alternative predictors, and job performance. *Personnel Psychology, 52, 561-589.*

Breaugh, J. A. (2009). The use of biodata for employee selection: Past research and future directions. *Human Resource Management Review,*19, 219-231.

Cortina, J. M., & Luchman, J. N. (2013). Personnel selection and employee performance. In N. W. Schmitt and S. Highhouse (Eds.), *Handbook of Psychology, vol 12. Industrial and Organizational Psychology* (pp. 143-183). Hoboken, NJ: John Wiley & Sons, Inc.

Cucina, J. M., Caputo, P. M., Thibodeaux, C. N. (2012). Unlocking the key to biodata scoring: A comparison of empirical, rational, and hybrid approaches at different sample sizes. *Personnel Psychology, 65,* 385-428.

Furnham, A. (2008). HR professionals' beliefs about and knowledge of assessment techniques and psychometric tests. *International Journal of Selection and Assessment,* 16, 300-305.

Gatewood, R. D., Feild, H. S., & Barrick, M. (2011). *Human resource selection*. Mason, Ohio: Thompson Southwestern.

Graham, K. E., McDaniel, M. A., Douglas, E. F., & Snell, A. F. (2002). Biodata validity decay and score inflation with faking: Do item attributes explain variance across items? *Journal of Business and Psychology,* 16 ,573-592.

Harold, C. M., McFarland, L. A., Weekley, J. A. (2006). The validity of verifiable and non-verifiable biodata items: An examination across applicants and incumbents. *International Journal of Selection and Assessment,* 14, 336-346.

Hom, P. W. (2011). Organizational exit. In S. Zedeck (Ed.). *APA Handbook of Industrial and Organizational Psychology*, vol. 2, (pp. 325-375). Washington, D. C.: American Psychological Association.

Kethley, R. B. & Terpstra, D. E. (2005). An analysis of litigation associated with the use of the application form in the selection process. *Public Personnel Management,* 34, 357-376.

Lautenschlager, G. J. (1994). Accuracy and faking of background data. In G.A. Stokes, M. D. Mumford, & W. A. Owens (Eds.), *Biodata handbook* (pp. 391-419). Palo Alto, CA: Consulting Psychologists Press.

McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology,* 85, 812-821.

McManus, M. A., & Masztal, J. J. (1999). The impact of biodata item attributes on validity and socially desirable responding. *Journal of Business and Psychology,* 13, 437-446.

Mount, M. K., Witt, L. A., & Barrick, M. R. (2000). Incremental validity of empirically keyed biodata scales over GMA and the five factor personality constructs. *Personnel Psychology,* 53, 299-323.

Mumford, M. D., Barrett, J. D., & Hester, K. S. (2012). Background data: Use of experiential knowledge in personnel selection. In N. Schmitt (Ed.), *The Oxford handbook of personnel assessment and selection.* (pp. 353-382). New York: Oxford University Press.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin,* 124, 262-274.

Schmitt, N., & Golubovich, J. (2013, in press). Biographical information. In K. F. Geisinger (Ed.), *APA Handbook of Testing and assessment in psychology*. Washington, D. C.: American Psychological Association.

Stokes, G. S., Hogan, J. B., & Snell, A. F. (1993). Comparability of incumbent and applicant samples for the development of biodata keys: The influence of social desirability. *Personnel Psychology,* 46, 739-762.

Van Iddekinge, C. H., Eidson, C. E., Kudisch, J. D., & Goldblatt, A. M. (2003). A biodata inventory administered via interactive voice response (IVR) technology: Predictive validity, utility, and subgroup differences. *Journal of Business and Psychology,* 18, 145-156.

Zibarras, L. D., & Woods, S. A. (2010). A survey of UK selection practices across different organization sizes and industry sectors. *Journal of Occupational and Organizational Psychology,* 83, 499-511.