An Assessment of Biodata Predictive Ability across Multiple Performance Criteria

Michelle A. Dean
San Diego State University

Although selection tests are typically validated against only a single criterion, calls in the personnel selection literature urge the examination of selection devices across multiple criteria. This study examined biodata criterion-related validity across multiple performance criteria in a sample of 6,036 automated systems controller applicants. The performance measures examined included a paper and pencil knowledge test, behavioral ratings of simulation performance, and a paper and pencil test of simulated problems. The results suggested that the biodata instrument was useful in predicting performance across a number of criteria. The use of a shortened biodata scale captured 89% of the corrected criterion validity obtained from the original 142-item biodata instrument. Practical implications of these findings are discussed.

Biographical data (biodata) is a paper-and-pencil personnel selection technology in which applicants are asked about previous life experiences that are presumed to influence their personal development. Items are usually constructed in multiple-choice format and optimally weighted to predict criteria of interest (Mumford & Owens, 1987; Owens, 1976). Biodata can be likened to a more standardized version of a paper and pencil behavioral interview in that it focuses on applicants' past experiences. Each applicant is presented with each biodata item in an identical manner and is given the same set of possible response options from which to choose to best describe either the frequency or magnitude of his or her previous experiences. Meta-analytic reviews report average biodata cross-validities between $\bar{r} = .30$ and .40 (Beall, 1991; Hunter & Hunter, 1984; Reilly & Warech, 1990; Schmidt & Hunter, 1998; Schmitt, Gooding, Noe, & Kirsch, 1984). Biodata meta-analytic cross-validities compare well to tests of general mental ability, which are generally recognized as superior selection device in terms of criterion-related validity. Hunter and Hunter (1984) obtained mean biodata and general mental ability test criterion validities of $\bar{r} = .34$ and .38, respectively. Schmitt et al. (1984) reported slightly lower criterion validities for tests of biodata and general mental ability ($\bar{r} = .25$ and .24, respectively).

Biodata systems are described above as a personnel selection "technology" because no two biodata instruments necessarily contain the same items, tap the same constructs, or are scored the same. Thus, unlike meta-analytic estimates of criterion validity generated for cognitive ability tests, only coarse inferences about the strength of biodata's relationships with criteria can be drawn from meta-analytic estimates of biodata criterion validity. Absent measurement of constant "biodata" constructs across

studies, no inferences related to construct validity or generalizability of predictor construct and criterion construct relationships can be drawn. At best, meta-analytic estimates of biodata criterion validity strongly suggest that traditional applications of paper-and-pencil biodata inventories yield relatively high criterion validities.

Calls for the development of theoretical explanations to accompany biodata criterion validity have been made for almost 50 years (Guion, 1965). Authors have started to address these needs by developing models and theories with testable hypotheses (e.g., Dean, Russell, & Muchinsky, 1999; Mael & Hirsch, 1993; Mumford, Stokes, & Owens, 1990). Common to these calls and models is the need to explore relationships between biodata and multiple performance outcomes (see Guion, 1976, and Smith, 1976, for general discussion of this need across all personnel selection technologies).

Findings from previous research suggest that it is important to examine selection test effectiveness across a number of performance criteria. In a 1995 meta-analysis examining the relationship between subjective ratings and objective measures of job performance, Bommer, Johnson, Rich, Podsakoff, and MacKensie found an average correlation of .39 between subjective and objective ratings. Additionally, Sackett, Zedeck, and Fogli (1988) noted the importance of distinguishing between prediction of typical and maximal performance. They reported that maximal and typical job performance were only modestly correlated (.16 and .36 in new employee and current employee samples, respectively). This finding suggests typical and maximal job performance measures tap different construct domains and, in turn, may have different causal influences and predictors. In a study that examined biodata against multiple performance criteria, Russell, Mattson, Devlin, and Atwater (1990) found moderate variation in biodata criterion validities across traditional academic and military performance measures obtained on Naval Academy midshipmen. Muchinsky (2003) suggested that the more complex the job, the more performance criteria are needed in order to capture all aspects of job performance.

Employers with large numbers of applicants need efficient screens to uncover their most qualified applicants. Traditional means of applicant-employer information transfer (i.e., cover letters, resumes, paper application blanks, written responses to each applicant, off-site interviews, and on-site interviews) are labor intensive and become increasingly expensive as the number of applicants increases. Biodata instruments are primarily used as an initial screening device, most appropriately used early in the selection process. However, biodata instruments are typically quite long, containing 100 or more items, thus possibly discouraging their use in practice even though they yield relatively high criterion validities relative to other selection devices. A short biodata scale used at initial stages of recruiting/selection sequences could provide a cost effective means of reducing extremely large applicant pools before resorting to more costly, labor intensive hurdles.

The primary purpose of the current study is to empirically examine differential relationships between biodata and 1) behavioral performance ratings, 2) job knowledge tests, 3) general knowledge paper-and-pencil tests, and 4) applied knowledge paper-and-pencil tests. A secondary purpose is to determine how well a short biodata scale created using a subset of biodata items might predict performance relative to the entire biodata inventory.

Method

Sample

The sample consisted of 6,036 applicants for automated systems controller positions for a governmental agency. Applicants were 80% white, 74% men, and averaged 25.8 years of age ($SD = 3.8$ years). These participants were randomly divided into two groups: 80% into a key development sample and 20% into a cross-validation sample for purposes of empirically scoring and cross-validating the biodata instrument. A cross-validation sample was used to ensure that the biodata empirical keys' predictive ability held up in an independent sample, thus guarding against capitalizing on chance predictor-criterion associations in the key development sample.

Predictor Measures

Biodata. The 142-item biodata questionnaire was administered to applicants for research purposes (i.e., the biodata inventory was not used in the selection of these candidates). The biodata instrument was developed from reviews of 1) qualification standards for the position, 2) job analysis information, 3) previous biodata efforts at this governmental agency, 4) interviews with training personnel to determine characteristics of successful candidates, and 5) interviews with supervisors to ascertain characteristics differentiating good and poor employees. In general, the items tapped previous life events related to high school, college, and previous work experience. The following are two example items from the biodata instrument:

The number of different high school sports I participated in was:
a.	4 or more
b.	3
c.	2
d.	1
e.	Didn't play sports

During my last year in college as a full-time student, my average number of hours of paid employment per week was:
a.	More than 20 hours
b.	10 - 20 hours
c.	Fewer than 10 hours
d.	None
e.	Didn't go to college

The biodata questionnaire was scored empirically at the response option level. Each of the 142 items had five response options yielding 710 (142 x 5) response options total. Each response option was weighted using its correlation with the criterion of interest in the key development sample. Participants' biodata scores in the cross validation sample were then set equal to the sum of the correlation weights associated with the 142 unique response options that each participant selected. The biodata questionnaire was administered to the applicants after they had taken a general mental ability test. The internal consistency reliability for this biodata instrument was .79.

To examine the viability of a shorter biodata scale, an additional biodata scoring key was developed using only those response options that yielded a correlation of greater than $\pm .10$ with the composite training score in the key development sample. Of the 710 response options in the instrument, 21 had correlations greater than $\pm .10$. These 21 response options came from a total 18 items. These 18 items were subsequently included in the short form assessment.

General mental ability test. A 110-item Office of Personnel Management (OPM) general mental ability test was the first hurdle in the selection process. It was designed to measure traditional cognitive aptitudes such as arithmetic reasoning, data interpretation, table reading, and spatial relations. Manning, Della Rocco, and Bryant (1989) found that this test's scores were significantly related to performance in a nine-week training/screening program and on-the-job training performance across many candidate cohorts for this particular position. The cut score used for this test was 85 points or higher on a 100 point scale as determined by previous applicant groups' performance. Those "passing" this test were then invited to attend the above-mentioned training program. An invitation to attend training represented a contingent job offer, the contingency being successful completion of the program by achieving acceptable scores on the various performance tests in the program. Available data suggested the general mental ability test had acceptable reliability but was vulnerable to practice effects, thus only first time applicants were included in the current sample.

Performance Criteria

Successful applicants were invited to attend a nine-week training/screening program. This program provided instruction in basic job rules and procedures and tested candidate knowledge through written exams and laboratory simulations. Four categories of performance assessment examined included:

1) *Skills test.* This test measured the job candidate's ability to apply job principles to resolve simulation problems in an objective paper and pencil format.

2) *Mid-term exam.* This was an objective paper-and-pencil multiple choice test that examined each candidate's ability to learn factual information administered mid-way through the training program.

3) *Comprehensive final exam.* This objective paper-and-pencil multiple choice test was administered at the end of the training program to measure knowledge acquisition at the end of the course.

4) *Behavioral ratings.* Instructors systematically evaluated trainee performance in six 30 minute observation-based ratings of performance in laboratory high fidelity simulations of actual job performance. Simulation scores consisted of an instructor technical assessment of the number of errors observed as well as a subjective assessment of trainee performance. The overall simulation score was based on the average of the top 5 out of 6 simulation scores.

5) *Composite performance score.* This score was a composite of the above individual assessments in which each test score was weighted as follows: skills test, 20%; instructor's behavioral ratings of candidate simulation performance, 60%; and final exam, 20%. The weighted scores were then summed to form the composite performance measure. The composite score was the best job performance measure available for this position due in part because job performance variability is minimal given the public safety nature of this job and union agreements mandating that employees' performance could only be evaluated on a dichotomous, satisfactory/non-satisfactory scale. Nonetheless, governmental studies have documented this training program's ability to predict subsequent job performance (Della Rocco, Manning, & Wing, 1990). No reliability data were available on the criterion measures.

Analyses

Correlational analyses were performed on the cross-validation sample to determine the criterion-related validity of biodata for each performance criterion. These correlations were also corrected for indirect range restriction due to prior selection on the general mental ability measure. This served to restrict the sample available for these analyses and likely resulted in correlations that were underestimates of the true population correlations for biodata with each respective criterion. The corrected correlations are provided for informational purposes in terms of the impact of range restriction of a previous selection hurdle on subsequent selection device criterion-validities. The uncorrected correlations can be seen as conservative estimates of biodata criterion validity for these data.

Hierarchical regression analyses were performed to examine whether biodata provided incremental validity above the contribution of cognitive ability. Thus cognitive ability was entered in step one of the regression analysis, and biodata was entered in step two. The relative contributions of these variables were examined by inspecting their standardized regression coefficients (βs). The change in variance accounted for in step two of the regression analyses was examined for evidence of the incremental validity of biodata.

Results

Criterion validities obtained for each biodata scale and the criterion measures in the cross-validation sample are reported in Table 1. Correlations corrected for indirect range restriction on the general cognitive ability test (r_c) are reported in Table 2. The biodata scales showed a consistent pattern of prediction across the various criteria (yielding correlations between .22 and .33). Specifically, scores on the biodata and the composite performance measure correlated .33 ($r_c = .45$), biodata and mid-term academic knowledge exam correlated .23 ($r_c = .35$), biodata and the paper and pencil simulated problems test correlated .33 ($r_c = .45$), biodata and the final exam correlated .22 ($r_c = .32$), and biodata and the behavioral performance ratings measure correlated .29 ($r_c = .37$).

The short biodata scale, which was scored using only items with response options correlated with the composite performance measure greater than +/- .10, yielded respectable levels of criterion-validity. The short biodata scale correlated .27 ($r_c = .39$) with the composite performance measure. Interestingly, the short biodata scale also fared well in predicting other criterion measures with correlations ranging from .17 to .31. The 18-item short biodata scale captured 89% of the corrected criterion validity obtained from the original 142 item biodata instrument. This decrease is surprisingly low considering 689 response options (97% of the entire biodata instrument) were removed from the original scoring key.

The biodata inventory, the short biodata scale, and general mental ability correlated .33, .27, and .16, respectively, with the composite performance measure. However, a more fair comparison of the biodata and general mental ability criterion validities requires correcting these correlations for range restriction on the general mental ability test (which was a hurdle administered prior to the biodata instrument). The small correlation between the general mental ability test and the composite criterion was likely due to the fact that the cognitive ability measure was used as an initial hurdle on these applicants, so only those passing the cognitive ability test were available for further analysis. Correcting the correlation between cognitive ability and the composite criterion yields a correlation of .40. This correlation is much closer to what one would expect to see and would have likely seen in these data had the cognitive ability test not been used as an initial selection hurdle on these applicants. The corrected correlations for the biodata long and short forms were .45, and .39, respectively. Corrected correlations for all criterion validities are reported in Table 2.

There was some shrinkage comparing the validities yielded in the development and hold out samples. This is expected given the key development sample validities are probably taking advantage of some chance associations in the data given the weights used to score the biodata instrument were calculated off of this group. For the full biodata form, the average validity shrinkage across all criteria was 15%, for the short form, the average percent shrinkage was 6%.

Hierarchical regression analysis results suggest that biodata does add incremental validity beyond that accounted for by the cognitive ability test. Both regression equations were significant ($p < .01$). The first regression equation ($R^2 = .03$, $p < .01$) indicated that cognitive ability ($\beta = .17$, $p < .01$) was significantly associated with the composite criterion. . In the second step of the regression equation (overall $R^2 = .12$, $p < .01$), cognitive ability ($\beta = .12$, $p < .01$) and biodata ($\beta = .31$, $\underline{p} < .01$) were significantly associated with the composite criterion. The addition of biodata accounted for a significant increase in variance in the criterion ($\Delta R^2 = .09$, $p < .01$). The results of the hierarchical regression analyses are presented in Table 4.

Table 1
Descriptive Statistics and Correlations for Cross-Validation Sample

Variable	Mean	SD	N	1	2	3	4	5	6	7	8	9	10	11
Predictors														
1. Biodata-Composite score[1]	100.5	1.04	980											
2. Biodata-Midterm exam	100.7	1.11	980	.71										
3. Biodata-Skill test	100.4	1.18	980	.89	.83									
4. Biodata-Final exam	100.4	1.02	980	.83	.93	.87								
5. Biodata-Behavioral ratings	100.5	0.92	980	.97	.53	.76	.70							
6. Biodata-Short form/composite	100.2	0.59	974	.81	.65	.87	.70	.72						
7. Cognitive ability test	91.6	5.04	2181	.14	.17	.21	.14	.10	.19					
Training Criteria														
8. Composite measure	71.1	11.48	1989	**.33**	.23	.29	.25	.31	.27	.16				
9. Midterm exam	92.6	7.95	2192	.21	**.23**	.19	.24	.18	.17	.12	.55			
10. Skills test	74.9	14.30	1989	.32	.24	**.33**	.25	.28	.31	.18	.83	.45		
11. Final exam	89.5	8.10	2187	.24	.20	.24	**.22**	.22	.21	.10	.55	.48	.46	
12. Behavioral ratings	63.0	14.21	1993	.29	.19	.24	.21	**.29**	.23	.14	.98	.45	.70	.43

[1] Information reported after each biodata scale is the criterion used to empirically key that particular scale

[2] Correlations in bold face are cross-validities between each biodata scale and its targeted criterion

Table 2: Corrected and Uncorrected Correlations

Criterion	Biodata - Full Form		Biodata - Short Form		Cognitive Ability	
	r	r_c	r	r_c	r	r_c
Composite training measure	.33	.45	.27	.39	.16	.40
Training mid-term exam	.23	.35	.17	.29	.12	
Skills test	.33	.45	.31	.44	.18	
Comprehensive final exam	.22	.32	.21	.28	.10	
Behavioral ratings	.29	.37	.23	.31	.14	

Table 3: Development Sample Validity Coefficients (N = 3375)

Criterion	Biodata - Full Form	Biodata - Short Form
Composite training measure	.37	.28
Training mid-term exam	.29	.18
Skills test	.37	.30
Comprehensive final exam	.28	.19
Behavioral ratings	.33	.24

Table 4: Hierarchical Regression Analyses (N = 686)

Predictors	β	Total R^2
Step 1		.03**
Cognitive Ability	.17**	
Step 2		.12**
Cognitive Ability	.12**	
Biodata	.31**	

** $p < .01$

Discussion

This study examined whether biodata exhibited different criterion validities across several measures of training performance. It also examined the validity yielded when scoring only a small fraction of the biodata instrument. The results suggested that biodata predicted performance well across the criteria examined. Closer examination of the criterion-related validities showed that the biodata inventory tended to predict criterion performance requiring the ability to apply facts to solve problems better than criterion performance measuring knowledge acquisition. The performance measures requiring knowledge application—the skills test, behavioral ratings, and the composite performance measure (which was 80% applied problems)—were correlated .33, .33, and .29, respectively, with the biodata instrument, whereas the knowledge acquisition performance measures (the mid-term and final examination) were correlated .23 and .24 with the biodata instrument.

The results of this study also suggested that it may be possible to shorten typically lengthy biodata inventories with minimal loss of predictive validity, making biodata an effective means of decreasing selection system costs as part of the initial screening process. A qualitative content analysis of the 21 items used in the short scale (i.e., those items that had the strongest predictive ability for these data) showed that they tended to tap previous life experiences relating to general mental abilities displayed during high school. The pattern of correlations found for the short biodata scale was similar to those found when the entire biodata instrument was scored in that the short

form tended to do a better job of predicting criterion performance requiring the ability to apply facts to solve problems over knowledge acquisition. Future research should validate these particular items for other jobs to determine if the previous life experiences captured in these items are developmental for more than one type of occupation/job.

Both biodata scales fared well in terms of performance prediction relative to the general mental ability test. These findings were consistent with previous meta-analytic studies that suggested that biodata and general mental ability tests yield comparably high criterion validities (Hunter & Hunter, 1984; Schmitt, Gooding, Noe, & Kirsch, 1984). Given that biodata traditionally displays little adverse impact against protected subgroups (Dean, 1999; Pace & Schoenfeldt, 1977; Reilly & Chao, 1983; Reilly & Warech, 1990) unlike tests of general mental ability (Gottfredson, 1986; Sackett & Wilk, 1994; Schmitt, Clause, & Pulakos, 1996), use of a short biodata inventory early in the recruiting/selection sequence may decrease selection system costs while also lowering exposure to possible Equal Employment Opportunity litigation.

In sum, this study found that biodata yielded respectable criterion validities across a number of performance criteria and provided initial evidence that the same biodata instrument can be used for prediction across a variety of performance measures. The results also suggested that selection systems may be simplified through the use of a short biodata inventory with negligible loss in predictive power. Future research should target items for shorter biodata forms and validate them across multiple criteria in assorted occupations. Such research would facilitate the understanding of which items predict best for which criteria and will help determine if there are generalized developmental life experiences that are useful for performance prediction across different occupations.

References

Beall, G. E. (1991). Validity of the weighted application blank across four criteria. *Applied H.R.M. Research, 2*(1), 18-26.

Bommer, W. H., Johnson, J., Rich, G. A., Podsakoff, P. M., & MacKensie, S. B. (1995). On the interchangeability of objective and subjective measures of employee performance: A meta-analysis. *Personnel Psychology, 48*, 587-605.

Dean, M. A. (1999). *Response option examination of biodata criterion-related validity and adverse impact*. Paper presented at the annual International Personnel Management Association Assessment Council Meeting, St. Petersburg, FL.

Dean, M. A., Russell, C. J., & Muchinsky, P. M. (1999). Life experiences and performance prediction: Toward a theory of biodata. In G. Ferris (Ed.), *Research in personnel/human resources management* (Vol. 17: 245-281). Stamford, CT: JAI Press.

Della Rocco, P. S., Manning, C. A., & Wing, H. (1990). *Selection of controllers for automated systems: Applications from current research* (Report No. 90-13). Washington, DC: Office of Aviation Medicine.

Gottfredson, L. (1986). Societal consequences of the g factor in employment. *Journal of Vocational Behavior*, *29*, 379-410.

Guion, R. M. (1965). *Personnel testing*. New York: McGraw-Hill.

Guion, R. M. (1976). Personnel selection. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology*. Chicago, IL: Rand McNally.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96,* 72-98.

Mael, F. A., & Hirsch, A. C. (1993). Rainforest empiricism and quasi-rationality: Two approaches to objective biodata. *Personnel Psychology*, *46*, 719-738.

Manning, C. A., Della Rocco, P. S., & Bryant, K. (1989). *Prediction of success in air traffic control field training as a function of selection and screening performance* (Report No. 89-6). Washington, DC: Office of Aviation Medicine.

Muchinsky, P. M. (2003). Psychology applied to work (7th Ed.). Belmont, CA: Wadsworth.

Mumford, M. D., Stokes, G. S., & Owens, W. A. (1990). *Patterns of life history: The ecology of human individuality*. Hillsdale, NJ: Lawrence Erlbaum.

Mumford, M. D., & Owens W. A. (1987). Methodological review: Principles, procedures, and findings in the application of background data measures. *Applied Psychological Measurement*, *11,* 1-31.

Pace, L. A., & Schoenfeldt, L. F. (1977). Legal concerns with the use of weighted application blanks. *Personnel Psychology*, *31*, 243-247.

Reilly, R. R., & Chao, G. T. (1982). Validity and fairness of some alternative employee selection procedures. *Personnel Psychology*, 35, 1-62.

Reilly, R. R., & Warech, M. A. (1990). *The validity and fairness of alternative predictors of occupational performance*. Paper invited by the National Commission on Testing and Public Policy, Washington, DC.

Rock, D. B., Dailey, J. T., Ozur, H., Boone, J. O., & Pickrel, E. W. (1981). Conformity of the new experimental test battery to the uniform guidelines on employee selection requirements. In S. B. Sells (Ed.) *Selection of air traffic controllers* (Report No. 84-2). Washington, DC: Office of Aviation Medicine.

Russell, C. J., Mattson, J., Devlin, S. E., & Atwater, D. (1990). Predictive validity of biodata items generated from retrospective life experience essays. *Journal of Applied Psychology, 75,* 511-520.

Sackett, P. R., & Wilk, S. L. (1994). Within-group norming and other forms of score adjustment in preemployment testing. *American Psychologist*, *49*, 929-954.

Sackett, P. R., Zedeck, S., & Fogli, L. (1988). Relations between measures of typical and maximum job performance. *Journal of Applied Psychology*, *73,* 482-486.

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., Clause, C. S., & Pulakos, E. D. (1996). Subgroup differences associated with different measures of some common job relevant constructs. In C. L. Cooper and I. T. Robertson (Eds.), *International Review of Industrial and OrganizationalPsychology, 11*, 115-137.

Schmitt, N., Gooding, R. Z., Noe, R. D., & Kirsch, M. (1984). Meta-analyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology, 37*, 407-422.

Smith, P. C. (1976). The criterion problem. In M.D. Dunnette (Ed.), *Handbook of Industrial and Organizational Psychology*. Chicago, IL: Rand McNally.

Author Notes

Questions about this manuscript should be addressed to the author at:
Michelle A. Dean
Department of Management
College of Business Administration
San Diego State University
5500 Campanile Drive
San Diego, CA 92182-8238
Voice: (619) 594-6839
Fax: (619) 594-3272
Email: michelle.dean@sdsu.edu