

Non-Cognitive Predictors of Air Traffic Controller Performance

Michelle A. Dean
Department of Management
University of North Texas
Denton, TX 76203-5429
(940) 565-4487
Fax: (940) 565-4394
e-mail: deanm@unt.edu

Craig J. Russell
Division of Management
Price College of Business
University of Oklahoma
Norman, OK 73019-0450
(405) 325-2458
Fax: (405) 325-1957
e-mail: cruss@ou.edu

William L. Farmer
FAA Civil Aeromedical Institute
PO Box 25082
Oklahoma City, OK 73125-0082
(405) 954-5954
Fax: (405) 954-4852
e-mail: Bill_Farmer@mmacmail.jccbi.gov

Non-Cognitive Predictors and Air Traffic Controller Performance

Abstract

Theory and empirical evidence suggest cognitive factors are viable predictors of performance of air traffic controller functions. Non-cognitive predictors historically played a secondary role in U.S. air traffic controller selection, but have earned renewed interest in other employment contexts. This chapter presents an overview of two important non-cognitive predictors of job performance: biographical information (biodata) and personality measures. Each predictor domain will be briefly reviewed and practical guidelines for future selection practice offered.

Non-Cognitive Predictors of Air Traffic Controller Performance

Cognitive tests yield among the highest criterion-related validities available in personnel selection settings. Unfortunately, reliance on cognitive ability tests may not capture the whole performance picture. Use of cognitive predictors has well documented adverse impact against minority groups. Continued sole reliance on cognitive predictors as employment screens in an increasingly diverse work force may serve to stifle diversity and violate Federal equal opportunity regulations. Non-cognitive predictors promise comparable criterion validities in personnel selection though they are used infrequently relative to cognitive predictors. Research suggests non-cognitive selection devices display meaningful incremental criterion-related validities in combination with cognitive ability predictors. This chapter discusses the merits of including two non-cognitive predictors in the air traffic controller selection process: biographical information (biodata) and personality. Literature addressing both predictors will be briefly reviewed, with particular emphasis given to findings reported with air traffic controllers. Practical guidelines for biodata and personality inventory use in air traffic controller selection are presented.

Biodata

Biodata is a paper and pencil selection technique using questions focusing on previous life experiences presumed to causally influence personal development. This assumption is implicit in the “consistency principle,” which argues that the best predictor of future behavior is past behavior (Owens, 1976, Wernimont & Campbell, 1968). Items included on biodata inventories typically emphasize the magnitude or frequency of a

previous behavior. Items are usually constructed in multiple choice format and are optimally weighted to predict criteria of interest (Mumford & Owens, 1987; Owens, 1976).

Biodata selection procedures consistently demonstrate among the highest criterion validities available. Meta-analytic reviews report average biodata cross-validities between $\bar{r} = .30$ and $.40$ (Hunter & Hunter, 1984; Reilly & Chao, 1982; Schmidt & Hunter, 1998; Schmitt, Gooding, Noe, & Kirsch, 1984). Biodata meta-analytic criterion validity estimates compare well to more frequently used tests of general cognitive ability (g). Hunter and Hunter (1984) obtained mean biodata and g criterion validities of $\bar{r} = .34$ and $.38$, respectively, correcting g for measurement and sampling error, while correcting biodata only for sampling error. Schmitt et al. (1984) reported nearly identical criterion validities for g and biodata ($\bar{r} = .243$ and $.248$, respectively) correcting both for sampling error.

Minorities typically score one standard deviation below majority applicants on standardized g tests (Hunter & Hunter, 1984; Sackett & Wilk, 1994; U. S. Employment Service, 1970). Hence, the need to incorporate other equally valid predictors with lesser adverse impact becomes an important practical issue. Several literature reviews concluded biodata inventories tend to display low adverse impact (Barge & Hough, 1988; Mitchell, 1994; Mumford & Stokes, 1992; Reilly & Chao, 1982; Reilly & Warech, 1990). Pace and Schoenfeldt (1977) noted biodata achieves low adverse impact through application of empirically derived biodata scoring keys. Items demonstrating differential criterion prediction in cross-validation samples are simply deleted from the scoring key.

Biodata and Air Traffic Controller Selection

The Federal Aviation Administration Civil Aeromedical Institute first examined biodata as a possible selection device for air traffic controllers in the mid 1980s. Specifically, two biodata inventories were initially designed and administered: the FAA Biographical Questionnaire (BQ) and Applicant Background Assessment (ABA). Items on the BQ inventory tapped eight content areas: 1) educational background, 2) prior military or civilian experience in air-traffic related work, 3) importance placed on various factors (e.g., salary, benefits, job security), 4) time expected to become an effective air traffic controller specialist (ATCS), 5) commitment to an ATCS career, 6) work-related attitudes, 7) expected satisfaction with aspects of ATCS careers, and 8) general personal information (e.g., socioeconomic status growing up, alcohol and tobacco usage; Collins, Manning, & Nye, 1990).

Preliminary research conducted on the BQ suggested the following biographical arenas were most useful in predicting air traffic controller success: pre-FAA air traffic control experience, high school grades in math and science, self-assessment of performance potential, previous military ATC experience, and tendency to help friends with problems (Collins, Manning, & Taylor, 1984; Taylor, VanDeventer, Collins, & Boone, 1983).

Collins, Manning, and Taylor (1984) examined biodata's relationship to training performance (pass/fail) and differential prediction for minority versus non-minority groups. Variables most predictive of majority candidate job performance included (in order

of effect size) age, high school physical science grades, and self-assessment of future ATCS performance. Three significant predictors of Academy pass/fail status for non-minority candidates included (in order of effect size) high school math grades, age, and self-assessment of future ATCS performance.

Collins et al. (1990) examined the individual and joint predictive power of biodata and a general cognitive ability measure. Biodata yielded a cross-validated correlation of $r = .34$ ($r_c = .37$, $N = 2,766$, where r_c = criterion validity corrected for range restriction) with a training performance score. A cognitive ability measure correlated $r = .22$ ($r_c = .51$, $N = 2,766$) with training performance. Multiple regression analysis suggested addition of biodata significantly increased the range restriction corrected multiple correlation from .51 to .58 ($N = 3,156$; Collins et al., 1990). In sum, use of the BQ biodata inventory initially appears very useful in predicting air traffic controller specialist performance, adding predictive power to existing forecasts drawn from cognitive ability measures.

Recently, Dean (1999) analyzed the Applicant Background Assessment (ABA) biodata inventory for its criterion-related validity in predicting air traffic controller training performance and racial adverse impact on a sample of 6,035 FAA air traffic controller specialist candidates. ABA inventory was based on reviews of: 1) qualification standards for ATCS, 2) job analyses conducted by the FAA, 3) previous biodata work done at the FAA, 4) interviews with training staff members to determine ATCS characteristics differentiating those who perform better in training and those who fail training, and 5) interviews with ATCS supervisors to ascertain characteristics differentiating good and poor ATCSs. ABA

items were limited to those dealing with experiences under the control of applicants (Dr. Dana Broach, personal correspondence, October, 1997) and covered areas such as high school, college, and previous work experience.

Dean assessed the ABA's simple criterion validity in predicting training performance and incremental criterion validity when compared it to the existing FAA cognitive ability measure. The empirically-keyed biodata instrument correlated $r = .37$ ($r_c = .44$, corrected for indirect range restriction on the general cognitive ability composite score) and the FAA cognitive ability composite score correlated $r = .16$ ($r_c = .42$, corrected for direct range restriction) with performance on an FAA air traffic controller training program in the cross-validation sample, suggesting both were valid predictors of performance. Relative contributions of biodata and cognitive ability to performance prediction was assessed using hierarchical multiple regression. Biodata yielded a change in R of .113 when added to a regression equation with a cognitive ability measure, while cognitive ability added a change in R of .071 when added to the biodata inventory.

Dean (1999) also examined the ABA biodata instrument for possible adverse impact against Blacks. Two separate biodata empirical keys that either included or excluded adverse impact response options (i.e., response options chosen by blacks less than 80% of the majority's selection rate) were constructed and compared. Deleting adverse impact response options caused a large decrease in standardized mean differences between Black and White group means and no significant decrease in biodata criterion-related validity.

Specifically, the Black/White standardized mean difference (uncorrected \underline{d}) and standardized mean difference corrected for range restriction on the cognitive ability measure (\underline{d}_c) for both biodata keys and a cognitive ability composite score were calculated. Deletion of adverse impact response options resulted in a 32% decrease in corrected mean difference in biodata scores of from $\underline{d} = .364$ to $.117$. A biodata cross-validity of $\underline{r} = .34$ ($\underline{r}_c = .424$) was obtained from a key excluding adverse impact response options, while a cross-validity of $r = .37$ ($\underline{r}_c = .44$) was obtained for a biodata key with all response options scored. The two correlations were not significantly different. In sum, the standardized mean subgroup difference on the biodata instrument decreased by 32% when adverse impact response options were removed from the scoring key, yet criterion-related validity decreased by only 5.7%.

Constructing Biodata Instruments

We now turn to guidelines for development and use of biodata in ATCS selection scenarios drawn from the literature on biodata item content, development, and scoring.

Biodata item content. Mael (1991) summarized previous biodata taxonomic work (e.g., Asher, 1972) and grouped typical biodata item attributes into three general categories: 1) historical, 2) methodological, and 3) controllability/job relevance. His taxonomic effort pinpointed types of items typically found on biodata instruments and attempted to differentiate biodata items from closely related, but conceptually distinct, non-cognitive measures such as personality tests.

Mael (1991) suggested the historical nature of items constituted biodata's defining characteristic. Biodata items typically concern actual past events that have taken place in one's life and do not include hypothetical scenarios such as found in situational interviews (cf., Latham, Saari, Pursell, & Campion, 1980). Mael (1991) suggested questions of general attitudes (e.g., "Would you describe yourself as shy?") not relating to a specific past event are outside the realm of biodata and are more closely aligned with dispositional or personality measures.

Certain methodological attributes such as externally focused, objective, first-hand, and verifiable are thought to aid in obtaining accurate biodata responses (Asher, 1972; Mael, 1991). Externally focused items tap some action or event in which an individual was involved and are not merely opinions or reactions to an event. Many researchers advocate biodata items be objective rather than subjective. Mael suggested biodata items should ask for respondent's first-hand knowledge, avoiding asking individuals about how others would evaluate the respondent. For example, asking, "How did your parents evaluate your academic achievement?" would be second-hand information in which the respondent is asked to speculate. Another recommended biodata item attribute is that they be verifiable. Verifiable items are believed to result in more honest, accurate responses by discouraging socially desirable responses or faking (Atwater, 1980; Cascio, 1975; Mosel & Cozan, 1952). Interestingly, Hough, Eaton, Dunnette, Kamp, and McCloy (1990) found simply warning respondents that answers can be verified may act as a faking deterrent.

Items may also vary in terms of controllability and visible job relevance.

Controllability refers to the degree to which a person chose to perform or not to perform an action (e.g., behaviors in which a person chooses to engage, such as playing sports in high school, versus circumstances that happen beyond a person's control, such as parental socioeconomic status). Mael (1991) suggested all life events (consciously chosen or not) have the ability to shape a person's future behavior and should be included on a biodata instrument. There are conflicting opinions regarding whether to include items such as parental behavior and socioeconomic status due to applicants' lack of control over their early environment (cf., Mael, 1991; Stricker 1987, 1988). Biodata items may also vary in visible job relevance. Some researchers prefer only items with point-to-point relationships with job content to increase item face validity.

Biodata performance prediction generally does not involve literally predicting future performance by measuring identical past performance (Dean, Russell, & Muchinsky, 1999). Wernimont and Campbell's (1968) "samples" versus "signs" distinction defined samples as representing past behaviors used to predict future behaviors drawn from a single performance domain. Behavioral signs are not equivalent to criterion domain behaviors, instead being drawn from domains hypothesized to 1) causally influence subsequent performance or 2) be highly correlated with those causal influences.

Biodata instruments generally use both signs and samples of past behavior to predict future performance outcomes (Dean et al., 1999; Russell, 1996). However, biodata is often used in scenarios where applicants may not have previous identical work experience (e.g.,

entry level jobs), and therefore, no past “sample” behaviors are available resembling desired future behaviors (cf., Russell, Mattson, Devlin, & Atwater, 1990). “Sign” biodata items are critical in situations where applicants have demonstrated no prior work-related behavior. Russell et al. (1990) faced this scenario when developing a biodata instrument to predict performance of high school applicants as midshipmen at the Naval Academy. High school seniors had no opportunities to exhibit “samples” of Naval Officer behaviors up to that point in their lives, making it necessary to find adolescent and pre-adolescent experience “signs” predicting future success as a Naval officer. Items focused on school, social, and employment experiences in predicting Academy success. Russell et al. found these experiences resulted in accurate prediction of subsequent academic and non-academic performance criteria. The empirically keyed biodata scales demonstrated incremental criterion-related validity when combined with a measure of cognitive ability. Similarly, the majority of recent applicants for Federal Aviation Administration air traffic controller positions have no previous experience air traffic controller experience (Collins, Manning, & Taylor, 1984; Collins, Nye, & Manning, 1990), necessitating greater reliance on signs than samples of desired future behavior. We now turn to a discussion of methods for generating item content for biodata inventories.

Biodata Item Development. It is useful to recall the defining characteristic of any biodata item is the reference to some historical event (Mael, 1991). However, this defining characteristic still leaves the biodata item developer with a virtually unlimited variety of item content. We briefly describe common item generation procedures found in the

literature, though the reader is referred to Mumford and Owens (1987) and Russell (1994) for a comprehensive discussion. For our purposes, we will focus discussion on different item generation techniques targeting criterion construct prediction. Note, this is a very different focus than that found in typical psychometric scale development efforts. No latent “biodata” constructs need be hypothesized.

Mumford and Owens (1987) described six sources of biodata items: the human development literature, life history interviews with incumbents, typical factor loadings of biodata items, known life history correlates with the criterion domain, existing biodata items with known criterion validities, and items generated from investigators’ general psychological knowledge. The latter four sources rely on investigators’ subjective judgments or existing biodata inventories (Russell, 1994). Use of existing biodata item pools and expert judgments by investigators will undoubtedly play a key role in ongoing validation research and subsequent item revision. However, since only one biodata instrument has been used and validated in one sample drawn from the air traffic controller applicant population, the latter four techniques for generating biodata items are of little use.

The first two sources capture a basic distinction in approaches to biodata research. Specifically, one expects the literature on human development to provide theories, models, associated constructs, and operationalizations that might suggest which prior life experiences to target with biodata item content. Alternatively, one would also expect interviews targeting prior life experiences of high and low performing incumbents to generate criterion valid biodata items. Unfortunately, theories linking content of the ATCS

job domain (or any other job domain) to individual difference characteristics are not abundant (e.g., Burke & Pearlman, 1988). We briefly describe alternate techniques targeting ATCS performance criteria in generating biodata items. Readers are referred to Dean et al. (1999), Mumford, Costanza, Connelly, and Johnson (1996), Russell (1994) for descriptions of theory-based item generation efforts.

Biodata items targeting ATCS performance criteria could be developed by harvesting information about prior life experiences from job incumbents and subject matter experts (e.g., superiors who previously held the position). Such information might be harvested using a number of methods, though answers to structured questions delivered in the context of an interview, focus group discussion, or writing assignment seem to be the most efficient. A basic approach would involve the following steps. First, identify key job requirements using standard job analysis procedures. Currently, there is no reason to believe behavioral, task, or skill requirement information is more or less appropriate for the steps to come.

Second, communicate key job requirements to subject matter experts (SMEs). SMEs should have first hand knowledge of job performance and be as representative of the future applicant pool as possible. Again, no one method is clearly preferable to another - focus group discussions, one-on-one interviews, or standardized written descriptions have been used. Third, ask SMEs to describe prior life experiences they feel influenced performance of these key job dimensions. Russell et al. (1990) asked midshipmen at the U.S. Naval Academy to write life history essays describing such experiences. Russell

(1990, 2000) used structured one-on-one interviews. Again, there is no evidence suggesting one particular method of “harvesting” prior life experiences is more or less deficient or contaminated.

Fourth, extract critical incidents from SME prior life event descriptions. Items are generated from these critical incidents to reflect a) behaviors engaged in by the candidate (or others) in the incident, b) key aspects of the situation circumstances (i.e., sources of assistance, obstacles, etc.), c) affect, attitudes, or feelings associated with the event, d) role responsibilities held by the candidate during the event, or e) task outcomes accomplished. As noted above, there is some debate in the literature regarding alternate taxonomies of biodata “types,” with some authors taking strong positions for or against certain types. For example, Stricker (1987) suggested items tapping experiences over which candidates had no control are not “fair,” while Gandy, Outerbridge, Sharf & Dye (1989) advocating using only those items for which candidate answers could be verified. We adopt Mael’s (1991) position, i.e., that any item demonstrating criterion validity and does not demonstrate differential prediction (i.e., is not biased under the Cleary, 1968, model used by the EEOC Uniform Guidelines) should be retained.

Empirical keying options. Once items have been developed, it is necessary to determine how to score the biodata instrument. We will focus our discussion on the scoring technique of empirical keying. Empirical keying assigns weights to individual items or response options, summing their product to form an overall biodata score used to predict a criterion of interest (e.g., job performance, absenteeism, etc.). Empirical keying

is used quite frequently in biodata selection systems, though there is no reason this scoring technique could not be applied to other selection devices.

A wide variety of empirical keying methods are available (Devlin, Abrahams, & Edwards, 1992). Regardless of the method, there are a number of issues regarding the sample used to empirically score biodata. First, large samples are needed to obtain reliable results. There are no rules of thumb to estimate sample size, though Hogan (1994) recommended that there be at least 5 to 10 persons per item or response option. The larger the sample used, the more stable the weights generated. Russell and Dean (1995) showed the rate of decrease in r when the sample size used was gradually decreased. They found samples size in the range $N = 500$ to $1,000$ yielded cross-validities approximately 75-85% as large as those found with $N = 5000$. Second, the sample used should come from a group similar to the reference group of interest. For example, an empirical key for air traffic controller applicant selection should use a group of air traffic controller applicants or possibly new air traffic control hires. Third, the sample should be as demographically diverse as possible to determine if items predict equally well across demographic subgroups.

The sample completing a biodata instrument is typically randomly divided into two sub-samples: a key development sample and a cross-validation sample. Key development samples are typically larger (e.g., approximately 3/4 of total sample) and used to develop item or response option weights. The cross-validation sample (e.g., remaining 1/4) is held aside and used to estimate a cross-validity that is independent of the original subsample

from which the weights were developed. Cross-validation is necessary to avoid inflation of criterion validity estimates due to capitalization on chance variations in the key development sample from which the weights were derived.

A brief review of response option criterion-referenced empirical keying at the response option level is presented below. Response option empirical keying treats each individual item response option as a single dichotomous “item.” A criterion-referenced response option-based key weights each individual response option by its correlation with the criterion as estimated in the key development sample. These weights are typically re-validated at least every three years. The most efficient means of capturing the strength of this relationship is the point biserial correlation (r_{pb}). The r_{pb} a special case of the Pearson product-moment correlation (r) applicable when correlating a truly dichotomous variable (e.g., response options either chosen or not) with a continuous variable (e.g., a performance measure). An illustrative scoring equation yielding a respondent’s score using this keying procedure is shown below for a 100-item biodata inventory:

$$\begin{aligned} \text{Individual Biodata Score} = & (r_{1,1} * ro_{1,1}) + (r_{1,2} * ro_{1,2}) + (r_{1,3} * ro_{1,3}) + (r_{1,4} * ro_{1,4}) + \\ & (r_{2,1} * ro_{2,1}) + (r_{2,2} * ro_{2,2}) + (r_{2,3} * ro_{2,3}) + (r_{2,4} * ro_{2,4}) + \dots + (r_{100,1} * ro_{100,1}) + \\ & (r_{100,2} * ro_{100,2}) + (r_{100,3} * ro_{100,3}) + (r_{100,4} * ro_{100,4}) \end{aligned}$$

Where $r_{A,B}$ = correlation between item A’s response option B and the criterion in the key development sample
 $ro_{A,B}$ = Which equals: 0 if respondent did not chose item A’s response option B, or 1 if respondent chose item A, response option B

Very simply, if a respondent does not choose a response option, that response option’s weight does not enter into the respondent’s biodata score. Weights of response options that were selected (i.e., correlations between the response option and the criterion of interest)

are added to all other response options chosen to obtain the respondent's overall biodata score. The overall biodata scores for individuals in the hold out sample are then correlated with the criterion to obtain cross validities.

Recommendations for future practice

We have summarized below what we feel constitute the more useful practices and guidelines for development and use of biodata instruments gleaned from the literature:

Item Development

1. Develop items focus on past life experiences, which ideally can be verifiable (to decrease the likelihood of faking), and are job relevant.
2. Provide a rationale explaining one or more links between biodata item content and the criterion.
3. Identify key job requirements using job analysis procedures
4. Ask subject matter experts (SMEs) to describe prior life events that they feel influenced performance on the key job dimensions
5. Generate items based on critical incidents from SME prior life event descriptions.

Key Development

1. Obtain a sample large enough to use at least $N = 500$ subjects in the key development subsample.
2. Key development sample should be as demographically diverse as possible
3. Derive response option weights using each response option's correlation with the criterion of interest.
4. Re-validate weights in empirical key at least every three years.
5. Estimate criterion validity for all response options. Exclude responses demonstrating differential response patterns for minority and majority groups. If reduction in adverse impact is high and decrement in criterion validity is low, use key with only non-adverse impact response options.

Application

1. Criterion validity is likely to be maximized if biodata (and other non-cognitive ability measures) are combined with a cognitive ability measure.
2. Deletion of response options with differential minority/non-minority selection frequencies tend to drastically reduce adverse impact.
3. Addition of a cognitive ability measure will increase overall criterion validity and likelihood of adverse impact.

Personality

Unlike cognitive ability measures' long tradition in air traffic selection and performance prediction, evidence supporting use of personality measures in personnel selection is not as prevalent. Regardless, individuals who work closely with air traffic controllers would likely agree there are some common temperamental qualities. This observation prompts the assumption that if these aspects of personality can be measured, we should be able use this information to select air traffic controllers.

Two things have prevented personality from being used for personnel selection generally. The first is a controversy over whether personality characteristics shared by successful air traffic controllers are a function of the people doing the job, are driven by job characteristics, or some combination thereof. At the extremes, one explanation implies individuals bring certain temperamental elements to the work environment while the other posits individual difference characteristics are generated by the work environment. Second, construct definition and measurement have also complicated personality inventory use.

This section first reviews issues associated with personality construct definition and theory development, with a focus on current trends. Second, we present an overview of research and practice regarding air traffic controller selection and personality tests. Research summarized will tend to rely heavily on efforts initiated in United States perspective, though we also touch on personality test use in systems elsewhere.

Personality Measurement and Theory. Modern personality theory is roughly 50 years old. Hall and Lindzey (1959) noted significant influences came from four key areas

of psychology: clinical, Gestalt, experimental, and psychometric. Allport (1937) defined personality as a new field of study with the defined objective of clarifying things that make individuals different from one another as opposed to the then popular view of “controlling” individual differences to study that which was common to us all. Allport’s (1937) own definition of personality emphasized the dynamic component of individuals’ psychophysical adjustment to the environment. Murray (1938) focused more on the importance of emotion and motivation to personality. Today, the defining aspects of personality are encompassed by these two seminal approaches.

Pervin (1990) noted many topics have historically occupied central positions in personality research. For example during the 1950’s a large body of literature accumulated defining the nature and antecedents of authoritarian personalities, though this topic receives relatively little attention today. Other topics endured as central questions for personality research. One of these, the idiographic vs. nomeothetic controversy pits methods of examining personality against one another. Idiographic approaches use the extensive study of individuals over time. Nomeothetic approaches examine key personality characteristics or traits from psychometric assessment of cross-sectional sample’s responses to personality inventories. Other important topics include the definition and differentiation of self from others, the extent of internal vs. external control, levels of conscious vs. unconscious involvement, and the impact of heredity and the environment. Personality mutability (a form of the old “state vs. trait” debate) and the connection with motivated thought and behavior are two especially popular issues at present (Pervin, 1990).

Interactional perspectives (Magnusson, 1990) and links between personality and purposive behavior (Cantor & Zirkel, 1990) and motivation (Koestner & McClelland, 1990) are extensively reviewed elsewhere.

Guion (1998) noted “personality is a mixture of values, temperament, coping strategies, character, and motivation, among other things” (p. 134). Guion (1998) also suggested it is relatively easy to identify theory-based linkages between personality constructs and job characteristics, though operational definitions are often difficult to obtain. This is due in part to conclusions drawn by Guion and Gottier (1965) over 35 years ago suggesting personality measure criterion validity for personnel selection had not been demonstrated, subsequently stunting research examining new operationalizations, incremental criterion validity, etc. (Guion, 1998).

More recently, Barrick and Mount (1991) hypothesized the primary reason for consistently low personality criterion validities was due to a lack of a well-accepted taxonomy for classifying personality traits. Similar to limitations imposed by biodata item taxonomies, absence of an underlying model of personality characteristics also slowed research and application of personality tests in personnel selection. Historically, numerous personality taxonomies have been proposed, perhaps the most notable of which are those of Cattell (1943), Eysenck (1944), and Guilford (1975). Unfortunately, these taxonomies were not accompanied by strong empirical support.

A new taxonomy has gained widespread support over the last 20 years, viewing personality in terms of five factors that have been repeatedly observed across investigators

and datasets. Barrick and Mount (1991) noted the “Big-5” are robust across different instruments, cultures, sources of ratings, and samples. Importantly, McCrae and Costa (1987) found these factors were relatively independent of cognitive ability. Researchers, however, are not in agreement on the universality of the Big-5 (Hogan, 1986) or this framework’s ability to predict criteria (Hough, 1992), Guion (1998) noted evidence suggests personality measures may constitute a useful predictor domain targeting important criteria that are not captured by more traditional abilities measures.

Various names have been applied to each of the Big-5 factors. The first factor is most often labeled “Extroversion” or “Surgency,” encompassing traits contributing to individuals’ social adaptability and interpersonal involvement. Other trait-labels often applied to the first factor include assertiveness, power, and activity. The second factor also pertains to social traits, though it focuses on individuals’ likeability, friendliness, and sociability, and is often referred to as “Agreeableness” (Guion, 1998).

The last three Big-5 factors relate to more internal individual difference characteristics. The third factor, “Conscientiousness,” is made up of traits including the will to achieve, dependability, task interest and dedication, and personal constraints. As Barrick and Mount (1991) and Guion (1998) pointed out, Conscientiousness is in many ways the central personality component in predicting subsequent task performance.

A fourth factor is generally referred to as “Emotional Stability” or “Neuroticism” and pertains to individuals’ levels of emotional control, anxiety, or general affect. Ability to cope with life stress is related more to this factor than the others. The last Big-5 factor

has the most variation in labels due in large part to a variety of associated traits. The factor, “Openness to Experience,” is characterized by “a liking for thinking about things” of a cultural or personal matter (Guion, 1998, p. 137) and problems to be solved or things to be created. Cattell (1949) argued this factor encompassed aspects of general mental ability (Pervin, 1990). Guion (1998) suggested the term “Intellectance” may be the most appropriate label for this trait.

Personality Criterion Validity Evidence. Recent meta-analyses by Barrick and Mount (1991, 1993) summarized results of 162 Big-5 criterion validities reported over a thirty year period (samples ranged from N = 13 to 1,401, total N = 23,994). Barrick and Mount (1991) sorted criterion validities across three job performance criteria and five occupational groups. Conscientiousness demonstrated consistent relationships with all performance criteria for all groups. Extroversion/ Surgency was related to all performance criteria for management and sales occupations, while Extroversion/Surgency and Intellectance were valid predictors for training performance. In a separate meta-analysis, Tett, Jackson, and Rothstein (1991) found Agreeableness to exhibit low to moderate levels of prediction for many job relevant criteria.

A number of authors noted the Big-5 do not constitute a comprehensive model of personality (Guion, 1998; Hough, 1992). The main limitation identified by these authors focused on the Big-5 as a system for description of personality structure - Big-5 models do not describe processes by which personality characteristics influence relevant job performance criteria or processes by which Big-5 personality characteristics evolve and

develop. The Big-5 model of personality does provide a starting point for future research and application. Perhaps most important has been its stimulation of renewed interest in the value of personality as a performance predictor.

Personality Research in Air Traffic Controller Selection

A number of studies using personality as a predictor of performance-based criteria for air traffic controllers have been reported over the last thirty years. Karson and O'Dell (1970) examined relationships between personality factors measured by Cattell's 16PF (Cattell, 1949) and job performance ratings for a group of 264 controllers. They reported no significant correlations between personality profile scores and job performance ratings. In light of the small sample size and possible range restriction on the profile scores, the poor observed criterion validities were consistent with other studies being conducted around the same period (i.e., studies yielding the pattern of results which prompted Guion & Gottier's, 1965, conclusion).

More recent efforts suggest a more favorable verdict for personality as a performance predictor. Using the State-Trait Personality Inventory (Spielberger, 1979), Nye and Collins (1991) found male and female Air Traffic Control trainees (N = 1,284) exhibited less anxiety and anger than normative groups of college students and Navy recruits. Another important finding was students who had higher than average anxiety and anger scores were more likely to fail at the FAA air traffic controller academy.

A similar study (Nye, Schroeder, & Dollar, 1994) investigated scores from Jenkins Activity Survey (Jenkins, Zyzanski, & Rosenman, 1979) for 474 Air Traffic Control

trainees, focusing on prevalence of Type A behavior patterns in air traffic control students. Though the study found no relationship between achievement striving and FAA Academy performance, students in Air Traffic Controller training courses demonstrated higher incidence of Type A behavior than a normative sample.

A more recent study investigating 16PF scores of post-strike FAA Academy trainees (Schroeder & Dollar, 1997). Air Traffic students exhibited less anxiety, higher self-discipline, higher emotional stability, and were more self-reliant and assertive than normative samples. In the same study, data originally gathered by Karson and O'Dell (1970) were reexamined. The same pattern of ATC student profile characteristics was found.

Schroeder, Broach, and Young (1993) examined relationships between personality and FAA Academy performance using a measure explicitly developed to tap Big-5 construct domains. Using the NEO Personality Inventory (Costa & McCrae, 1985), Schroeder et al. found Air Traffic students (N=1,030) exhibited lower than average Neuroticism scores and higher than average Extroversion, Intellectance (Openness to Experience) and Conscientiousness scores than normative samples. They also found Big-5 measures predicted significant incremental performance variance over measures of cognitive ability.

The bulk of these studies suggest Big-5 personality dimensions constitute a viable avenue for additional research and application in Air Traffic Controller selection. Personality measures are currently being used for Air Traffic Controller selection in the United States, Germany, Sweden, and the United Kingdom (Broach & Manning, 1997). In the United States, personality assessment has been

formally used since 1965 as part of the medical screening program (Convey, 1984) for Air Traffic Controllers. An empirical key using 38 items from the Sixteen Personality Factor Questionnaire (16PF) was designed to target potential anxiety disorder and used to refer screened applicants for more extensive psychiatric and psychological evaluation. Importantly, Pickrel (1984) reported between one and two percent of all applicants warranted closer examination and that subsequently half of these were medically disqualified from service.

Personality measures are being used successfully in the selection of air traffic controllers because of their criterion validity with performance measures and as a “flag” for those who might have difficulty succeeding in an occupation where stress levels can be high. Research suggests personality may prove valuable in additional areas. Where cognitive abilities may be more predictive of core technical competence, personality (and biodata) may be more relevant to what Borman and Motowildo (1993) termed “contextual performance.” Pending further research, personality may prove more useful in career counseling situations.

Conclusion

Biodata and personality measures both appear to hold promise as performance predictors. Progress has been made in both predictor literatures, making them viable contributors to efforts aimed at predicting air traffic controller job performance. Additional research on these non-cognitive devices is warranted to examine how well biodata and personality measures complement or supplement general cognitive ability measures in selection of air traffic controllers.

References

- Allport, G.W. (1937). Personality: A psychological interpretation. New York: Holt, Rinehart, & Winston.
- Asher, J. J. (1972). The biographical item: Can it be improved? Personnel Psychology, 25, 251-269.
- Atwater, D. C. (1980). Faking of an empirically keyed biodata questionnaire. Paper presented at the annual meeting of the Western Psychological Association, Honolulu, HI.
- Barge, B. N. & Hough, L. M. (1988). Utility of biographical data for the prediction of job performance. In L. M. Hough (Ed.), Literature review: Utility of temperament, biodata, and interest assessment for predicting job performance (ARI Research Note 88-020). Alexandria, VA: U. S. Army Research Institute.
- Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. Personnel Psychology, 44, 1-26.
- Barrick, M. R., & Mount, M. K. (1993). Autonomy as a moderator of the relationships between the Big Five personality dimensions and job performance. Journal of Applied Psychology, 78, 111-118.
- Borman, W. C., & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt, W. C. Borman, & Associates (Eds.), Personnel selection in organizations (pp. 71-98). San Francisco: Jossey-Bass.

Burke, M. J. & Pearlman, K. (1988). Recruiting, selecting, and matching people with jobs. In J. P. Campbell & R. J. Campbell (Eds.), Productivity in organizations. (pp. 97-142). San Francisco: Jossey-Bass.

Broach, D., & Manning, C. A. (1997). Review of air traffic controller selection: An international perspective. (DOT/FAA/AM-97/15). Washington, DC: FAA Office of Aviation Medicine.

Cantor, N., & Zirkel, S. (1990). Personality and purposive behavior. In L. A. Pervin (Ed.), Handbook of personality (pp. 369-395). Palo Alto, CA: Guilford.

Cascio, W. F. (1975). Accuracy of verifiable biographical information blank responses. Journal of Applied Psychology, 60, 767-769.

Cattell, R. B. (1949). The 16 PF Questionnaire. Champaign, IL: Institute for Personality and Ability Testing. Cascio, W. F. (1975). Accuracy of verifiable biographical information blank responses. Journal of Applied Psychology, 60, 767-769.

Cleary, T. A. (1968). Test bias: Prediction of grades of Negro and white students in integrated colleges. Journal of Educational Measurement, 5, 115-124.

Collins, W. E., Manning, C. A., & Taylor, D. K. (1984). A comparison of prestrike and poststrike air traffic control specialist trainees: Biographic factors associated with academy training success. In Studies of poststrike air traffic control specialist trainees: I. Age, biographic factors, and selection test performance related to academy training success (Report No. FAA-AM-84-6). Washington, DC: Federal Aviation Administration, Office of Aviation Medicine.

Collins, W. E., Nye, L. G., and Manning, C. A. (1990). Studies of poststrike air traffic control specialist trainees: III. Changes in demographic characteristics of academy entrants and biodemographic predictors of success in air traffic controller selection and academy screening (Report No. FAA-AM-90-4). Washington, DC: Federal Aviation Administration, Office of Aviation Medicine.

Convey, J. J. (1984). Personality assessment of ATC applicants. In S. B. Sells, J. T. Dailey, E. W. Pickrel (Eds.) Selection of air traffic controllers (pp. 323 - 351) (DOT/FAA/AM-84/2). Washington, DC: FAA Office of Aviation Medicine.

Costa, P. T., & McCrae, R. R. (1985). The NEO Personality Inventory manual. Odessa, FL: Psychological Assessment Resources.

Dean, M. A. (1999). On biodata construct validity, criterion-related validity, and adverse impact. Unpublished doctoral dissertation. Louisiana State University, Baton Rouge.

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 resource management. Greenwich, CT: JAI Press.

Devlin, S. E., Abrahams, N. M., & Edwards, J. E. (1992). Empirical keying of biographical data: Cross-validity as a function of scaling procedure and sample size. Military Psychology, 4, 119-136.

Gandy, J. A., Outerbridge, A. N., Sharf, J. C., & Dye, D. A. (1989). Development and initial validation of the Individual Achievement Record. Washington, D. C: U. S. Office of Personnel Management.

Guion, R. M. (1998). Assessment, measurement, and prediction for personnel decisions. Mahwah, NJ: Erlbaum.

Guion, R. M., & Gottier, R. F. (1965). Validity of personality measures in personnel selection. Personnel Psychology, 18, 135-164.

Hall, C. S., & Lindzey, G. (1959). Theories of personality. New York: Wiley.

Hogan, J. B. (1994). Empirical keying of background data measures. In G. S. Stokes, M. D. Mumford, & W. A. Owens (Eds.), Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction. (pp. 69-108). Palo Alto: Consulting Psychologists Press.

Hogan, R. (1986). Manual for the Hogan Personality Inventory. Minneapolis: National Computer Systems.

Hough, L. M. (1992). The “Big Five” personality variables-construct confusion: Description versus prediction. Human Performance, 5, 139-155.

Hough, L., Eaton, N. K., Dunnette, M. D., Kamp, J. D., McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. Journal of Applied Psychology, 75, 581-595.

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

Jenkins, C. D., Zyzanski, S. J., & Rosenman, R. H. (1979). Jenkins Activity Survey (Form C). San Antonio, TX: Psychological Corporation.

Karson, S., & O'Dell, J. W. (1970). Performance ratings and personality factors in radar controllers. (DOT/FAA/AM-70/14). Washington, DC: FAA Office of Aviation Medicine.

Koestner, R., & McClelland, D. C. (1990). Personality and intrinsic motivation. In L. A. Pervin (Ed.), Handbook of personality (pp. 445-496). Palo Alto, CA: Guilford.

Latham, G. P., Saari, L. M., Pursell, E. D., Campion, M. A. (1980) The situational interview. Journal of Applied Psychology, 65, 422-427.

Mael, F. A. (1991). A conceptual rationale for the domain and attributes of biodata items. Personnel Psychology, 44, 763-792.

Magnusson, D. (1990). Personality from an interactionist perspective. In L. A. Pervin (Ed.), Handbook of personality (pp. 347-367). Palo Alto, CA: Guilford.

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. Journal of Personality & Social Psychology, 52, 81-90

Mitchell, T. W. (1994). The utility of biodata. In G. S. Stokes, M. D. Mumford, & W. A. Owens (Eds.), Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction (pp. 485-516). Palo Alto, CA: Consulting Psychologists Press.

Mosel, J. L. & Cozan, L. W. (1952). The accuracy of application blank work histories. Journal of Applied Psychology, 36, 356-369.

Mumford, M. D., Costanza, D. P., Connelly, M. S., & Johnson, J. F. (1996) Item generation procedures and background data scales: Implications for construct and criterion-related validity. Personnel Psychology, 49, 361-398.

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.

Mumford, M. D. & Stokes, G. S. (1992). Developmental determinants of individual action: Theory and practice in applying background measures. In M. D. Dunnette (ed.), Handbook of Industrial and Organizational Psychology (2nd ed., vol. 3, pp. 61-138). Palo Alto, CA: Consulting Psychologists Press.

Murray, H. A. (1938). Explorations in personality. New York: Oxford University Press.

Nye, L. G., & Collins, W. E. (1991). Some personality characteristics of air traffic control specialist trainees: Interactions of personality and aptitude test scores with FAA Academy success and career expectations. (DOT/FAA/AM-91/8). Washington, DC: FAA Office of Aviation Medicine.

Nye, L. G., Schroeder, D. J., & Dollar, C. S. (1994). Relationships of Type A behavior with biographical characteristics and training performance of air traffic controllers. (DOT/FAA/AM-94/13). Washington, DC: FAA Office of Aviation Medicine.

Owens, W. A. (1976). Background data. In M. D. Dunnette (Ed.), Handbook of industrial and organizational psychology (pp. 609-644). Chicago: Rand McNally.

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

Pervin, L. A. (1990). Personality: A historical perspective. In L. A. Pervin (Ed.), Handbook of personality (pp. 3-43). Palo Alto, CA: Guilford.

Pickrel, E. W. (1984). Research contributions at the Office of Aviation Medicine (OAM). In S. B. Sells, J. T. Dailey, E. W. Pickrel (Eds.) Selection of air traffic controllers (pp. 113-117) (DOT/FAA/AM-84/2). Washington, DC: FAA Office of Aviation Medicine.

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.

Russell, C. J. (1990). Selecting top corporate leaders: An example of biographical information. Journal of Management, 16, 71-84.

Russell, C. J. (1994). Generation procedures for biodata items: A point of departure. In G. S. Stokes, M. D. Mumford, & W. A. Owens (Eds.). Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction. (pp. 17-38). Palo Alto: Consulting Psychologists Press.

Russell, C. J. (1996). Toward a model of life experience learning. In R. B. Stennett, A. G. Parisi, & G. S. Stokes (Eds.) A compendium: Papers presented to the first biennial biodata conference (pp. 17-31), Applied Psychology Student Association, University of Georgia: Athens, GA.

Russell, C. J. (2000). A longitudinal study of top-level executive performance. In C. J. Russell (chair), Identification and Development of Top Level Organizational Leaders. Symposium conducted at the 13th annual meetings of the Society of Industrial and Organizational Psychology, New Orleans, LA.

Russell, C. J. & Dean, M. A. (1995, May). Bootstrapping estimates of response option validity: Theory development and theory testing through empirical keying. In C. J. Russell (chair), Leveraging empirical keying procedures advance theory in performance prediction. Symposium presented at the meeting of the Society for Industrial and Organizational Psychology, Orlando, FL.

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*, 569-580.

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

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., 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.

Schroeder, D. J., Broach, D., & Young, W. C. (1993). Contribution of personality to the prediction of success in initial air traffic control specialist training. (DOT/FAA/AM-93/4). Washington, DC: FAA Office of Aviation Medicine.

Schroeder, D. J., & Dollar, C. S. (1997). Personality characteristics of pre/post-strike air traffic control applicants. (DOT/FAA/AM-97/17). Washington, DC: FAA Office of Aviation Medicine.

Spielberger, C. D. (1979). Preliminary manual for the State-Trait Personality Inventory. Tampa, FL: Human Resources Institute, University of South Florida.

Stricker, L. J. (1987). Developing a biographical measure to assess leadership potential. Paper presented at the annual meeting of the Military Testing Association, Ottawa, Ontario.

Stricker, L. J. (1988). Assessing leadership potential at the Naval Academy with a biographical measure. Paper presented at the annual meeting of the Military Testing Association, San Antonio, TX.

Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analysis. Personnel Psychology, *44*, 703-720.

U. S. Employment Service (1970). Manual for the USES General Aptitude Test Battery, Section III: Development. Washington, DC: U. S. Department of Labor, Manpower

Administration.

Wernimont, P. F. & Campbell, J.P. (1968). Signs, sample, and criteria. Journal of Applied Psychology, 52, 372-376.