Psychology from the Margins

Volume 3 *Psychology from the Margins: Volume 3 (2021)*

Article 5

2021

Biographical Data and Black Box Empiricism: Lessons Learned for Algorithmic Assessments in Personnel Selection

Ketaki Sodhi
The University of Akron

Marc Cubrich
The University of Akron

Introduction

The practice of using biographical data in personnel selection, commonly referred to as biodata, has a history that spans well over 100 years (Stokes, 1999). Although there has historically been definitional disagreement about what constitutes biodata, in the hiring or employee selection context it has traditionally been defined as historical and verifiable pieces of information about an individual's personal background and life history that are typically collected in the application process (Anderson & Shackleton, 1990). For example, factual information about an applicant can include their age, marital status, previous jobs, years of experience in previous roles, and level of education. As personnel selection practices, technology, as well as legal, technical, and practical requirements have changed dramatically over the past 100 years, so has the measurement, application, and theory surrounding biodata.

Although biodata items can be developed using rational or theory-driven approaches that focus on specific constructs (Mumford et al., 1996; Oswald et al., 2004; Reiter-Palmon & Connelly, 2000; Stricker & Rock, 1998), empirically-derived keys are often used to score items and improve prediction because of their relationships with important criteria such as job performance (Cucina et al., 2012; Cucina et al., 2013; Karas & West, 1999; Mitchell & Klimoski, 1982; Reiter-Palmon & Connelly, 2000). Such an approach is often criticized as employing "black box" or "dust bowl" empiricism, therefore lacking in psychological meaning, job relatedness, and legal defensibility in the hiring or employee selection context. The term "black box" is used in the sciences to describe a complex device for which we know the inputs and outputs, but not the mechanisms or innerworkings that produce the outcomes. As it relates to applications of artificial intelligence (AI) and machine learning, we may know the inputs (i.e., training data used to build these models) and the resulting output (i.e., predictions based on the model), but the actual innerworkings of the machine remain opaque and unknowable by human observers.

As it relates to biodata, for instance, a certain zip code reported on a biodata item could be correlated with job performance and based on that criterion, only candidates who live in that area may be selected. Not only can the use of such information result in discrimination, but it also goes against several employment laws**.** As the popularity of biodata in selection assessments grew in the 1980s and 1990s, many attempts to develop biodata theories and guide the development of biodata items were introduced (e.g., Hough & Paulin, 1994; Mael, 1991; Mumford & Owens, 1987; Mumford et al., 1990; Nickels, 1994; Schoenfeldt, 1999; Schmitt et al., 1999; Steinhaus & Waters, 1991). The insights gleaned from this body of research are increasingly relevant in the new age of big data, AI, and machine learning. More than ever, AI and machine learning are being used to score

candidates and make hiring recommendations. Although there are several ways to integrate AI and machine learning, many organizations are using data-driven approaches, which are frequently atheoretical and usually based on correlations or pattern recognition. As a result, even if information about protected classes is removed (e.g., race and gender), algorithms often detect factors that correlate with these protected groups and can indirectly propagate bias in personnel selection.

These issues speak directly to the role of industrial and organizational (I-O) psychologists to promote fairness, equity, and unbiased decision making in organizations. Although more overt discriminatory treatment such as segregation and wage disparities based on protected class (e.g., sex and race) have declined since the introduction of the Civil Rights Act of 1964, certain groups remain underrepresented in high-paying and high-status jobs (Berdahl & Moore, 2006). For instance, Fortune magazine reported that 73% of senior executives at Fortune 500 companies are White compared to the 21% who are Asian, 3% Latino, and 0.6% Black (Jones, 2017). Additionally, women as well as racial and ethnic minorities have been shown to consistently earn less compared to their male and white counterparts. Studies suggest that on average women earn 82 cents for every dollar men earn for comparable work, and men of color on average earn less than White men but outearn women within their race or ethnic groups for comparable work (Payscale, 2021). Inherent in the selection process is the consideration of several legal issues, regulatory bodies, and guidelines. The Civil Rights Acts of 1964 deals with a variety of discrimination issues, while Title VII of this act is directed specifically at employment issues. According to Title VII, employers cannot discriminate based on color, national origin, race, religion, sex. In order to help organizations, abide by Title VII, the Equal Employment Opportunity Commission (EEOC) was created. In addition to key laws regarding employment practices, The Uniform Guidelines on Employee Selection Procedures (1978) and Principles for the Validation and Use of Personnel Selection Procedures (1987, 2003) provide guidance on the implementation of scientifically valid assessments and fair practices in the workplace. I-O psychologists are often directly responsible for ensuring that organizational practice abide by existing employment discrimination laws and the Uniform Guidelines.

Ostensibly, the inherent problems with AI-powered assessments and algorithmic scoring are akin to the historical problems with biodata. Similar to the way in which I-O psychologists help ensure the appropriate use of biodata, I-O psychologists can help ensure the appropriate use of data, legal defensibility, and optimal outcomes for organizations using these novel approaches. In light of these parallels, the present paper examines the long history of biodata use in personnel assessment, drawing parallels between past and present, and identifying lessons learned and their implications for applications of machine learning and AI in the

hiring process. Finally, we provide recommendations for today's problems based on the lessons learned from past biodata research.

Overview of the Hiring Process

While every organization has their own unique hiring process, there are generally four key stages in most hiring processes: sourcing, screening, interviewing, and selection (Bogen & Rieke, 2018). As noted, I-O psychologists play a critical role in helping organizations design hiring practices that are not only valid and predict who will be successful on the job but also minimize bias at each one of these steps. Sourcing involves generating a pool of candidates for a specific job. This is usually done by posting job requisitions on multiple different platforms (e.g., job boards, LinkedIn) as well as searching for, identifying, and then reaching out to desirable candidates. Screening is the next step which consists of pre-employment assessments that can help screen out unsuitable applicants and refine the candidate pool. In other words, pre-employment screening tools and assessments include procedures that can help make preliminary evaluations of candidates in order to narrow the hiring funnel. The next step is interviewing, which involves conducting a more in-depth evaluation of candidates through structured conversations with multiple key stakeholders (e.g., manager, team, etc.). This process can help determine who the best candidates for the job are.

Finally, there is the selection or final evaluation stage, which consists of incorporating the results of all the pre-employment assessments and interviews to make a final decision about whom to select for the role. It is important to remember that the hiring process is cumulative, and decisions made at every step of the process can affect the quality of hire and overall fairness of the process. Historically, many groups have been systematically underrepresented within higher levels of the workforce and this is partly a result of unintended bias baked into commonly used pre-employment assessment or evaluation processes from the past. While several new high-tech tools are emerging to help tackle this problem, many of them also have the potential to propagate unintended bias, in turn undermining their value as fair and reliable.

Given the potential for unintended bias, we believe that revisiting the critiques of some of these older tools and incorporating the lessons learned into design and administration of new pre-employment assessment procedures is key to actually increasing fair and diverse representation in the workplace. As noted, in this paper we will focus primarily on pre-employment assessments, highlighting the similarities between historically popular, yet highly controversial methods like biodata assessments and new methods like algorithmic pre-employment assessments. In doing so, we will focus on how their development can influence potential for unintended bias or discrimination and the ethical and legal

implications of this practice. Lastly, we will use lessons learned from the use of biodata assessments to inform our recommendations for algorithmic and AI-powered assessments.

The History of Biodata in Personnel Assessment

Although the form and function of biodata assessments have changed drastically over the past century, the use of biodata is not new. For example, an early application of biodata was the selection of successful sales personnel in the life insurance industry (Goldsmith, 1922). According to a report from the International Personnel Management Association (IPMA; van Rijn, 1992), biodata has been collected in the private sector to predict a diverse set of constructs over the years including job performance (Walther, 1961), employee turnover (Cascio, 1975), managerial effectiveness (Laurent, 1962), creativity (Buel et al., 1966), vocational interests (Mumford & Owens, 1982), student achievement (Freeberg, 1967), credit risk (Moran et al., 1968) honesty (Rosenbaum, 1976), training success (Drakeley et al., 1988), and career success (Childs & Klimoski, 1986).

Scholars have pointed to the Weighted Application Form (WAF) as an early predecessor to biodata assessments as we know them today (Anderson & Shackleton, 1990). Similar to biodata items, the WAF is a form whose responses are assigned a numerical value. The items used in a WAF of course depend on their context and application, but these items typically seek to gather information and assign values in order to make a decision. Over 90 years ago, research demonstrated that the systematic analysis of items on a standard job application form could be used to predict the chances of candidates being successful in the job (Anderson & Shackleton, 1990). This procedure is similar to what is used for credit card applications or the determination of insurance rates. For example, insurers may empirically derive that young drivers, people with fast cars, and those who live in large cities are all at greater risk of an accident, and thereby charge them higher premiums. These assumptions are based on systematic evidence, and WAFs work on the same principle. Biodata assessments operate on a similar principle but differ in that they are typically tailored for a particular job and use a multiple-choice questionnaire that allow for the execution of detailed statistical analyses (Anderson & Shackleton, 1990).

An analysis of the type of information that has historically been collected in biodata questionnaires reveals biodata topics that are not in line with current guidance surrounding employee selection procedures and legal protections for protected classes in the United States. Drawing from a report on the challenges and potential uses of biodata in the public sector from the IPMA in 1992, Table 1 provides a summary of topic areas historically included in biodata questionnaires. Most of the topic areas included in Table 1, such as marital status, national origin,

sex, and family adjustment, would not be suitable or legally acceptable in selection procedures as they currently stand (Civil Rights Act, 1964, 1991; EEOC et al., 1978). The fact that many of the topics represented in Table 1 are not used in current hiring practices is indicative of the role of I-O psychologists in creating assessments that are job-relevant, theoretically grounded, and legally defensible. In other words, historically, I-O psychologists have led the charge to help organizations better understand the tools at their disposal and only use those that are valid, fair, and legal.

Table 1

Topic Areas Historically Included in Biodata Instruments

Personal	Skills
Age	Read/speak non-native language
Marital status	Read blueprints
Number of years married	Ability to type
Dependents, number of Birth order	Repair work on cars
Physical health	Possession of job skills
Time lost from job	Training for target job
Size of hometown	Machines/tools/equipment
Number of times moved	Employment
Time at last address	Type of previous experience
Nationality	Worked while in high school
Weight and height	Number of previous jobs
Sex	Specific work experiences
Background, General	Self-employment
Occupation of parent	Seniority
Military discharge record	Reason for leaving last job
Early family responsibility	Social
Parent family adjustment	Club memberships
Education	Attendance at group meetings
Highest level of education	Offices held
Education level of spouse	Leadership experience
Major field of study	Interests
Subjects liked, disliked	Preference for outside work
Recency of education	Hobbies
Grades, honors, awards	Sports
Socioeconomic	Sources of entertainment
Financial responsibility	Personal/Attitudinal
Phone number for creditors	Willingness to relocate
Loans as a portion of income	Willingness to travel
Monthly mortgage payment	Self-confidence
Debts	Basic personality needs
Net worth	Drive or energy level
Amount of life insurance	Job preferences
Properties/Investments	
Current living expenses	
Earnings expected	

Note. Adapted from "Biodata: Potentials and challenges in public sector employee selection," by P. van Rijn, 1992, *Personnel Assessment Monographs of International Personnel Management Assessment Council*, p. 8. Most of the topic areas included in this table would not be suitable, acceptable, or legally defensible if used for personnel selection procedures.

The 1980s and 1990s marked a turning point in the popularity and use of biodata assessment in personnel decisions. A summary of staff selection decisions

leading into the 1990s remarked the following on biodata assessments, "Despite any problems they are highly cost effective and deserve to be more widely utilised than at present" (Anderson & Shackleton, 1990, p. 6.). Demonstrating the lack of utilization at the time, they pointed to a survey of major companies indicating that only eight percent were using biodata for selection at the time (Robertson & Makin, 1986). The growth and popularity of biodata assessments in the 1980s and 1990s can be attributed to several factors. In 1992, a report from the IPMA indicated that reasons for increased interest in biodata include: (a) the rates of adverse impact associated with traditional written tests, (b) calls to expand the arsenal of predictors beyond the domain of cognitive ability, and (c) promising advances in the understanding and development of biodata questionnaires (van Rijn, 1992).

Further, the IPMA pointed to at least eight additional reasons for the growing interest in biodata assessments at the time: (a) high test-retest reliability, (b) high criterion-related validity, especially for objective criteria such as employee turnover, (c) generalizable validities that are relatively stable over time, across different subgroups, occupations, organizations, and situations, (d) relatively little adverse impact, (e) high cost-effectiveness and easy administration, especially when administered to a large group of applicant, (f) acceptable and non-threatening to most job applicants, and (g) a broad and flexible basis for assessment and increased composite validity when used with other assessments (van Rijn, 1992).

Studies examining the benefits of biodata assessments have continued to show impressive results that support these notions. Research suggests that biodata assessment methods are commonly found to be related to job performance, turnover, and other life outcomes (Barrick & Zimmermann, 2009; Breaugh, 2014; Breaugh et al., 2014; Gessner et al., 1993; Hunter & Hunter, 1984; Mumford et al., 1996; Mumford & Owens, 1987; Oswald et al., 2004; Ployhart et al., 2006; Reilly & Chao, 1982; Rothstein et al., 1990; Schmitt et al., 1984; Schmidt & Hunter, 1998; Siegel, 1956; Stricker & Rock, 1998), and tend to exhibit lower levels of adverse impact than many other selection tools (Hough et al., 2001; Pulakos & Schmitt, 1996; Sharf, 1994).

Algorithmic Pre-Employment Assessments, Artificial Intelligence, and Machine Learning

With advancements in technology, algorithmic pre-employment assessments are becoming increasingly common. For instance, a recent Mercer study (2020) found that 88% of firms from a global sample use AI as a part of their HR processes. Further, the Global Digital Talent Acquisition Industry is estimated to grow at a rate of 8.8%, reaching almost 74 billion USD by 2025 (MarketWatch, 2021). This makes it all more critical to address and refine AI and Machine Learning approaches used to build employment assessments. Such approaches include using

computer-based algorithms to score and screen candidates on various factors in order to decide if candidates should progress through the hiring pipeline. This technology can help automate parts of the hiring process. Several companies now use AI and machine learning to further increase the predictive power of their algorithmics and decide which candidates have the greatest likelihood of success. In other words, these assessment technologies use data (often data that already exists within a company's HR systems) to train algorithms to "learn" what differentiates successful candidates from unsuccessful ones and apply this information to make decisions about which candidates should move on to the next step of the hiring process.

There are several reasons why organizations choose to use algorithmic assessments. Firstly, using algorithmic assessments can help companies save time and money. On average, it takes about 6 weeks for an employer in the U.S. to fill a role (LinkedIn, 2017). Moreover, recruiters and hiring managers often spend several hours on each step in the hiring process. Together, this not only leads to a poorer candidate experience but also to a high cost per hire. Indeed, research suggests that it costs around $4000 to hire a new employee (Society for Human Resource Management, 2016) in the U.S. and talent acquisition departments often feel constrained when it comes to their budgets. In fact, even when projecting an increase in volume from one year to the next, talent acquisition departments often do not expect an increase in funding. All of these concerns increase the desirability of any process that can help make the hiring process more cost effective and less labor intensive.

Moreover, several vendors that design algorithmic or AI-powered assessments suggest that their processes can lead to hiring a better-quality candidate as well as an increase in diversity based on factors such as race, gender, age, etc. They also often suggest that machine learning algorithms can help predict fit which can increase engagement and reduce turnover (Bogen & Rieke, 2018). This potential benefit is important because turnover is an incredibly costly problem for companies to have.

However, while I-O psychologists were central to the conversations around appropriate use of biodata, the world of algorithmic pre-employment assessments is currently dominated by data scientists and computer scientists. This has led to a situation where algorithmic pre-employment assessments are not always the best solution and can be problematic in several ways. In fact, some of the criticisms of biodata assessments are actually relevant in the case of algorithm assessments as well, especially their ethical and legal implications related to diversity and inclusion. In the proceeding sections, we will start by discussing key practical and legal considerations when developing pre-employment assessments, as recommended by extensive research in the I-O psychology literature. Next, we will discuss the main criticisms of biodata-based screening and highlight the similarities

between biodata and algorithmic assessment tools. Finally, we will use lessons learned by I-O psychologists from the historic practice of biodata-based assessments to inform our recommended solutions for algorithmic assessments.

Key Considerations for Developing Pre-Employment Assessments

When developing or selecting a pre-employment assessment approach and considering their implications for diverse representation in the workplace, it is important to first understand the primary pre-employment assessment development approaches as well as the legal foundations of selection assessment.

Data-driven versus theory-driven approaches

Irrespective of the way in which a test is administered (e.g., paper-pencil, game-based), it needs to meet the standards set by these laws and follow development guidelines provided by the EEOC. Because of these requirements, no single "type" of test or assessment inherently increases or decreases discrimination and legal liability. Instead, the way in which a test is developed (including the data used as a part of test development and the scoring rules or algorithms applied to selects candidates) determine the legal implications of the selection test or assessment. In other words, *test development* often plays a bigger role in impacting the validity and bias present in a test than the type of assessment or selection test (i.e., biodata, paper-pencil, game-based, AI-powered, etc.). Assessment development approaches or techniques usually fall somewhere on the continuum from theory-based to data-driven.

Theory-driven techniques utilize organizational psychology theory to identify and measure constructs that are necessary for the job (i.e., job-relatedness) and should be related to job performance. From a practical perspective, this would first involve conducting job analyses or developing competency models to determine the knowledge, skills, abilities, and other key characteristics (KSAOs) required to successfully perform the job. Next, this information would be used to develop assessments that measure these KSAOs in a way that is valid, reliable, and reduces adverse impact. On the other hand, data-driven techniques utilize big data approaches to identify what correlates most strongly with job performance, irrespective of how job-related they are. In other words, large amounts of data on current high performers and candidates can be mined to maximize the ability to predict if a candidate will fall into a certain group of interest. For instance, high performers may be used as benchmarks against which candidates are scored or scoring algorithms may measure in game behavior of high performers and use similarity in playing patterns as a criterion to select incumbents. However, when an algorithm replied on the extent to which a construct correlates with job

performance, "to ask whether the model is 'job related' in the sense of 'statistically correlated' is tautological. The more important question in the context of data mining is what does the correlation mean?" (Kim, 2016, p. 866).

Legal foundations for selection procedures

When it comes to using any form of assessments to make employee selection decisions (e.g., hiring, promotions) there are legal implications. Within the United States, there are currently several laws and guidelines in place to ensure that organizations use valid, reliable, and fair assessment and selection processes. Firstly, the *Uniform Guidelines for Employee Selection Procedures* (EEOC et al., 1978) provide a framework that can help employers not just determine the proper use of selection instruments but also comply with the Federal law which prohibits discrimination in employment practices based on race, color, religion, sex, and national origin. Next, Title VII of the Civil Rights Act of 1964 established anti-discrimination laws by forbidding discrimination based on race, color, religion, sex, and national origin. Additionally, the Age Discrimination in Employment Act (1967) makes it illegal to discriminate on the basis of age and the Americans with Disabilities Act (1990) makes it illegal to discriminate against people with disabilities. Finally, in the last decade and a half, privacy concerns led to additional legislature such as the Genetic Information Nondiscrimination Act (2008) in the United States which prohibits employers from asking about or acquiring genetic information from applicants or employees, and the General Data Protection Regulation (GDPR; 2018) which aims to strengthen right to privacy by regulating the data that organizations can collect about candidates and employees present in the European Union (EU).

Criticisms of Biodata and Algorithmic-based Assessments

Many of the criticisms associated with biodata are dependent upon the approach used to develop, weight, and score biodata items. The current practice of scoring biodata can be broadly differentiated into the three major categories: (a) rational keying, (b) empirical keying, and (c) hybrid blends of these respective approaches (Cucina et al., 2012; Speer et al., 2019). Rational approaches to developing and scoring biodata items use theory to guide item development, selection, and scaling (Reiter-Palmon & Connelly, 2000). Rational approaches seek to ensure relevance, generalizability, and long-term prediction by identifying specific psychological constructs that result in performance and writing or selecting items to reflect those constructs (Reiter-Palmon & Connelly, 2000). Although rational approaches have a clearer theoretical foundation and are potentially more legally defensible, this

method generally produces weaker criterion-related validity when compared to empirical approaches (Cucina et al., 2012; Speer et al., 2019).

Considering the incremental validity gained through empirical keying (Cucina et al., 2012, 2013), this approach remains a widely used method of scoring biodata in personnel selection (Hogan, 1994; Reiter-Palmon & Connelly, 2000; van Rijn, 1992). Although there are a variety of methods to create empirical keys, generally speaking, empirical keys are created by weighting and combining items as a function of their relationships with important criteria such as job performance (Reiter-Palmon & Connelly, 2000; Speer et al., 2019). This approach can maximize the size of the validity coefficients by capitalizing on item-criterion relationships (Mumford & Owens, 1987). Finally, hybrid keying is another common approach that uses both rational and empirical relationships to determine the final response weights (Speer et al., 2019). Such an approach can capitalize on the benefits of each of the aforementioned approaches.

As it relates to implications for today's hiring practices, the lessons learned from research on empirical keying are particularly useful. Akin to many criticisms of AI-based assessments, biodata assessments have historically been criticized as employing "black box" or "dust bowl" empiricism, therefore lacking psychological meaning. As such, the primary concern and criticism of biodata assessments is a theoretical one. An understanding of the theoretical link between a person's life history and their success in different jobs has not been adequately understood (Stokes, 1999). Empirical keying results in robustly predictive assessments at the expense of a theoretically grounded understanding of what the scale is measuring and generalizability to other contexts is generally considered problematic (Speer et al., 2019). In fact, for this exact reason, an inability to explain why a biodata test is related to job performance can create concerns for legal defensibility.

This concern is mirrored in the use of AI and algorithmic-based assessments. Although these novel assessments may maximize prediction, the theoretical basis of the information being used to make decisions is often unclear. In other words, data-driven assessment scoring algorithms based on AI and machine learning are often atheoretical, lack sufficient transparency, and are based on correlations between any number of candidate or employee data points and high performance. For instance, the algorithm may detect that having a certain eye color is correlated with high job performance and in turn recommend selecting candidates who have that eye color. However, for most jobs, it is unlikely that eye color is actually "job related" or a cause for high performance. This example illustrates the importance of remembering that correlation is not causation. In other words, just because something like zip code or eye color is correlated with high performance, it may not cause or lead to high performance and thus should not be a factor considered while making selection decisions. Not only can it discriminate against protected groups, it can also have legal implications. Specifically, not being able to

demonstrate that selection criteria is "job related for the position in question and consistent with business necessity" can in some cases make companies vulnerable to legal action under the Civil Rights Act (1991).

The use of empirical criterion-keying can also result in biodata items that lack face validity (i.e., assessments do not appear to be relevant to the job, irrespective of if they actually are or not). Research has demonstrated that applicants react negatively to selection tests with no obvious relationship to job performance, and meta-analytic evidence has demonstrated that applicant reactions are significantly and meaningfully associated with performance on selection tests (Hausknecht et al., 2004). The same concerns can be applied to AI-based assessments. It may be unclear how organizations are using AI or machine learning to make personnel decisions, and applicants may have a general mistrust or unawareness of these practices, thereby introducing the potential for increased anxiety into the hiring process (van Esch et al., 2019). Additionally, research suggests that lack of face validity and the inability to understand how an assessment process relates to the job can actually increase the chances of legal action from applicants (Smither et al., 1993).

Although the use of the biodata topics contained in Table 1 are not commonplace in practice today, an examination of the past reveals the types of extraneous and non-job relevant information that has been used historically to inform personnel decisions. The use of extraneous information, or information surrounding protected group status, is a major concern as AI is being increasingly used in hiring practices. As just one example, facial recognition can determine a candidate's sexual orientation with a great deal of accuracy (Rule et al., 2016). Organizations have the potential to collect a number of additional characteristics during the recruitment process such as age, body image, socioeconomic status, gender, health condition, race, and sexual orientation (van Esch et al., 2019). When used incorrectly or inappropriately, this information could then be used to catalogue job candidates further and to discriminate where possible (van Esch et al., 2019).

The potential for such a problem speaks directly to the role of I-O psychologists in maintaining fairness and legal defensibility. Similar to what was witnessed in the 1980s and 1990s surrounding biodata, I-O psychologists introduced biodata theories and methods to guide the development of biodata items to address potential problems (Hough & Paulin, 1994; Mael, 1991; Mumford & Owens, 1987; Mumford et al., 1990; Nickels, 1994; Schmitt et al., 1999; Schoenfeldt, 1999; Steinhaus & Waters, 1991). Even more recently, Speer et al. (2019) introduced a model that seeks to apply theory to empirically scoring biodata. As previously mentioned, development of AI-powered solutions for pre-employment testing is dominated by data scientists and computer programmers. However, as AI and machine learning approaches continue to gain traction in personnel decision making, I-O psychologists need to assume a seat at the table to

help ensure the appropriate use of data, legal defensibility, and optimal outcomes for organizations. While many of the problems arising from the use of AI are novel, a retrospective look at the problems associated with biodata can inform many of these issues. The proceeding sections delve more deeply into the use of AI, machine learning, and algorithmic scoring, with a focus on insights to guide today's personnel selection problems.

Additional criticisms of algorithmic assessments

Even though AI and machine learning algorithms may detect seemingly neutral criteria to select employees based on, it could lead to unintended bias, also referred to as classification bias. Classification bias can be defined as the use if "classification schemes have the effect of exacerbating inequality or disadvantage along line of race, sex, and other protected characteristics" (Kim, 2016, p. 911). For instance, an algorithm may be trained on data from high performers who are historically white. Based on this training dataset, the algorithm may determine that a certain criterion like address is highly correlated with job performance and recommend that candidates who live in a certain area should be selected because they are similar to current high performers. However, addresses are often conflated with socio-economic status and race. Thus, even though the algorithm wasn't intentionally identifying race as the basis for selection, an unintended consequence was to select a proxy for race as criterion.

These effects are often seen in the real-world, especially given the history of systematic oppression, segregation, and discrimination based on demographics that is a part of our history. In other words, systematic barriers and lack of access to opportunities for people belonging to certain groups is a key precedent to consider when thinking about bias in hiring today. As previously mentioned, even though overt discriminatory treatment has declined since the Civil Rights Act (1964), there is still potential for more subtle and systematic forms of discrimination to enter the system because of historic instances of bias. For instance, in 2015, a large technology company developed a machine learning algorithm to screen resumes in order to automate part of their talent acquisition process. However, in practice they found that instead of screening candidates in an unbiased way based on their skills or experience, the algorithm learned to rate women's resumes lower (identified when words such as "women's organization" were detected), presumably because the training set (in other words, the current demographic makeup of those who submit their resumes and get selected to work for the company) primarily consisted of men (Dastin, 2018).

While discrimination of this kind is possible when using other forms of assessment, the use of machine learning algorithmics that "learn" and update themselves can often exacerbate these issues and make them hard to identify. In

addition to the ethical implications of this, using such an assessment can also leave employers vulnerable to legal action under section 703(a)(2) of Title VII of the Civil Rights Act (1964) which states that "it shall be an unlawful employment practice for an employer to limit, segregate, or *classify* his employees or applicants for employment in any way which would deprive or tend to deprive any individual of employment opportunities or otherwise adversely affect his status as an employee, because of such individual's race, religion, sex, or national origin." This section was not written with machine learning algorithms and classification biases in mind, the language can enable employees or candidates to take legal action against algorithmic assessments that may propagate unintended biases.

Lastly, when using machine learning algorithms, there may be legal implications if the model continuously updates as and when new data becomes available. This is because with each model "revision", the selection criteria may change, making it inconsistent across candidates (Lundquist et al., 2019). Additionally, each modified model is considered a separate selection instrument and requires validity evidence and adverse impact analysis to ensure that the updated criterion is still job-related, valid, and fair. A related concern is also the accuracy and completeness of the data used to train the algorithm or evaluate candidates. For instance, when data from resumes or social media profiles are used along with the way candidates play a game, it can lead to varying amounts of missing data, and in turn inconsistency in the criteria on which candidates are selected. Since the consistency or reliability of a measure in assessing candidates is critical, this is an important concern to keep in mind when evaluating game-based assessment vendors.

Lessons from Biodata and Solutions for Algorithmic Assessments

Key lessons

One of the biggest lessons learned from the popularity of biodata assessments is the importance of using organizational psychology research to understand what can best predict performance on the job. Related to this notion is the idea that development of any assessment should be a cross-functional effort and it is critical to have I-O psychologists at the table to help make key decisions to ensure that there is intentional or unintentional discrimination baked into pre-employment assessments.

Role of the I-O psychologist

As noted, I-O psychologists were critical in advancing the extensive work done on biodata assessments and continue to produce guidelines to ensure that any criteria

used to make personnel decisions is driven by theory. Unfortunately, a vast majority work in the AI and machine learning space is predominantly conducted by engineers and data scientists. Although they bring an immense amount of technical expertise, the lack of I-O psychologists often leads to approaches that are atheoretical and can propagate bias.

A major tenet of the field of I-O psychology is to promote fairness, equity, and unbiased decision making in the world of work. For example, a core responsibility of I-O psychologists is creating assessments that maximize validity while minimizing personnel decisions that perpetuate bias or inequality. Adverse impact is a legal concept, concerned with ensuring the equality of outcomes resulting from the implementation decision rules that are applied to real-world scores (Arthur et al., 2013). The real-world scores in this case would be used to make personnel selection decisions. Not only are I-O psychologists for implementing the decision rules, but they are also responsible for designing assessments that maximize validity while minimizing adverse impact. Techniques for reducing adverse impact include post-test methods such as cut scores, banding, and empirically derived keys that ensure validity while minimizing adverse impact (Arthur et al., 2013).

Thus, we recommend that assessment development teams should be cross-functional and include I-O psychologists who have deep knowledge of what makes people successful on the job as well as data scientists and designers who have the technical skills to build the assessment technology. At every step of the hiring process, I-O psychologists can help ensure the appropriate use of data, legal defensibility, and optimal outcomes for organizations using these novel approaches.

Using theory to guide assessment development

As is evident from the discussion above, one of the biggest lessons learned from the practice of biodata-based assessments and one of the biggest critiques of AI and machine learning based assessments is the danger of using purely data-driven approaches to assessment development. Since data-driven approaches use pattern recognitions and correlations in the existing data to predict what behaviors are related to important outcomes like job performance, they can increase discrimination, especially if existing data does not reflect diverse populations. It is imperative to remember that correlation does not equal causation. In other words, just because something correlates with job performance does not mean it can actually lead to increased performance. Instead, psychological theories and research should be used to determine exactly what should be measured or used as criteria to evaluate candidates. This can be done through job analysis or competency

modeling which help identify knowledge, skills, abilities, and other characteristics that are related to performing a job well.

Custom versus neural networks

The use of theory driven methods for developing assessment does not mean that AI and machine learning cannot be leveraged to increase its predictive power. Instead, it is a question of using the right AI systems or algorithms so that existing bias does not get propagated through a new assessment. There are two key types of AI systems used to develop scoring algorithms for assessments: neural networks and custom systems. Neural network AI systems are systems that analyze large volumes of incoming data, "learn" patterns, and adapt their behavior accordingly. This can lead to a situation where the algorithm may select or reject a candidate but is unable to explain why. Moreover, if archival or existing employee data is used for training the algorithm, it can result in propagation of unintentional bias. Conversely, custom networks observe best practices from human raters who are taught how to rate specific competencies. In order to minimize bias, multiple raters can be used, and the system can be trained to "learn" from behaviors patterns that are common between the way in which individual human raters score candidates. While generating training data in this way takes longer, the use of custom systems not only minimizes the chance of bias in the algorithm but increases transparency since criteria on which selection or rejection decisions have been made are clear.

Conclusion

In summary, the inherent problems with AI-powered assessments and algorithmic scoring are akin to the historical problems with biodata. Similar to the way in which I-O psychologists help ensure the proper use of biodata, I-O psychologists can help ensure the appropriate use of theory, legal defensibility, and optimal outcomes for organizations using these novel assessment approaches. As a field, we must not forget the lessons learned from our past as we continue to integrate AI-powered assessments.

References

Anderson, N., & Shackleton, V. (1990). Decision making in the graduate selection interview: A field study. *Journal of Occupational Psychology*, *63*(1), 63–76.

Arthur, W., Doverspike, D., Barrett, G. V., & Miguel, R. (2013). Chasing the Title VII holy grail: The pitfalls of guaranteeing adverse impact elimination. *Journal of Business and Psychology*, *28*(4), 473-485.

Barrick, M. R., & Zimmerman, R. D. (2005). Reducing voluntary, avoidable turnover through selection. *Journal of Applied Psychology*, *90*(1), 159−166.

Berdahl, J., & Moore, C. (2006). Workplace Harassment: Double Jeopardy for Minority Women. *Journal of Applied Psychology*, *91*(2), 426-436.

Breaugh, J. A. (2009). The use of biodata for employee selection: Past research and future directions. *Human Resource Management Review*, *19*(3), 219–231.

Breaugh, J. A. (2014). Predicting voluntary turnover from job applicant biodata and other applicant information. *International Journal of Selection and Assessment*, *22*(3), 321–332.

Breaugh, J., Frye, K., Lee, D., Lammer, V., & Cox, J. (2014). The value of biodata for selecting employees: Comparable results for job incumbent and job applicant samples? *Journal of Organizational Psychology*, *14*, 40–51.

Bogen, M., Rieke, A. (2018, December). *Help Wanted: An Exploration of Hiring Algorithms, Equity, and Bias*. https://www.upturn.org/static/reports/2018/hiring-algorithms/files/Upturn%20--%20Help%20Wanted%20-%20An%20Exploration%20of%20Hiring%20Algorithms,%20Equity%20and%20Bias.pdf

Buel, W.D., Albright, L.E., & Glennon, J.R. (1966). A note on the generality and cross-validity of personal history for identifying creative research scientists. *Journal of Applied Psychology*, *50*(3), 217–219.

Cascio, W.F. (1975). Accuracy of verifiable biographical information blank responses. *Journal of Applied Psychology*, *60*(6), 767–769.

Childs, A., & Klimoski, R.J. (1986). Successfully predicting career success: An application of the biographical inventory. *Journal of Applied Psychology*, *71*(1), 3–8.

Civil Rights Act of 1964, Pub. L. No. 88-352, 78 Stat. 241 (1964). https://www.govinfo.gov/content/pkg/STATUTE-78/pdf/STATUTE-78-Pg241.pdf

Civil Rights Act of 1991, Pub. L. No. 102-66, 105 Stat. 1071 (1991).

Cucina, J. M., Caputo, P. M., Thibodeaux, H. F., & MacLane, C. N. (2012). Unlocking the key to biodata scoring: A comparison of empirical, rational,

and hybrid approaches at different sample sizes. *Personnel Psychology*, *65*(2), 385–428.

Cucina, J. M., Caputo, P. M., Thibodeaux, H. F., MacLane, C. N., & Bayless, J. M. (2013). Scoring biodata: Is it rational to be quasirational? *International Journal of Selection and Assessment*, *21*(2), 226–232.

Dastin (2018, October 18). *Amazon scraps secret AI recruiting tool that showed bias against women.* https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

Drakeley, R.J., Herriot, P., & Jones, A. (1988). Biographical data, training success, and turnover. *Journal of Occupational Psychology*, *61*(2), 145–152.

Equal Employment Opportunity Commission (EEOC), Civil Service Commission, Department of Labor, & Department of Justice. (1978). Uniform guidelines on employee selection procedures. *Federal Register*, *43*, 38290–39315

Freeberg, N.E. (1967). The biographical information blank as a predictor of student achievement: A review. *Psychological Reports*, *20*, 911–925.

Gessner, T. E., O'Connor, J. A., Clifton, T. C., Connelly, M. S., & Mumford, M. D. (1993). The development of moral beliefs: A retrospective study. *Current Psychology*, *12*, 236–254.

Goldsmith, D.B. (1922). The use of the personal history blank as a salesmanship test. J*ournal of Applied Psychology*, *6*, 149–155.

Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, *57*(3), 639–683.

Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: Issues, evidence and lessons learned. *International Journal of Selection and Assessment*, *9*(1-2), 152–194.

Hough, L., & Paullin, C. (1994). Construct-oriented scale construction. In G. A. Stokes, M. D. Mumford, &W. A. Owens (Eds.), *Biodata handbook* (pp. 109−145). Palo Alto, CA: Consulting Psychologists Press.

Hogan, J. B. (1994). Empirical keying of background data measures. In G. A. Stokes, M. D. Mumford, &W. A. Owens (Eds.), *Biodata handbook* (pp. 69−107). Palo Alto, CA: Consulting Psychologists Press.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*(1), 72–98.

Jones (2017, June 9). *Fortune 500: 7 in 10 Senior Executives are White Men.* https://fortune.com/2017/06/09/white-men-senior-executives-fortune-500-companies-diversity-data/

Karas, M., & West, J. (1999). Construct-oriented biodata development for selection to a differentiated performance domain. *International Journal of Selection and Assessment*, *7*(2), 86–96.

Kim, P. T. (2016). Data-driven discrimination at work. *William. & Mary L. Review*, *58*(3), 857–936.

Laurent, H. (1962). Early identification of management talent. *Management Record*, *24*, 33–38.

LinkedIn (2017). *LinkedIn Global Recruiting Trends.* https://www.slideshare.net/pedrooolito/linkedin-global-recruiting-trends-report-2017

Lundquist, K. K., Scott J. C., Nyugen, R., Locklear, T. S., Tomczak, K., Winter, J., & Foster, K. E. (2019). *Survival Guide to the Bright, Shiny Trends in Talent Acquisition.* https://aptmetrics.com/wp-content/uploads/2019/11/Survival-Guide-to-the-Bright-Shiny-Trends-in-Talent-Acquisition-2019-SIOP-LEC-1.pdf

Mael, F. A. (1991). A conceptual rationale for the domain and attributes of biodata items. *Personnel Psychology*, *44*(4), 763–792.

Mitchell, T. W., & Klimoski, R. J. (1982). Is it rational to be empirical? A test of methods for scoring biographical data. *Journal of Applied Psychology*, *67*, 411–418.

Moran, G., Walsh, J.A., Clement, W., & Bumbeck, J. (1968). Personal data as a predictor of small finance company credit risk. *Personnel Psychology*, *21*, 245–253.

Mumford, M.D., & Owens, W.A. (1982). life history and vocational interests. *Journal of Vocational Behavior*, *21*(3), 330–348.

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

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*(2), 361–398.

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

Nickels, B. J. (1994). The nature of biodata. In G. A. Stokes, M. D. Mumford, & W. A. Owens (Eds.), *Biodata handbook* (pp. 1−16). Palo Alto, CA: Consulting Psychologists Press.

Oswald, F. L., Schmitt, N., Kim, B. H., Ramsay, L. J., & Gillespie, M. A. (2004). Developing a biodata measure and situational judgment inventory as predictors of college student performance. *Journal of Applied Psychology*, *89*(2), 187–207.

Payscale (2021). The State of the Gender Pay Gap in 2021. Racial and Gender Pay Gap Statistics for 2021 (payscale.com)

Pulakos, E. D., & Schmitt, N. (1996). An evaluation of two strategies for reducing adverse impact and their effects on criterion-related validity. *Human Performance*, *9*(3), 241–258.

Reilly, R. R., & Chao, G. T. (1982). Validity and Fairness of Some Alternative Employee Selection Procedures. *Personnel Psychology*, *35*(1), 1–62.

Reiter-Palmon, R., & Connelly, M. S. (2000). Item selection counts: A comparison of empirical key and rational scale validities in theory-based and non-theory-based item pools. *Journal of Applied Psychology*, *85*(1), 143–151.

Robertson, I. T., & Makin, P. J. (1986). Management selection in Britain: A survey and critique. *Journal of Occupational Psychology*, *59*(1), 45–57.

Rosenbaum, R. W. (1976). Predictability of employee theft using weighted application blanks. *Journal of Applied Psychology*, *61*(1), 94–98.

Rothstein, H. R., Schmidt, F. L., Erwin, F. W., Owens, W. A., & Sparks, C. P. (1990). Biographical data in employment selection: Can validities be made generalizable? *Journal of Applied Psychology*, *75*(2), 175–184.

Rule, N. O., Bjornsdottir, R. T., Tskhay, K. O., & Ambady, N. (2016). Subtle perceptions of male sexual orientation influence occupational opportunities. *Journal of Applied Psychology*, *101*(12), 1687–1704.

Smither, J. W., Reilly, R. R., Millsap, R. E., Pearlman, K., & Stoffey, R. W. (1993). Applicant reactions to selection procedures. *Personnel Psychology*, *46*(1), 49–76.

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*(2), 262–274.

Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. (1984). Meta-analyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology*, *37*(3), 407–422.

Schmitt, N., Jennings, D., & Toney, R. (1999). Can we develop measures of hypothetical constructs? *Human Resource Management Review*, *9*, 169−183.

Schoenfeldt, L. F. (1999). From dust bowl empiricism to rational constructs in biographical data. *Human Resource Management Review*, *9*, 147−167.

Sharf, J. C. (1994). The impact of legal and equal employment opportunity issues on personal history inquiries. In G. A. Stokes, M. D. Mumford, &W. A.

Owens (Eds.), *Biodata handbook* (pp. 351−390). Palo Alto, CA: Consulting Psychologists Press.

Siegel, L.A. (1956) Biographical inventory for students: Validation of the instrument. *Journal of Applied Psychology*, *40*, 122–126.

Steinhaus, S. D., & Waters, B. K. (1991). Biodata and the application of a psychometric perspective. *Military Psychology*, *3*, 1−23.

Stokes, G. S. (1999). Introduction to special issue: The next one hundred years of biodata [Editorial]. *Human Resource Management Review*, *9*(2), 111–116.

Stricker, L. J., & Rock, D. A. (1998). Assessing leadership potential with a biographical measure of personality traits. *International Journal of Selection and Assessment*, *6*(3), 164–184.

Society for Human Resource Management. (2016, November). *2016 Human Capital Benchmarking Report.* https://www.shrm.org/hr-today/trends-and-forecasting/research-and-surveys/Documents/2016-Human-CapitalReport.pdf

van Esch, P., Black, J. S., & Ferolie, J. (2019). Marketing AI recruitment: the next phase in job application and selection. *Computers in Human Behavior*, *90*, 215–222.

van Rijn, P. (1992) Biodata: Potentials and challenges in public sector employee selection. *Personnel Assessment Monographs of International Personnel Management Assessment Council*, *2*(4), Alexandria, VA: Assessment Council of the International Personnel Management Association.

Walther, R. H. (1961). Self-description as a predictor of success or failure in foreign service clerical jobs. *Journal of Applied Psychology*, *45*(1), 16–21.