HIRING

Big data solutions to candidate screening and recruitment are attractive to employers because they are cheap, fast, easy to administer, and capable of narrowing a pool of thousands of candidates within seconds. But are the new approaches legally defensible? In this Bloomberg Law Insights article, authors Esther Lander, Ashley Keapproth, and David Jones examine the legal principles involved with employee selection screening.

Tweet This: Legal Considerations for Big Data Hiring

ESTHER LANDER, ASHLEY KEAPPROTH AND DAVID JONES

Introduction

Technology is not new to the world of hiring. In the 1990s, for example, technology-based solutions could read resumes and determine if they matched an employer's job description. Today, as individuals are continuously creating data about themselves on social media, new analytic tools are being developed that scour the Internet, examine an applicant's online footprint, and identify those with potential. Some vendors promise to identify a company's best applicants using data algorithms that screen applicants' social media

Esther G. Lander and Ashley J. Keapproth are partner and counsel, respectively, in the labor and employment practice at Akin Gump Strauss Hauer & Feld. Ms. Lander is a former principal deputy chief of the Employment Litigation Section within the Civil Rights Division of the Justice Department.

David Jones Ph.D. founded HRStrategies, a candidate assessment firm acquired by Aon-Hewitt. He is currently president of Growth Ventures Inc., a technology-based candidate assessment company with global clients, and a frequent expert witness.

profiles, online footprints, or resumes for words that they claim predict successful on-the-job performance.

As the world of technology grows, employers are seeking higher-volume, faster-paced, and lower-cost approaches to candidate screening and assessment. Big data solutions are attractive to employers because they are cheap, fast, easy to administer, and capable of narrowing a pool of thousands of candidates within seconds. Most importantly, the vendors selling these solutions claim technology is better at making good selection decisions than people, therefore giving companies that use them a competitive edge.

But are the new approaches legally defensible? Before implementing a big data hiring solution, employers must insist that providers of these analytical solutions demonstrate *not only* improved hiring results, *but also* that (1) minorities or women are not adversely impacted by the screening method, or (2) that the employer can meet its burden under Title VII of proving that the screening procedure is job related and consistent with business necessity. To date, little has unfolded in this area.

Basic Legal Principles of Employee Selection Procedures

Under Title VII, a selection device, such as a test, interview, background check, or other tool used to screen applicants, must be job related and consistent with business necessity if it has a disparate impact on members of a protected group. A disparate impact occurs when a selection device disproportionately excludes minority groups or women from advancing to the next

stage of the selection process. Once disparate impact is established, the employer must demonstrate that the selection procedure is job related and consistent with business necessity.

Job relatedness typically must be established through a validation study. The EEOC's Uniform Guidelines on Employee Selection Procedures recognize three forms of validation: content, criterion-related, and construct. Content validity demonstrates that the content of the selection procedure represents the content of the job, such as a pilot simulator for pilots or a typing test for administrative positions. Construct validity demonstrates that a selection procedure measures a construct or underlying human trait (such as conscientiousness or adaptability), and that the trait is important to successful job performance. Construct validity is seldom used because obtaining empirical support requires a series of arduous and expensive research studies.

Criterion-related validity asks an empirical question: Is performance on the selection procedure predictive, or significantly correlated with, subsequent performance on the job? When validating a selection procedure using criterion-related validity, there must be a demonstrated relationship statistically between scores on the selection procedure and job performance or other important job behaviors.

Assuming an employer can meet its burden of proving job relatedness, employers are still obligated to explore equally valid but less discriminatory alternatives if disparate impact is found. For example, employers should ask whether a different way of scoring, combining, or standards-setting among components of a selection process would still identify qualified candidates, but reduce any adverse impact that results under different ways of using the tools.

Considering the Adverse Impact of Data Analytics

In the absence of a study that demonstrates otherwise, employers should assume that protected groups will be adversely impacted by using quantitative-driven screening algorithms. This is particularly true where the vendor claims that its algorithms are screening for things like educational attainment and/or past work experience, which are commonly known to negatively impact racial and national origin groups. Similarly, if job tenure is within the algorithm, women who leave the labor force to have children, or persons with disabilities who have had periods of medical absence from work, could be adversely impacted by the algorithm.

Another very important consideration for adverse impact is the applicant pool size. As a matter of statistics, the larger the applicant pool, the more likely adverse impact will be identified. Any data solution that considers information about candidates to determine who should advance in the hiring process will likely convert all such individuals into "applicants" for purposes of adverse impact analysis. For instance, if an algorithm is used to filter *all* applicants, a company's pool of applicants for purposes of assessing adverse impact will arguably include everyone subject to the algorithm, not just those applications or resumes that were reviewed or considered by a person. This remains true regardless of whether the employer is using a data tool to screen the resumes themselves for particular words, or using a

data tool to scour the internet for information about all if its applicants to narrow the pool.

Therefore, an employer must consider whether to use data screening tools on the entire applicant population or on a smaller randomly selected subset of applicants, taking into account the size of the applicant pool, the strength of validation evidence, and any adverse impact it is likely to produce. Bottom line, using screening algorithms grows the size of applicant groups, which, in turn, grows the likelihood of the employer experiencing evidence of adverse impact, evidence that then needs to be defended.

Considering the Validity of Data Analytics

Any data-analytic screening process that results in adverse impact must be properly validated. Since data-analytic screening tools do not simulate the job or test for human traits, a criterion-related validity study should be conducted to document a statistical relationship between the screening algorithm and successful job performance. Most criterion-related validation studies rely on a sample of incumbent employees who are given the screening procedure and whose job performance is rated. Statistical analyses are then performed to see whether success on the screening procedure predicts success on the job. While many vendors claim to have conducted this type of study, employers should ensure that the tools, data, analytics, and methods for summarizing the study undertaken can satisfy the rigorous process under the EEOC's Uniform Guidelines, including the following.

First, has there been a job analysis? Most criterion-related validity studies begin with a job analysis to determine what performance-based criteria should be measured to see if they correlate with success on the screening procedure. This requires some review of job requirements to determine the work behaviors that are relevant and important to the job in question.

Second, once criteria for documenting successful job performance have been identified, employers must then collect supervisor ratings or other metrics, such as sales levels, production rates, error frequency, customer survey results, etc., for those individuals in the study to see if their job performance scores correlate with how well they do on the data analytic screening device. This requires careful calibration and training of performance raters to ensure the consistency and accuracy of the data being collected, as well as the inclusion of any job-related data that may correlate with the screening procedure scores, such as turnover, absenteeism, or disciplinary history.

Third, is the screening device itself competently designed? Designing a competent selection procedure requires a background, qualifications, and expertise in testing validation, typically with a degree in Industrial-Organizational Psychology. Many of the vendors selling "big data" hiring solutions, however, do not have this background. Their backgrounds are only in data analytics. As a result, tools are being marketed that analyze applicants' online footprints and rate positively or negatively words or phrases taken from the online footprint of high performing incumbents. In this way, the vendors claim successful job performance has been correlated with screening words or phrases. Although this "big data" approach for developing an algorithm based on high performing incumbents *sounds like* criterion-

related validity, in reality the words and phrases selected may not be job related at all. For example, perhaps the high performing incumbents happen to regularly follow or discuss sports or sports teams online. As a result, an algorithmic hiring solution based on these incumbents could screen for applicants who follow the same sports teams as the incumbents. This could result in an adverse impact against women or minority groups without any job-related justification. Thus, it is critical for employers to fully understand how any data-analytic tool works and what job-related qualifications, terms, or phrases the tool may home in on when making selections.

Fourth, is the procedure fair? Any properly conducted validation study must consider ''test fairness.'' Among other things, this means that the sample of individuals chosen for the study must represent a large and diverse sample of current employees to be able to test for bias in how different groups are being rated by supervisors, or in how they perform on the selection procedure. Thus, the vendor should address how data will be collected on incumbent employees, whether the collection will be representative and diverse, and how, if at all, it will test for bias in the rating process, and adverse impact in how the sample of incumbents performs on the selection procedure.

Finally, are the ratings, cutoff scores, or rankings validated? Tools may rank applicants or set a passing cutoff score. Either method must be validated if it results in adverse impact. For example, if the data-analytic algorithm ''scores'' applicants against a threshold passing score, that passing score must be validated. If the algorithm ''ranks'' candidates in a particular order for selections, the ranking practice must be validated. Validating passing scores and ranking approaches can be done by using the performance data from the incumbent samples and showing statistically that score differences predict meaningful performance differences.

Considering Alternatives

When a selection procedure results in adverse impact on a protected group, Title VII requires employers to consider equally valid alternatives that reduce adverse impact. For this reason, criterion-related validity studies typically look at the adverse impact created by the selection procedure and consider modified uses of that procedure to see whether the adverse impact can be reduced without compromising validity. A data-analytic algorithm could be modified, for example, by tinkering with the words and word choices that are driving adverse impact to see if the correlations with performance remain strong when those words are removed.

Additionally, two alternatives that typically have little adverse impact and correlate well with certain entry-level performance are personality inventories and biographical surveys. And, one alternative is always the employer's existing process for screening applicants. In litigation, all of these alternatives would need to be considered. Again, these steps are unaddressed in the current data-analytic solutions being advertised.

Conclusion

The unfolding data-analytic algorithm products on the market have great potential for quickly analyzing hundreds or thousands of applicants, eliminating significant manual labor, and offering an advantage over competitors in identifying and hiring the best employees. However, as described above, responsible employers should consider whether the procedure they use is legally defensible before implementation. This can only be done by launching a sound, professionally developed validation study, which many vendors on the market have not done. Also important can be setting up data tracking and new hire monitoring to examine, analyze, and determine whether what has been expected regarding use of the online data-analytic process actually occurs. For example, if the process has been validated to project new hire sales increases, production improvements, and lower turnover, has it? In effect, the employer's Six Sigma approach to evaluating the new selection procedure's results can offer the most valuable defensibility needed.

In conclusion, employers who proceed with purchasing big data hiring solutions should seriously consider retaining an experienced professional who specializes in employment selection procedures to review the validation work *before* the product is used on real applicants. Absent such an assessment, any potential benefits of these screening processes are greatly outweighed by the legal risk of a successful challenge under Title VII.