Chair Yang and Commissioners Barker, Feldblum, Lipnic and Burrows: Good afternoon. Thank you for the opportunity to share some of my thoughts with you about the challenges and opportunities for employers as they begin to use big data to recruit and select candidates for employment. Specifically, I will address the validation issues associated with the use of big data.

The framework for my remarks is that of an Industrial Organizational psychologist - one who studies the relationship between people and jobs - ensuring that the way we make predictions about who is best suited for a job accurately and fairly assesses the skills, abilities and other characteristics needed for success in the job.

**Uses of Big Data in the Employment Setting**

Big data, predictive analytics or talent analytics - all terms used to describe the harvesting of a wide range of empirical data for HR decision making - is the inevitable future of HR. It presents a future that is both promising and scary.

The wide variety and accessibility of data today has introduced the opportunity to develop predictive analytics for HR decision-making in much the same way that retailers are able to target ads to you online based on your browsing history. These algorithms may rely on such diverse data as:

- social media habits;
- twitter feeds;
- specific combinations of words on resumes;
- personality test results;
- facial recognition software, and
- individual performance ratings.

Algorithms may also incorporate such internal company information as the frequency and location of team meetings, the recipients and content of employee emails, and records of employee participation in wellness programs.

By combining and analyzing trends in these data, companies can develop predictive algorithms to assist in a wide range of HR decision processes, e.g., to identify possible recruits, evaluate talent, reduce turnover or build leadership pipelines. Such methods have also been used to identify the drivers of specific behaviors, such as diversity, innovation, and collaboration. The information thus gained forms the basis for implementing targeted programs to support, increase and leverage the desired behaviors (Bersin, 2016).

**The Nature of the Algorithms**

Where predictive analytics are used to recruit or select candidates for jobs, the algorithms are typically constructed by compiling a wide range of data and mining that data to maximize the ability to predict membership in some criterion group, such as high-performing current employees. Through successive approximations (i.e., machine learning), data patterns from diverse and seemingly unrelated variables (such as social media behavior, GPA and personality test results) are
analyzed, combined and weighted to best predict the characteristics of the high-performing group. These weighted characteristics of the high-performing group may then be used to score future job candidates as to their likelihood of success on the job.

Hundreds or even thousands of data points may be combined in the algorithm, but for any individual candidate many of the data points may not be populated. For example, in one algorithm over 100,000 individual data points were potentially scoreable. Due to missing data and the nature of the characteristics that were scored, the average candidate was scored on results for only 500 data points, but those 500 data points were unlikely to be the same from candidate to candidate. Hence, the specific data on which two candidates would be compared very likely would not be the same. Moreover, algorithms used for recruiting often include data obtained by searching publicly available databases where the accuracy or completeness of the data may be questionable, leading to more missing and incorrect data in the selection process.

Consider also the situation where the algorithm only draws information from resumes provided by job applicants. The algorithm may give credit (positive or negative) for the inclusion of certain specific words in the resume (e.g., "technical" or "football") or the absolute number of words in the resume. Two equally qualified candidates might describe their experience and backgrounds using different words or different styles of expression and be scored quite differently by the algorithm.

The data-driven weights may also sometimes be counterintuitive or non-linear. For example, GPAs of 2.0 to 2.9 may be considered more valuable in the predictive algorithm than GPAs of 3.0 to 3.5. In these situations, further research is necessary to properly understand and interpret the meaning of these results for prediction.

Finally, algorithms may be trained to predict outcomes which are themselves the result of previous discrimination. The high-performing group may be non-diverse and hence the characteristics of that group may more reflect their demographics than the skills or abilities needed to perform the job. The algorithm is matching people characteristics, rather than job requirements.

While these methods present the possibility of identifying previously unrecognized talent and reducing some aspects of subjective decision making, their use can be compromised by errors inherent in the data and the decisions that underlie the design of the algorithms themselves. Moreover, despite the ease with which these data are collected, there is surprisingly little real validation evidence being collected to substantiate the job relatedness of the algorithms used.

Validation Challenges

Underlying any process used for employment decision-making is the assumption that it is measuring the characteristics necessary for successful job performance. Validation is the process of examining whether a particular device used to make a selection decision is job related. This generally requires a showing that the selection procedure measures important aspects of the job and is predictive of future job success. According to the Uniform Guidelines, selection procedures must be validated if they have adverse impact when used as a basis for an employment decision.

By its nature, the process of selecting and weighting variables in an algorithm to best predict and correlate with successful job performance appears to be evidence of criterion-related validity. However, a conclusion that this is sufficient evidence of validity is misplaced. In an article entitled "Data-Driven Discrimination at Work", Pauline Kim points out:

> In the case of workforce analytics, the data algorithm by definition relies on variables that are correlated in some sense with the job. So 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. 7)

The process of developing an algorithm is inductive and atheoretical, often a "black box". It is driven by the data, rather than by any understanding of the causal relationship between the variables and the behavior it seeks to predict. The atheoretical basis of many predictive algorithms is inconsistent with professional standards for validation which specify the "variables chosen as predictors should have an empirical, logical, or theoretical foundation. The rationale for a choice of predictor(s) should be specified" (SIOP Principles, p. 17).

As a result of its atheoretical underpinnings, such analysis may produce unstable models which do not predict equally well for different populations (e.g., applicants vs. incumbents, more diverse applicant pools, or applicants with different levels of experience). Moreover, the ease with which statistically significant correlations with job performance can be achieved in large data sets challenges the interpretation and stability of such evidence. For this reason, it is essential to monitor the continued effectiveness of the algorithm and to cross-validate it on new data to ensure it is predicting accurately.

Multiple strands of validity evidence may be required to support the inference that the algorithm is valid. According to the Principles for the Validation and Use of Selection Procedures published by the Society for Industrial and Organizational Psychology (SIOP, 2003, p. 8), "Where the proposed uses [of a selection procedure] rely on complex, novel, or unique conclusions, multiple lines of converging evidence may be important". It may be necessary to do more than show a statistically significant correlation; it may be necessary to understand why that correlation exists and what it tells us about a person's ability to perform the job. It is
the accumulation of evidence about the relationship between performance on the selection procedure and job performance which strengthens the researcher's ability to assert an unambiguous conclusion of validity.

Where the algorithm is not fixed, but rather continues to iterate with additional data, validity must be established for each iteration or successive approximation that is used for decision-making. This presents a situation where the "test" is not only variable from person to person, but ever-changing.

Finally, for a selection procedure to be valid or job-related, it must first be reliable or relatively free from measurement error. Measurement error is produced by factors that influence candidates' scores or evaluations, but are unrelated to the attributes being measured by the selection procedure. To ensure reliability, there must be standardization in the execution of the selection procedure (including, for example, what specific criteria are used and how those criteria are scored) to minimize the extent to which measurement error impacts the results. Without reliability, interpretations about job performance are undermined since the selection procedure's results were contaminated by factors unrelated to the candidates' skills and abilities. Hence, a selection procedure must be BOTH reliable and valid to be job-related.

To the extent that a particular algorithm suffers from incomplete, inaccurate or missing data (as discussed above), the reliability of the instrument is subject to challenge. Without reliable data, the validity of a selection procedure cannot be supported.

Best Practices in the Use of Big Data

Employers may take a number of steps to implement big data analytics in ways that meet best practices. Five ways to implement effective predictive analytics solutions include:

1. Validate the predictive models' accuracy over time and with different employee segments;
2. Conduct a job analysis to ensure the algorithm is measuring the knowledge, skills and abilities related to the job performance, rather than reflecting the demographic characteristics of current employees;
3. Examine the representativeness of the populations included and the accuracy and fairness of the data inputs on which the algorithm is based to ensure that all relevant data are both correct and inclusive;
4. Train and support managers in how to interpret and use these solutions to make decisions, and
5. Inform candidates about the use of information to avoid privacy concerns.

The skills needed for success in the jobs of the 21st century will certainly provide a different and, in some cases, a more level playing field for ethnic and racial minorities competing for jobs. They will also require those designing selection processes to look more broadly at the tools available to measure these skills fairly and accurately.

Thank you.

References


