Journal of Applied Psychology

Using Machine Learning to Translate Applicant Work History Into Predictors of Performance and Turnover

Sima Sajjadiani, Aaron J. Sojourner, John D. Kammeyer-Mueller, and Elton Mykerezi

Online First Publication, March 25, 2019. http://dx.doi.org/10.1037/apl0000405

American Psychological Association

Journal of Applied Psychology

0021-9010/19/$12.00

http://dx.doi.org/10.1037/apl0000405

Using Machine Learning to Translate Applicant Work History Into Predictors of Performance and Turnover

Sima Sajjadiani
University of British Columbia

Aaron J. Sojourner and John D. Kammeyer-Mueller
University of Minnesota

Elton Mykerezi
University of Minnesota

Work history information reflected in resumes and job application forms is commonly used to screen job applicants; however, there is little consensus as to how to systematically translate information about one's work-related past into predictors of future work outcomes. In this article, we apply machine learning techniques to job application form data (including previous job descriptions and stated reasons for changing jobs) to develop interpretable measures of work experience relevance, tenure history, and history of involuntary turnover, history of avoiding bad jobs, and history of approaching better jobs. We empirically examine our model on a longitudinal sample of 16,071 applicants for public school teaching positions, and predict subsequent work outcomes including student evaluations, expert observations of performance, value-added to student test scores, voluntary turnover, and involuntary turnover. We found that work experience relevance and a history of approaching better jobs were linked to positive work outcomes, whereas a history of avoiding bad jobs was associated with negative outcomes. We also quantify the extent to which our model can improve the quality of selection process above the conventional methods of assessing work history, while lowering the risk of adverse impact.

Keywords: selection, occupational analysis, data mining

Researchers and practitioners increasingly use massive databases of job applications produced by electronic application systems. These systems present challenges, as organizations need to contend with many applicants in a systematic and efficient manner (Flandez, 2009; Grensing-Pophal, 2017). Both human resource (HR) departments and consulting firms are evaluated based on time-to-hire and volume of qualified candidates (Gale, 2017). To keep the best applicants, organizations need to respond rapidly to individuals who may be sending out dozens of online applications (Ryan, Sacco, McFarland, & Kriska, 2000). Time pressures, the large volume of applications, the complexity of the decision task, and recruiters' biases and heuristics increase the chance of overlooking or misinterpreting candidate qualifications (e.g., Converse, Oswald, Gillespie, Field, & Bizot, 2004; Tsai, Huang, Wu, & Lo, 2010).

While standardized tests and inventories speed the acquisition of data, they overlook individualized applicant information. Facets of work history, including relevant work experience, tenure in previous jobs, and reasons for leaving previous jobs, are empirically and conceptually distinct from either cognitive ability or personality, and so can add significant predictive power in a selection battery (Ryan & Ployhart, 2014). Although job-relevant experience predicts performance, it is difficult to track job-relevance systematically across applicants' idiosyncratic work histories (Tesluk & Jacobs, 1998). For example, recruiters struggle to quantify the difference between an individual with 5 years of experience in childcare relative to someone with 3 years of experience in corporate training, or between a person who quit a previous job because of insufficient administrative support relative to an intrinsic desire to share knowledge. Lacking a system for organizing job history information, many organizations evaluate qualifications using idiosyncratic and cumbersome processes (Brown & Campion, 1994).

Sima Sajjadiani, Organizational Behaviour and Human Resources Division, Sauder School of Business, University of British Columbia; Aaron J. Sojourner and John D. Kammeyer-Mueller, Work and Organizations Department, Carlson School of Management, University of Minnesota; Elton Mykerezi, Department of Applied Economics, University of Minnesota.

This project was supported in part by funding from a University of Minnesota Carlson School Dean's Small Grant and IES grant R305H160036. Earlier versions of this article were presented at the 2017 meeting of the Society for Industrial and Organizational Psychology, the 2017 meeting of the Association for Public Policy Analysis and Management, the 2019 meeting of the Allied Social Science Association, and the Northwestern University Institute for Policy Research (January 2019). Thanks to Arindam Banerjee, Alan Benson, Dan Goldhaber, Nathan Kuncel, Paul Sackett, and Connie Wanberg for helpful advice along the way. Thanks to district staff including Chris Moore, Maggie Sullivan, Scott Weber, and Daniel Glass and the many teachers whose work made this study possible.

Correspondence concerning this article should be addressed to Sima Sajjadiani, Organizational Behaviour and Human Resources Division, Sauder School of Business, University of British Columbia, 663-2053 Main Mall, Vancouver, BC V6T 1Z2, Canada. E-mail: sima.sajjadiani@sauder.ubc.ca

To circumvent these problems, more organizations use large-scale data analytic techniques to comb through open-ended text fields in electronic applications. Most HR professionals are familiar with automated keyword searches of applications, a method that far predates the use of electronic systems (e.g., Peres & Garcia, 1962). The development of these lists, however, is seldom linked to a conceptual or theoretical understanding of qualifications. In the absence of this knowledge, the cognitive and information limitations of decision makers are built into the system in the stage of building keyword lists (Bao & Datta, 2014). Keywords are also often applied in a rudimentary scorekeeping method, with each word that matches the keyword list receiving equal weight independent of the context in which the word is used.

Recent developments in machine learning provide opportunities to summarize work history as rapidly as keyword methods, but in a more rigorous and comprehensive manner that can be linked to other research and practice. Broadly defined, machine learning consists of prediction algorithms, including text mining techniques that classify chunks of text into categories or order them based on a criterion (Mohri, Rostamizadeh, & Talwalkar, 2012). Unlike keyword searches focused on individual words, machine learning techniques find terms that co-occur, enabling better incorporation of context. These techniques can also calculate the importance of words, and algorithmically calculate the probabilities that a statement belongs across multiple categories. These methods are increasingly used measure individual differences and work motivation (de Montjoye, Quoidbach, Robic, & Pentland, 2013; Doyle, Goldberg, Srivastava, & Frank, 2017). Despite calls to apply these methodological developments in employee selection (Campion, Campion, Campion, & Reider, 2016; Chamorro-Premuzic, Winsborough, Sherman, & Hogan, 2016), to the best of our knowledge, machine learning and text-mining have not been systematically applied to translate information from standard application forms into predictors of subsequent work outcomes.

In an attempt to find low-cost and systematically assessed predictors of performance and turnover from applications, we draw on existing theories and use machine learning techniques to develop useful and novel measures of work history. In purely empirical applications of machine learning, researchers rely on hundreds of variables, train different algorithms using a training sample of the data and evaluate the performance of these algorithms on a test sample. Then, they regularize these algorithms and evaluate how the algorithm can improve employee selection in the specific organization (e.g., Aiolli, de Filippo, & Sperduti, 2009; Chalfin et al., 2016). They do not examine the importance of particular predictors, nor attempt to evaluate mechanisms driving the prediction. Further, lack of interpretability can make these selection models vulnerable to challenge from stakeholders within the organization and from legal cases alleging that the method is disconnected from job requirements (Klehe, 2004). Our approach contrasts with this atheoretic, "black box" prediction approach that may not generalize to other contexts, sheds little light on the underlying mechanisms, and is difficult to explain to decision makers, the public, courts, or other stakeholders.

We focus on three aspects of work history from application forms: (a) *work experience relevance* incorporating correspondence of knowledge, skills, abilities, and other attributes (KSAOs) from previous job titles and job descriptions with the current job; (b) *tenure history,* incorporating length of tenure in previous jobs; and (c) *attributions for previous turnover,* including a history of involuntary turnover, leaving to avoid bad jobs, and leaving to approach better jobs. This mix of predictors represents components of skill development and job requirements embodied in the U.S. Department of Labor's Occupational Information Network (O*NET; Peterson, Mumford, Borman, Jeanneret, & Fleishman, 1999), patterns of behavior and attitudes, and general motivation for work based on approach-avoidance models (e.g., Elliot & Thrash, 2002; Elliot & Thrash, 2010; Maner & Gerend, 2007). We use these prehire measures to predict performance across multiple domains, and both voluntary and involuntary turnover hazards (i.e., duration of employment until turnover occurs).

Besides these innovations on the predictor side, we are able to evaluate the proposed selection system using a broad set of outcome variables, and contrast this idealized system relative to the existing selection system. Barrick and Zimmerman (2009) opened out the criterion space to include both performance and turnover, and concluded that it is more cost-effective for organizations to assess candidates using constructs that predict both performance and turnover. Our data allow us to accomplish this by incorporating multiple perspectives on performance, including (a) client evaluation, (b) expert observations of performance, (c) more-objective measures based on standardized test results, and (d) voluntary and involuntary turnover hazards.

Linking Work History to Performance and Turnover

Most job applications request information related to work experience and history of job changes. In addition to these factual pieces of information, forms ask applicants about the content of previous job responsibilities and reasons for leaving previous jobs. Below we describe how we use these clues in work history to assess how well-acquainted applicants are with job-related KSAOs, their motivational tendencies linked to turnover, and the duration of previous employment.

Relevant Experience

Work experience is conceptualized in terms of whether the applicant has encountered work situations relevant to the requirements of the job for which s/he applies. Ployhart (2012, p. 24) proposed that "work experience is a broad, multidimensional construct that often serves as a proxy for knowledge." Quiñones, Ford, and Teachout (1995) and Tesluk and Jacobs (1998) emphasized the importance of the qualitative aspects of work experience, including the type of tasks performed which can be translated into work-related knowledge and skills.

The key factor we assess is work experience relevance, defined consistent with prior work (e.g., Dokko, Wilk, & Rothbard, 2009) as the correspondence between the KSAO requirements of applicants' previous jobs and the job that the applicant seeks. KSAO-based matching of experience is a better predictor of performance than using titles and employment durations in previous jobs (Quiñones et al., 1995). The training and development literature (Blume, Ford, Baldwin, & Huang, 2010; Saks & Belcourt, 2006) similarly argues that repeatedly doing tasks contextually similar to those required for the focal job develops competency. Relevant job experiences are considered socially acceptable hiring criteria by job seekers, organizations, and legal systems as they are factual and job related (Hausknecht, Day, & Thomas, 2004; Klehe, 2004). Relevant work experience signals applicants' fit with the focal job. Applicants possessing relevant experience can make more informed decisions relative to those who have not had such direct interaction with core job tasks (Jovanovic, 1984). Individuals also tend to gravitate toward jobs that match their KSAOs (Converse et al., 2004; Wilk, Desmarais, & Sackett, 1995). In sum, work experience relevance, weighted by recency and tenure in each previous job, provides valuable information about applicants' level of knowledge, skills, abilities, interests, and values. Despite this, there are still few efforts to build a truly systematic scoring method for evaluating work experience relevance. Large databases of job titles and relevant tasks have been used in synthetic validation to assess entire selection measures or systems (Johnson, Steel, Scherbaum, Hoffman, Jeanneret, & Foster, 2010; Steel & Kammeyer-Mueller, 2009), but such tools have not been used to predict individual applicant performance.

We propose that machine learning can be used to match occupationally relevant KSAOs to job requirements. We operationalize work experience relevance by measuring the similarity between the KSAOs required for applicants' previous jobs and the required KSAOs for the job for which the applicant has applied. Starting with the previous jobs' self-reported titles and job descriptions provided by the applicants in their application form, we categorize each previous job into the standard O*NET occupations. As a machine learning technique, words from self-described job titles and job descriptions are matched with best fitting O*Net job titles probabilistically. These probabilities form estimates of the level of different work characteristics the individual has encountered in previous jobs. Profile analysis techniques measure the similarity between each applicant's past profile and the profile of the focal job. We then weight the KSAO match based on the tenure in each previous job and the recency of each job to give a better sense of the quantity of experience (e.g., Quiñones et al., 1995).

> *Hypothesis 1:* Work experience relevance, assessed through machine learning, is (a) positively related to performance and (b) negatively related to turnover hazard.

Tenure History

There is a consensus among organizational psychology researchers and practitioners that past behavior is the best predictor of the future behavior (e.g., Barrick & Zimmerman, 2005; Owens & Schoenfeldt, 1979; Wernimont & Campbell, 1968). Job applications provide information regarding behavioral tendencies based on the applicant's average length of time spent in previous jobs, which we term "tenure history." A person with a questionable tenure history has a record of changing jobs after a relatively short period of time, whereas a more reliable tenure history is indicated by many spells of long tenure. The existence of different typical levels of tenure history across jobs has been noted in several theoretical and empirical works (e.g., Judge & Watanabe, 1995; Maertz & Campion, 2004). As noted by Barrick and Zimmerman (2005) "while most turnover models view intent to quit as an immediate precursor to actual turnover, some individuals may be predisposed to quit even before starting the job" (p. 164). Peripatetic tenure history could also signal other problems, such as poor skills or low motivation. Short tenure in previous jobs may reflect a generally poor work ethic, correlated with consistently lower levels of organizational commitment and a higher likelihood of turnover (Mathieu & Zajac, 1990). Also, job applicants with poor levels of skills or motivation are expected to have lower average tenure in their previous jobs as they either involuntarily or voluntarily leave the job as they lack dispositional conscientiousness for their work (Barrick & Zimmerman, 2009; Griffeth, Hom, & Gaertner, 2000).

> *Hypothesis 2:* Tenure in previous positions is positively related to (a) performance, and negatively related to (b) voluntary and (c) involuntary turnover hazard.

Attributions for Previous Turnover

Machine learning is well-suited to examining open-ended text regarding one's attributions for leaving past jobs. We start from the premise that turnover attributions extracted from job applications are valid signals of traits and dispositions toward work. Individuals vary greatly in the attributions they make for prior turnover (Lee, Mitchell, Holtom, McDaneil, & Hill, 1999). Attributions or motives related to turnover may be indicative of a persistent orientation toward work across multiple jobs, as noted in previous work on the consistency of job attitudes (e.g., Newton & Keenan, 1991; Staw & Ross, 1985). This approach to coding open-ended information as indicative of stable characteristics has a long history (Lee & Peterson, 1997; Spangler, 1992). Like previous text coding research, our approach looks at attributions for previous events (Burns & Seligman, 1989; Staw, McKechnie, & Puffer, 1983). Machine learning approaches circumvent the unreliability of alternative approaches through a standardized method of coding. Machine learning allows us to identify words or phrases that signal applicants' relatively stable psychological characteristics based on a priori categories. For example, an applicant can write that s/he left the previous job because of excessive stress or poor working conditions. This means the person was seeking to avoid a bad job, although he or she did not explicitly use words like "leaving a bad job," or more abstract theoretical terms like "avoidance motive." Although there are several contextually specific reasons for leaving previous jobs, such as continuing education, relocation, or caregiving, we only focus on attributed reasons that signal relatively stable behavioral and attitudinal characteristics.

To systematically evaluate reasons applicants describe for leaving previous jobs in our data, we use supervised machine learning techniques. We trained a small sample of data (3% of the data) and manually categorized applicants' reported reasons for leaving a past position into four categories: (a) involuntary, (b) avoiding bad jobs, (c) approaching better jobs, and (d) other reasons. This is shown in Table 1. We then trained the algorithm on these data and it "learned" to evaluate the probabilities that different words and word combinations predict belonging to each category. Then it read the remaining data, which we call the test sample (97% of the data), and applied what it learned from the co-occurrence of terms and phrases and their probability distributions over the four categories of reasons in the training sample to recognize the semantic patterns and themes in applicants' reported attributions for leaving previous jobs. The algorithm then delivered probabilities that each text aligns with each of the four categories.

It is reasonable to believe that applicants might not disclose the actual reason for their leaving. However, in this study, the main

Table 1
Sample of the Training Dataset

Attributions for turnover	Reasons for leaving
Involuntary	Low student enrollment budget hold back from state
Involuntary	School closed due to low enrollment
Involuntary	Reorganization after turnaround transferred management back to Dutch owners
Involuntary	Company went under due to economic situation
Involuntary	Position eliminated due to recession
Avoid a bad job	The school wasn't a good fit for my teaching style
Avoid a bad job	I was unhappy and I resigned my position
Avoid a bad job	I was pretty much burned out
Avoid a bad job	Air pollution no health insurance low pay
Avoid a bad job	Bad management not enough hours
Approaching a better job	Interested in having a more challenging position
Approaching a better job	I'm interested in education and am now pursuing my dream
Approaching a better job	I love working with kids my passion is in teaching and promoting learning
Approaching a better job	A new professional challenge and an opportunity for professional growth
Approaching a better job	Advancement in career opportunity to grow personally and professionally
Other	Birth of my daughter
Other	I had a baby
Other	Relocated for family illness
Other	Husband's job was transferring
Other	Began Master of Education program

question is not whether the applicants reported the verifiable reasons organizations record for turnover, but whether applicants' self-reported attributes provide information useful in predicting future performance and turnover. Our model focuses on psychological processes filtered through self-presentation motives, memory distortions, and attributional biases to produce a description of prior turnover. Moreover, our hypotheses focus not primarily on answering theoretical questions about the variables we have included in our model, but rather, whether machine learning algorithms grouped around theoretical concepts established in prior research can predict subsequent work behavior.

History of involuntary turnover. An attribution of involuntary turnover reflects a situation in which the applicant reports that the organization made the decision to end the employment relationship. As an example of our machine learning process, reasons related to involuntary turnover include "I was laid off due to a budget cut," "my position was eliminated because of budget cuts," or "my position was eliminated and I was excessed." Words and phrases associated with budget, cut, eliminate, position, and layoff are expected to co-occur in reasons related to involuntary turnover. In our machine learning approach, algorithms learn these words and phrases are related to one another in a training sample, and applies the rule on the rest of the data by searching for similar relationships among words in the test sample. The algorithm then categorizes each individual reason for leaving given by applicants into corresponding categories by calculating the probability distribution of that information. The algorithm repeats this process in the test sample and finds their probability distributions over the categories predefined in the training sample.

Although we cannot be certain that individuals in our sample did experience involuntary turnover, we can use their attributions of involuntary turnover to predict behavior. Such attributions are important because they may reflect administrative decisions, but also reflect a tendency for an individual to focus on the role of external actors in shaping their behavior. The first issue relates to the more factual component of involuntariness. Several studies have found that employees who involuntarily leave their jobs tend to be lower performers compared to those leaving voluntarily (Barrick, Mount, & Strauss, 1994; Barrick & Zimmerman, 2009). Even in the case of layoffs, the selection of which individuals are terminated is often reflective of poor performance. Davis, Trevor, and Feng (2015) further note that individuals who have a history of being laid off tend to have more negative attitudes toward subsequent jobs, and in turn, are more likely to quite these subsequent jobs. The second issue relates to internal mental processes which drive individuals to describe their prior turnover to involuntary components (Wang, Bowling, & Eschleman, 2010). Such descriptions are consistent with a lower sense of personal responsibility and agency. Machine learning scores differentiate a statement "I was laid off due to budget cuts in my last job," from "After I was laid off, I sought a job that better matches my career goals." The former statement is scored as a totally involuntary attribution whereas the latter statement is scored as only a partially involuntary attribution because it blends an external attribution with personal control. In many other cases, the extent to which the employer or employee is truly the initiator of a job separation is not so clear-cut. In such ambiguous situations, individuals who are less agentic will recall and describe the event as completely involuntary, whereas an agentic individual may recall and describe the event as purely voluntary. The combination of the characteristics that lead an employer to terminate an employment relationship as well as the characteristics of an individual to recall and report this termination as mostly or entirely involuntary point in the same direction:

> *Hypothesis 3:* Applicant attributions of previous turnover as involuntary, as assessed via supervised machine learning, is (a) negatively associated with performance, and (b) positively associated with voluntary turnover hazard.

History of avoiding bad jobs. An extensive research tradition has differentiated individuals based on their long-term, dispositional motivational orientations. Scholars have come to find an important distinction between an "avoidance" disposition and an "approach" disposition (Elliot & Thrash, 2010). An avoidance disposition is associated with a tendency toward noticing negative or threatening features of the environment, experiencing anxiety when confronted with negative information, and behavioral attempts to avoid the resulting negative emotional stimuli (e.g., Diefendorff & Mehta, 2007; Elliot & Harackiewicz, 1996; Ferris et al., 2011). A focus on avoiding negative outcomes has been linked to attention to minimal standards of job performance, characterized by trying to find "minimally sufficient" levels of effort (Förster, Higgins, & Idson, 1998). While individuals with a strong avoidance focus may be able to complete core job tasks at a very basic level by showing up on time and completing strictly defined duties, efforts to innovate, exert extra effort, or seek advancement in one's career generally suffer (Elliot & Harackiewicz, 1996; Elliot & Sheldon, 1997). Moreover, it is also possible that individuals who attribute previous quitting to problems with their former workplace are behaviorally prone to externalize blame for negative events (Maier & Seligman, 2016). An avoidance focused attribution for job changes will also be associated with higher probability of turnover. Evidence clearly suggests that a disposition toward avoidance motivation is associated with lower levels of job satisfaction (Lanaj, Chang, & Johnson, 2012) which is a key antecedent of turnover (e.g., Schleicher, Hansen, & Fox, 2011; Trevor, 2001). Individuals who are avoidance focused will also be more prone to exit a job when problems arise, based on their generalized tendency to cope with problems by avoiding them.

> *Hypothesis 4:* Applicant attributions of previous turnover to avoiding bad jobs, as assessed via supervised machine learning, is (a) negatively associated with performance, and (b) positively associated with voluntary turnover hazard.

History of approaching better jobs. Several studies describe how an approach motivational orientation is positively associated with positive work outcomes (e.g., Diefendorff & Mehta, 2007; Elliot & Harackiewicz, 1996; Ferris et al., 2011). The approach orientation can represent itself in seeking a better fit, following one's passion, or looking for opportunities for advancement and development. Wrzesniewski, Dutton, and Debebe (2003) identified that interpreting one's work as a calling to be sought out is linked to more enjoyment, greater satisfaction and spending more time at work which all result in better performance and lower levels of turnover. Other studies found that a positive desire for one's work can positively contribute to long-term performance (e.g., Baum & Locke, 2004; Bonneville-Roussy, Lavigne, & Vallerand, 2011; Vallerand et al., 2008). Other studies have found that people who framed their work positively (e.g., as having positive effects on others) were more effective and more resilient in the wake of setbacks (Blatt & Ashford, 2006).

> *Hypothesis 5:* Applicant attributions of previous turnover to approaching better jobs, as assessed via supervised machine learning, is (a) positively associated with performance, and (b) negatively associated with voluntary turnover hazard.

Method

Data and Sample

We used data from 16,071 external applicants for teaching positions at the Minneapolis Public School District (MPS) between 2007 and 2013. The district hired 2,225 of the applicants. Of these, 1,756 stayed with the district at least until the 2012–2013 academic year, when the district introduced its teacher-effectiveness evaluation system, data from which we obtained our performance measures. Institutional review board (IRB) approval was granted by University of Minnesota (IRB Protocol #: 1510S79046, study title: Improving Human-Resource Management in Urban School Districts).

MPS is one of the largest school districts in Minnesota, serving over 30,000 students each year and employing around 2,800 total teachers in recent years. The school district teaching staff is 86% White, 6% African American, 2% Hispanic, 4% Asian, and 1% Native American. It serves a diverse population of students, many of whom are economically disadvantaged. MPS serves about 70% students of color, 21% English language learners, and 65% students eligible for free or reduced-price lunch.

The district publicly posts vacancy announcements. Typical teaching positions include elementary, high school math, or special education. Applications are submitted via a series of electronic forms that elicit semistructured text similar to that commonly found on a resume. The central human-resources department does a light screening to ensure each applicant meets minimal qualifications, such as having required licenses. School-based hiring teams conduct interviews and make offers. According to the district, more than 90% of offers are accepted.

For each application, we have data on position and self-reported applicant characteristics. These included a detailed work history with job title, job description, reason for leaving, and start and end dates for each previous job. Half of applicants' past positions were in 51 teaching occupations (their titles include the stem "teach*"). The other half were in 663 nonteaching O*NET occupations. The three most frequent were first-line supervisors of office and administrative support workers, educational guidance school counselors and vocational counselors, and social and human service assistants. Some applicants also disclosed race and gender, although this was not required. For hires working in the 2012–2013 academic year or after, we were able to link application information to performance data. We have information on turnover for all participants who were hired.

Measures

Work experience relevance. We automated measurement using four steps: (a) map past position job-title and job-description text to O*NET occupation codes, (b) map occupation codes to O*NET KSAO space, (c) measure distance in KSAO space between the past and desired position, and (d) average this distance across all the applicant's past positions using a weighting function that favors more recent and longer-held positions.

For the first step, we used supervised machine learning techniques to develop an algorithm that classified self-reported job titles and job descriptions into an O*NET standardized occupation code. Supervised classification is recommended for theory-driven studies and for cases in which a strong, external, ground-truth dataset exists on which to train the algorithm. Such classifiers were developed by learning the characteristics of different classes from a training sample of preclassified documents (Feldman & Sanger, 2007; Mohri et al., 2012). We used a naive Bayes Classifier, the most prevalent text classifier in machine learning (Feldman & Sanger, 2007; Mohri et al., 2012). We trained the algorithm against the ground truth of the official O*NET (O*NET, n.d.) occupational descriptions and job titles. We trained the classifier using a dataset comprising O*NET's detailed job descriptions and alternative job titles for each of 974 occupations as displayed on O*NET's web page. From the text on each occupation's O*NET web page, we made a "bag of words" for each O*NET standard occupation containing its description and commonly reported job titles associated with the occupation. We then trained the classifier on these data. Technical details are included in the Appendix.

Next, we ran the trained algorithm on the self-reported job description and job title from each applicant regarding a past position. The algorithm mapped this to a standardized O*NET occupation. To validate the classification, we took a random sample from the self-reported previous jobs, and hired a research assistant to classify the job descriptions into O*NET occupations. A senior undergraduate student in HR classified a random sample of 500 self-reported job titles and job descriptions. She searched for each self-reported job in O*NET online database, read the descriptions of occupations, and decided which O*NET occupation is closest to the self-reported job title and job description. We compared the predicted occupation from the naive Bayes classifier with the Research Assistant' (RA)'s classification to calculate the agreement rate between human and machine classifications. They agreed in 92% of the cases in the sample.

In the second step, each past position's standard occupation was mapped to a point in KSAO space. O*NET provides detailed information about the required level and/or importance of different abilities, knowledge, skills, vocational interests, values, and styles for each occupation. This gave each occupation-*o* a profile, x_o, in a high-dimensional KSAO space.

In the third step, we operationalized work-experience relevance with a profile similarity index (PSI), measuring the similarity between an applicant's past occupation and the occupation sought. A PSI is a single value representing the extent to which a past occupation and the prospective one are (dis)similar across multiple variables (Converse et al., 2004; Edwards & Harrison, 1993). We used *profile level* which measures dissimilarity and measures the extent to which scores in one profile tend to be higher or lower than scores within another profile. As is common, we used the L2 (Euclidean) distance between the two profiles (Converse et al., 2004; Edwards & Harrison, 1993). Letting *a* index the past position, *b* index the desired position, *i* index the dimensions of KSAO space, the profile level measures dissimilarity as,

$$R_a = -\left(\sum_i (x_{ai} - x_{bi})^2\right)^{\frac{1}{2}} \tag{1}$$

To measure relevance R_a, we reverse coded distance, using the negative sign above.

Finally, to aggregate information across an applicant's entire work history, we take a weighted average of *R* across all the applicant's past jobs. Applicants in our sample average 3.18 previous jobs ($SD = 2.2$). To account for the duration and recency of experiences, we define a weight for each previous job as the integral of the decay function of both the elapsed time since the person left the previous job (E_a) and their tenure in that job (T_a). The weight accorded to past position-*a* is,

$$w_a = \int_{E_a}^{E_a+T_a} e^{-rx}dx \tag{2}$$

A higher decay parameter (*r*) implies a faster rate of decay in KSAOs over time. For $r = 1$, current KSAOs become irrelevant in about 5 years. For $r = .5$, it takes about 10 years. In Minnesota, where the school district is located, teaching licenses expire after 5 years, and renewal requires more than 100 hr of professional development. We chose $r = 1$ to match this 5-year decay. The correlation between weights yielded from a decay function with $r = 1$ and $r = .5$ is 0.95 and the correlation between work-experience relevance variables created using those weights is 0.97. Our aggregate measure of work-experience relevance is the w_a-weighted average R_a across the applicant's past positions, which we then standardized across applicants.

Tenure history. We defined tenure history as the average deviation of applicant's tenure in prior jobs from the median tenure in each occupation category. Barrick and Zimmerman (2005) used average tenure in previous jobs to measure this. However, average tenure differs across occupations because of occupational characteristics unrelated to individual tendencies. To address this issue, we collected median tenure in an applicant's relevant prior occupation category, reported on the department of labor's website (United States Department of Labor, Bureau of Labor Statistics, n.d.). For each past position, we computed the difference between the applicant's tenure and relevant median tenure. Each applicant's tenure history is the average deviation across prior positions.

Attributions for turnover history. To systematically categorize self-reported turnover reasons, we hand-coded a randomly selected set of attributions into each of the main three topics (i.e., involuntary, approaching a better job, and avoiding a bad job) or a fourth, residual category capturing all other topics. In the training data, if the applicant's turnover attributions provided more than one topic, we broke the text into specific topics (additional lines in the training sample). In the process of improving our training sample, we started with manually coding a random sample of 100 attributions, trained the algorithm on that sample, and ran the trained algorithm on a validation sample that we had set aside. We increased the size of the hand-coded training sample gradually and repeated the process until we obtained algorithm accuracy of 90%, which occurred with a training sample size of 1,000 attributions. To ensure that the final training sample was accurately hand-coded by the authors, the RA rated a random sample of 100 from the final training sample. The agreement between the RA rating and our rating was 100%. At this point we trained a supervised naive Bayes classifier on the final training sample. Next, we had the trained classifier algorithm score each of the 34,601 turnover attributions in the whole dataset. For each past job, this delivers a probability distribution that sums to one across the four possible attributed reasons for leaving. To check the accuracy of this classification, the RA categorized a random sample of 350 classified turnover attributions that were not in the training sample. There was a 93% agreement between the machine and RA classification.

To assess differences among attributions across past positions for applicants with multiple past positions, we split the sample between the current or most-recent position and the others. Compared with other prior positions, applicants attributed a higher share of the most-recent position exits to the reasons of approaching a better job (0.26 vs. 0.18) and avoiding a bad job (0.19 vs. 0.13). The share attributed to involuntary turnover does not differ. The share attributed to other causes is lower for the most-recent positions.

Table 2 presents examples of turnover attributions classified using supervised machine learning along with a probability distribution over these attributions. The four probabilities across the four attributions add up to 1 for each textual explanation provided by the applicant why they left a past job. This allows us to capture if explanatory text expresses a mix of attributions with probabilities between 0 and 1. For the applicant who left her previous job because she was "looking to return to public school employment the atmosphere and professional climate at a private parochial school does not fit with [her] views and philosophies of education," the algorithm assigns a probability of 0.47 to "approaching a better job" variable, and a probability of 0.53 to "avoiding a bad job" variable, and 0 for the other attributions. This model does not measure the extent to which or intensity with which one describes a particular attribution for turnover, but rather, the probability that the attribution fits into a category.

Performance. In the 2012–2013 academic year, the district adopted a well-tested, comprehensive system of multiple measures of performance to evaluate teaching performance (Kane, McCaffrey, Miller, & Staiger, 2012). These measures included the following.

Student evaluation. The district administered a survey to all students about their teachers twice each year starting in the 2013–2014 academic year. The questions asked students about the degree to which their teachers academically "engage," "illuminate," "manage," "relate," and "stretch" them and their peers. This survey is based on the Tripod Seven C's survey of teacher practice (Kane et al., 2012). Items and teachers are scored on a "favorability" metric. That is, items are scored 1 if a student responded "yes" in Grades K–2 or "mostly yes" or "yes, always" in Grades 3–12. Responses of "no" or "sometimes" in Grades K–2 or "no, never," "mostly no," or "maybe/sometimes" in Grades 3–12 were scored 0. A teacher's score is simply the mean of their dichotomous item scores multiplied times 100, resulting in a score between 0 and 100. Two examples of items used in the survey are: "This class makes me a better thinker" and "The teacher in this class really cares about me."

Expert observations. Measures of effective instruction were scored after classroom observations four times each year by trained, certified raters against a rubric of effective instruction based on the widely used Framework for Effective Teaching (Danielson, 2007). The raters evaluate teacher performance using a 20-item scale. All items used 4-point Likert-type scales with anchors of 1 (*strongly disagree*) to 4 (*strongly agree*). Examples of items included are "Plans units and lessons effectively" and "Uses relevant resources and technology."

Value-added. This measure of teacher performance was based on students' standardized achievement tests in reading and/or math and student–teacher links based on teacher-verified rosters controlling for each student's prior achievement level and other characteristics. This measure was only available for teachers who have taught math or reading since 2012. This measure was developed in an association between the school district and the Value-Added Research Center (VARC) at the University of Wisconsin (Value-Added Research Center & Minneapolis Public Schools, 2013). The model is based on a posttest-on-pretest regression, so the value-added scores represent a model of growth in student achievement over the course of a year of instruction. The

Table 2
A Sample of Classifying Reasons for Leaving Into Four Categories of Attributions for Turnover Using Supervised Machine Learning

	Probability distribution over attributions for turnover			
Reasons for leaving: Representative statements	Approach better job	Avoid bad job	Involuntary turnover	Other reasons
Interested in expanding my professional career in a diverse setting where my skills and commitment to education will serve the students, parents and district	1	0	0	0
I miss working with students face-to-face and would like to work in an urban setting	1	0	0	0
Was not satisfied with the high caseload and hours; on-call work	0	1	0	0
Dissatisfied with pay same as subbing and environment	0	1	0	0
Position was eliminated at the end of the school term due to budget cuts	0	0	1	0
My contract was not renewed	0	0	1	0
I am looking to return to public school employment the atmosphere and professional climate at a private parochial school does not fit with my views and philosophies of education	.47	.53	0	0
I moved on to a new employment opportunity at [name of the school] where I could learn more about serving clients with disabilities. [name of the school] did not provide this learning opportunity.	.67	.33	0	0
This is a one academic year position that is grant funded. I have a desire to return to the classroom as a teacher	.82	0	.18	0
The district did not renew my contract for the school year. I am interested in working with students in a diverse setting that is both challenging and rewarding	.86	0	.14	0
Not tenured after 3 years at XXX. Different supervisors during probationary period. Unclear how to meet expectations	0	.29	.71	0
Evaluation team was dissolved and the job duties changed	0	.46	.54	0
I was graduating from college and moving to a new location to begin graduate school	0	0	0	1
Sought employment closer to home after birth of child	0	0	0	1

model also includes controls for student characteristics and incorporates multiple pretests when available.

The district created a z-score for each teacher-year observed using the cross-sectional distribution of each measure among the district's teachers. Because our sample is new hires and there is a learning curve in teaching, the sample's average performance is below the district average. Each score also has a standard error, which depends on the reliability of the measure and the amount of information available for that teacher's measure, such as the number of a teacher's students responding to the surveys or taking the standardized tests. To aggregate information across measures, the district uses a composite measure of teacher performance computed with inverse-variance weighting. We used all of these four measures of performance (student evaluation, expert observations, value-added, and the composite) as dependent variables to compare the predictive validity of our predictors for various performance measures.

Our analysis is fundamentally cross-sectional because we are studying a one-time hiring decision. We constructed a measure incorporating information from many years of posthire performance. To compare hires' performance on an equal footing despite their being observed during different spells of experience, we residualized each performance score (Z_{itm}) conditional on a measure-specific quadratic regression model of teacher-i's years since hire in year-t (X_{it}) (Papay & Kraft, 2015; Wiswall, 2013) using all observed teacher-years of performance for the measure-m. Then we scored the residual for each observation: $Z_{itm} - E(Z_{itm} \mid X_{it}, X_{it}^2)$. For teacher-i with $N_{im} > 0$ observations on performance measure-m, we measure performance as the average of residualized performance:

$$Y_{im} = N_{im}^{-1} \sum_{t=1}^{N_{im}} [Z_{itm} - E(Z_{itm} | X_{it}, \quad X_{it}^2)] \tag{3}$$

Turnover. Among hires, we have access to the hire date and, if applicable, a turnover date and reason (voluntary or involuntary). Voluntary turnover reasons included "personal reasons," "not eligible extend LOA," "health reason," and "educator in another district or state." Involuntary turnover reasons included "discharged," "probationary release-performance," "resigned in lieu of termination," "discontinuance of contract," "end temp assignment," "inactive," "lay off," "license/certification require," or "probationary release, staff reduction." We used survival analysis to calculate (voluntary or involuntary) turnover hazard, defined as the expected speed of turnover (Dickter, Roznowski, & Harrison, 1996). To measure turnover hazard, we also used employment duration in years. Turnover hazard indicates whether and when the employee turned over (Dickter et al., 1996; Singer & Willett, 2003). When predicting voluntary (involuntary) turnover, the applicants who were terminated involuntarily (voluntarily) were treated as censored observations. A total of 349 individuals, or 16% of the sample of 2,225 applicants who were hired between 2007 and 2013 voluntarily turned over. The duration of employment for those who voluntarily turned over ranged from 1 to 9 years (M = 3.48 years, SD = 1.68 years). A total of 398 individuals, or 18% of the sample of 2,225 hires involuntarily turned over. The duration of employment for them ranged from 1 to 9 years (M = 4.08 years, SD = 1.85 years). Employees who had not turned over by 2017 are analyzed as right censored.

Control variables. In addition to our hypothesized predictors related to KSAOs, we also included related variables in our model that are potentially correlated with our measures. This was done to both guard against spurious inference compare the relative magnitude of our new predictors from the machine learning system to those that have traditionally been employed.

Applicant general writing skill might correlate with the quality of their application and job performance, but is not relevant to our core hypotheses. The qdap and hunspell packages in R were used to count spelling errors in each application. We reverse coded the resulting variable such that higher scores reflected fewer mistakes. We controlled for whether applicants have an advanced degree for similar reasons.

We included whether applicants had worked as a teacher in the past, because this may exert an influence on several performance ratings beyond the similarity of skills (e.g., teachers as raters may have in-group biases toward those who have prior teaching experience). For similar reasons, we also controlled for whether they have been the district's employee before in any position. We controlled for overall years of work experience, because this potentially relates to several of our central variables and performance. The average employment gap between previous jobs was included since it may relate to employment history but is not central to our hypotheses.

Due to differences in performance ratings across jobs, we incorporated the type of teaching position (special education, science, math, reading, elementary school teacher, social science, and others) in our regressions. Because our sample started working for the district at different points, and this influences duration of observation, we controlled for application year.

Demographic variables. We did not control for race or gender because we aim to introduce a selection model independent of these variables. As such, the main results presented in our regression analyses do not incorporate them. However, we note that we also ran contrasting models that did include these demographic factors, and found that the results were nearly identical to those from our selection model, with no changes in the pattern of significance and only small changes in the magnitude of effects for our hypothesized predictors.

To evaluate the effectiveness of our proposed model in reducing the risk of adverse impact, we assessed whether the predictive ability of demographic variables in determining selection changed under our proposed model relative to previous hiring methods. In our sample, 37% of applicants did not self-report their race and gender. Because we cannot determine whether the demographic values are missing completely at random or because a specific group of people chose not to reveal their demographic characteristics, we cannot drop applicants who did not report their demographic information. For that purpose, using Minnesota statewide administrative data that includes name, gender, and race for all teachers in the state, we built a reference database to train a supervised algorithm to classify applicants with missing gender into female and male categories, and classify applicants with missing race into white and nonwhite categories. Our data do not allow use of more-refined ethnic subgroups because our training sample is insufficiently diverse to develop reliable classifications. To validate the accuracy of our algorithm, we took a random sample of 100 hires who did not self-report their race and gender at application but did have demographic information in the district's administrative data. The race and gender retrieved from administrative data matched with the algorithm classification with 95% accuracy. We did not use race and gender in our predictive model, but we used them to evaluate whether our model reduces the risk of adverse impact.

Correction for Sample Selection Bias and Instrumental Variables

Because applicants went through a nonrandom selection process to be hired, estimates from an OLS regression of work outcomes on predictors *among hires only* might produce estimates suffering from omitted-variable bias and range restriction (Sackett & Yang, 2000). To correct for this, we used a Heckman selection correction (Heckman, 1979). As instrumental variables, we used the quality and quantity of the competition an applicant faced in applying for the position, both of which will affect an applicant's chance of being hired, but are uncorrelated with unobserved applicant characteristics. These instruments shift an applicant's probability of hire, but do not affect posthire performance or turnover. Similar instruments have been used previously in the context of teacher selection (Goldhaber, Grout, & Huntington-Klein, 2014).

To measure the quantity of an applicant's competition, we calculated the share of applicants hired for the position. To measure the quality of the competition, we ran a probit model using all predictors and control variables from the applicant pool to predict the likelihood of being hired. For each applicant, this yields a predicted probability of hire. The average predicted hire probability of their competitors indexes the quality of the competition.

Measurement Properties of Machine Learning-Derived Variables

We took several steps to ensure that our machine-learning approach generated measures of key constructs with desirable psychometric properties (McKenny, Aguinis, Short, & Anglin, 2018). First, we assessed test–retest reliability in applicants' attributions for turnover across jobs. For individuals who had more than one previous job (91% of our sample), we divided individual's jobs into two random subsets, computed the average within each person-attribution-subset, and calculated the correlation between the two subsets within person-attribution. This produced a correlation of $r = .58$ for involuntary turnover, $r = .56$ for avoiding bad jobs, and $r = .58$ for approaching better jobs across the two subsets. Given the fact that respondents are evaluating different prior jobs in each subset, this moderate level of correlation is consistent with a dispositional attribution style that may carry over within person across jobs.

To generate evidence on the reliability of our machine-learning classifications, we first tested the accuracy of several alternative classification algorithms. We randomly partitioned the training sample, where we have reliable measures of ground-truth from human coding, into 10 equally sized subsamples. We used each subsample as a test dataset once and trained all algorithms on the other 90%. We scored the test subsample under each trained algorithm and recorded the proportion of correspondence of the subsample to the human coding. For each algorithm, we averaged these proportions across the 10 test samples to measure correspondence. The naive Bayes estimator we used achieved a very high correspondence rate of 98.8%, which compares favorably relative to Logistic Regression (98.7% correspondence), a Decision Tree (89.7%), a Random Forest (88.5%), or K-Nearest Neighbor (76.3%). Consistency across algorithms reflects "inter-algorithm" reliability, akin to interrater reliability.

Results

Table 3 and 4 present the intercorrelations and the descriptive statistics for the study variables. All continuous variables are standardized. Table 4 reports variable summaries before standardization.

Linking Work History to Performance and Turnover

To predict each of the four measures of performance, we estimated a Heckman regression using Stata 14 using Maximum Likelihood. This is preferred over OLS analysis in the hired-only subsample due to the threat of omitted-variable or selection bias created by the fact that outcomes are observed only for those who are hired (Clougherty, Duso, & Muck, 2016; Wooldridge, 2010). If unobserved determinants of performance are correlated with predictors of hire, estimates from OLS will be biased. Our approach corrects for this by harnessing instrumental variables that shift each individual's probability of hire but are not related to unobservable determinants of her performance. Table 5 compares the first stage of Heckman model to a similar probit model excluding the instruments. The first column reports estimated effects of different predictors on the probability of getting hired from a probit. The second column adds the instruments we have defined, the quality and quantity of competition faced by each job applicant. These are strong predictors of the probability of getting hired but should not be related to unobserved determinants of individual performance or turnover conditional on hire.

Table 6 shows the estimated outcome models, the Heckman second stages, with columns varying only the posthire outcome. In partial support of Hypothesis 1a, they show that work experience relevance (Hypothesis 1a) is positively associated with expert observations, value-added, and the performance composite ($\beta_{\text{work experience relevance-Expert observation}} = 0.05$, $p < .01$; $\beta_{\text{work experience relevance-Value-Added}} = 0.11$, $p < .01$; $\beta_{\text{work experience relevance-Performance composite}} = 0.05$, $p < .01$), but not with the student evaluation of teacher performance. Also, in partial support of Hypothesis 2a, tenure history has a significant negative effect on expert observations, value added, and the performance composite ($\beta_{\text{Tenure history-Expert observation}} = 0.08$, $p < .01$; $\beta_{\text{Tenure history-Value-Added}} = 0.08$, $p < .05$; $\beta_{\text{Tenure history-Performance composite}} = 0.07$, $p < .05$). Again, there is no evidence supporting that tenure history has any impact on the students' evaluation of teacher performance. Leaving previous jobs due to involuntary turnover (Hypothesis 3a) only predicts expert observations and performance composite ($\beta_{\text{Involuntary turnover-Expert observation}} = -0.06$, $p < .05$; $\beta_{\text{Involuntary turnover-Performance composite}} = -0.07$, $p < .01$). Leaving to avoid a bad job (Hypothesis 4a) is negatively related to student evaluations, expert observations, value added, and the performance composite ($\beta_{\text{Avoid bad-Student evaluation}} = -0.14$, $p < .01$; $\beta_{\text{Avoid bad-Expert observation}} = -0.17$, $p < .001$; $\beta_{\text{Avoid bad-Value-Added}} = -0.11$, $p < .001$; $\beta_{\text{Avoid bad-Performance composite}} = -0.18$, $p < .01$). Finally, fully supporting H5a, leaving to seek a better job (H5a) is positively associated with all the performance measures ($\beta_{\text{Seek better-Student evaluation}} = 0.09$,

Table 3
Intercorrelations for the Study Variables

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
Outcome variables																			
1. Student evaluation	1.00																		
2. Expert observation	.35	1.00																	
3. Value-added	.13	.25	1.00																
4. Performance composite	.35	.96	.34	1.00															
5. Voluntary turnover	−.10	−.13	−.05	−.16	1.00														
6. Involuntary turnover	−.09	−.17	−.04	−.18	−.20	1.00													
7. Work experience relevance	−.05	.05	.06	.05	−.10	.02	1.00												
8. Tenure history	−.02	.11	.08	.15	−.12	.02	.09	1.00											
History of leaving previous jobs																			
9. Involuntary turnover	−.00	−.03	−.01	−.03	−.06	−.03	.10	−.08	1.00										
10. Avoiding bad jobs	−.14	−.22	−.13	−.22	.03	.12	−.01	.00	−.09	1.00									
11. Approaching better jobs	.13	.13	.10	.15	−.11	−.01	.04	.10	−.24	−.13	1.00								
Instruments																			
12. Competition-Quantity	.01	−.05	.02	−.07	.11	−.03	−.14	−.41	−.09	−.01	−.05	1.00							
13. Competition-Quality	−.03	.06	−.03	.06	−.07	.02	.08	.32	.03	−.04	.07	−.59	1.00						
Control variables																			
14. Spelling accuracy	.07	.03	.03	.04	.00	−.03	−.02	.11	−.09	−.07	.01	.06	−.00	1.00					
15. Years of experience	−.03	.03	.04	.07	−.14	.08	.13	.61	.03	.01	.06	−.41	.26	−.19	1.00				
16. Prior district employment	−.04	.05	−.01	.09	−.10	.03	.17	.27	.12	.00	.09	−.33	.25	−.06	.26	1.00			
17. Prior work as a teacher	.00	.05	−.02	.06	−.04	−.03	−.03	.02	.07	−.02	−.06	−.01	.06	−.00	−.04	.23	1.00		
18. Advanced degree	.03	.07	−.00	.09	−.07	.06	.08	.27	.04	−.02	.08	−.26	.22	−.06	.32	.16	−.01	1.00	
19. Employment gap	.03	.00	.02	.01	−.06	.01	.07	.02	.01	−.01	−.02	−.14	.06	.02	.27	.04	.01	.06	1.00

Note. Values greater than or equal to .07 are significant at $p < .05$.

Table 4
Descriptive Statistics for the Study Variables

Variable	*N*	Mean	*SD*
Outcome variables			
Performance composite	1,756	−.17	.75
Expert observation	1,728	2.92	.25
Student evaluation	1,342	82.71	6.14
Value-added	866	2.98	.63
Voluntary turnover	2,225	.16	.36
Involuntary turnover	2,225	.18	.38
Work experience relevance	16,071	16.07	4.93
Tenure history	16,071	−1.66	4.5
History of leaving previous jobs			
Involuntary turnover	16,071	.15	.23
Avoiding bad jobs	16,071	.13	.19
Approaching better jobs	16,071	.20	.26
Instruments			
Competition-quantity	16,071	.84	.13
Competition-quality	16,071	.14	.08
Control variables			
Spelling accuracy	16,071	.74	1.42
Years of experience	16,071	7.8	7.08
Prior district employment	16,071	.23	.42
Prior work as a teacher	16,071	.17	.38
Advanced degree	16,071	.47	.49
Employment gap	16,071	.44	.82
Demographic variables			
Female	16,071	.76	.42
White	16,071	.84	.37
Age	16,071	33.12	10.62

$p < .05$; $\beta_{\text{Seek better-Expert observation}} = 0.09$, $p < .01$; $\beta_{\text{Seek better-Value-Added}} = 0.09$, $p < .01$; $\beta_{\text{Seek better-Performance composite}} = 0.09$, $p < .01$). A negative coefficient on the inverse mills ratio (IMR), as in the models for expert observation and the performance composite ($\beta_{\text{IMR-Expert observation}} = -0.10$, $p < .05$; $\beta_{\text{IMR-Performance composite}} = -0.09$, $p < .001$), gives evidence that unobservable factors which increase hiring probability tend to push down these outcomes. Given conceptual similarities in our variables, we estimated variance inflation factors (VIFs) for the variables in our models to evaluate the risk of multicollinearity in our models (Wooldridge, 2010). All the VIFs ranged between 1.05 and 2.03, which is well below the range indicating strong collinearity problems.

To estimate the hazard function for voluntary and involuntary turnover, we use the Cox partial likelihood method (Singer & Willett, 2003). We also corrected for selection bias in these models by including the inverse Mills ratio from the Heckman model as a proxy for unobservable determinants of hire. Table 7 reports the results. The hazard function of the Cox model is given by $r(\text{t}, \text{x}) = \text{h(t)}\, e^{\hat{a}x}$, where h(t) is the baseline hazard, x is a vector of covariates, and β is a vector of regression coefficients. The Cox method is a semiparametric approach not requiring any assumption about the distribution of the hazard function. However, the hazard functions should be proportional for different covariates, so that the effects of the covariates on the criterion does not change over time (Cleves, Gould, & Marchenko, 2016). To test this assumption, we run the Grambsch and Therneau (1994) maximum likelihood test. We failed to reject the null hypothesis ($p = .14$) that the log hazard-ratio function is constant over time, which suggests our model did not violate the assumption required for Cox model.

Results presented in Table 7 show that one standard deviation increase in work experience relevance (Hypothesis 1b) is predicted to decrease voluntary turnover hazard by 8% ($H_{\text{Work experience relevance-Voluntary turnover}} = 0.92$, $p < .001$), whereas the same change in the tenure history (Hypothesis 2b) is predicted to decrease voluntary turnover hazard by 11% ($H_{\text{Tenure history-Voluntary turnover}} = 0.89$, $p < .05$). We found an opposite relationship as we hypothesized (Hypothesis 3b) between leaving previous jobs due to involuntary turnover and hazard of voluntary turnover, showing a negative link between leaving due to involuntary turnover and the hazard of voluntary turnover ($H_{\text{Involuntary turnover-Voluntary turnover}} = 0.87$, $p < .01$). We do not find any support for Hypothesis 4b, which predicted a positive relationship between leaving prior positions to avoid bad jobs and the hazard of voluntary turnover. Finally, we do not find any support for Hypothesis 5b that there is a negative relationship between approaching a better job and the hazard of voluntary turnover. The results for the hazards of involuntary turnover and overall turnover are also reported in Table 7. The results support a positive relationship between tenure history and the hazard of involuntary turnover, so that one standard deviation increase in tenure history is linked to 13% decrease in the hazard of involuntary turnover ($H_{\text{Tenure history-Involuntary turnover}} = 0.87$, $p < .05$). Our results show a positive relationship between avoiding a bad job and the hazard of involuntary turnover, so that one standard deviation increase in avoiding a bad job is associated with 10% increase in the hazard of involuntary turnover ($H_{\text{Avoid bad-Involuntary turnover}} = 1.10$, $p < .001$). We do not find any support for a relationship between our other predictors and the risk of involuntary turnover. The relationship between the predictors and overall turnover are very similar to those for voluntary turnover, but weaker. Taken as an aggregate, these results show that our work experience variables predict outcomes above and beyond commonly used selection measures such as years of work experience, field specific experience, and prior work in the same organization.

Table 5
Selection Model Without and With Instruments

Variable	Hired	Hired
Work experience relevance	.12*** (.03)	.09*** (.02)
Tenure history	.08*** (.03)	.05 (.03)
History of leaving previous jobs		
Involuntary turnover	−.01 (.01)	−.02 (.01)
Avoiding bad jobs	−.02** (.01)	−.02* (.01)
Approaching better jobs	.05*** (.01)	.03*** (.01)
Control variables		
Spelling accuracy	.04*** (.01)	.03** (.01)
Years of experience	.02 (.01)	−.02 (.01)
Prior district employment	.99*** (.06)	.83*** (.04)
Prior work as a teacher	.44*** (.04)	.44*** (.04)
Advanced degree	.09** (.03)	.02 (.03)
Employment gap	−.02 (.02)	−.01 (.02)
Instruments		
Competition-quantity		−.45*** (.02)
Competition-quality		−.07*** (.02)
Controlled for application year and position type	Yes	Yes
R-squared	.19	.26
Observations	16,071	16,071

Note. Standard errors (in parentheses) are adjusted for seven clusters in application years.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 6
Models Predicting Different Measures of Teacher Performance—Heckman Second Stage

Variable	Student evaluation	Expert observation	Value-added	Performance composite
Work experience relevance	−.04 (.04)	.05** (.02)	.11** (.03)	.05** (.02)
Tenure history	.00 (.05)	.08** (.03)	.08* (.03)	.07* (.03)
History of leaving previous jobs				
Involuntary turnover	.01 (.02)	−.06* (.03)	.00 (.01)	−.07** (.03)
Avoiding bad jobs	−.14** (.06)	−.17*** (.02)	−.11*** (.02)	−.18** (.02)
Approaching better jobs	.09* (.04)	.09** (.03)	.09** (.03)	.09** (.04)
Inverse mills ratio	−.11 (.09)	−.10* (.04)	.23 (.13)	−.09*** (.04)
Control variables				
Spelling accuracy	.04*** (.01)	.01 (.01)	.03 (.03)	.02 (.01)
Years of experience	−.08* (.03)	−.09* (.04)	.02 (.02)	−.06 (.03)
Prior district employment	−.19* (.08)	−.06 (.16)	.07 (.12)	−.01 (.18)
Prior work as a teacher	.05 (.05)	.07*** (.02)	.07 (.04)	.07*** (.02)
Advanced degree	.02 (.02)	.18*** (.05)	−.02 (.04)	.19*** (.05)
Employment gap	.01 (.01)	.01 (.02)	.02** (.01)	.01 (.02)
Controlled for application year and position type	Yes	Yes	Yes	Yes
Observations	1,342	1,728	866	1,756

Note. Standard errors (in parentheses) are adjusted for seven clusters in application years. The numbers of observations are different across models because different performance evaluations started at different times, and were used for different position types.
* $p < .05$. ** $p < .01$. *** $p < .001$.

As a contrast, we aggregated our separate classification of approach and avoidance turnover in previous jobs into a single combined measure. We then ran supplemental analyses comparing our original analyses with this new combined variable in place of our separate measures of approach and avoidance. The combined measure was not predictive of any of the four outcomes, supporting the utility of our classifications as indicators of work-relevant constructs. Results of this supplemental analysis are available from the first author on request.

To clarify the contribution of the predictors we introduced in this study, we conducted a Wald-test comparing the model only with variables already discussed in the literature (years of work experience, whether the applicant has teaching experience, employment gap, etc.) and a model that also included our new predictors (work experience relevance, attributed reasons for leaving, and tenure history). The postestimation tests for the model of each outcome confirmed that the predictors we introduced are significant additions to the model with only variables that were previously introduced in the literature at levels below $p < .01$.

Evaluating the Effectiveness of Our Proposed Model

We evaluated the effectiveness of our proposed models in terms of (a) lowering the risk of adverse impact and (b) selecting higher performers or longer-serving hires. To do so, we developed a list of model-recommended hires based on hiring applicants with the best predicted post-hire outcome. We “recommended” the same number of applicants as the district hired each year in each position type. For example, if the district hired 100 of 300 applicants in 2013 for the position of special-education teacher,

Table 7
Survival Models Predicting Voluntary and Involuntary Turnover

Variable	Voluntary turnover	Involuntary turnover	All turnover
Work experience relevance	.92*** [.89, .96]	.96 [.88, 1.04]	.94* [.89, 1.00]
Tenure history	.89* [.79, .98]	.87* [.73, .99]	.88* [.77, .98]
History of leaving previous jobs			
Involuntary turnover	.87** [.78, .97]	1.03 [.96, 1.10]	.95 [.88, 1.02]
Avoiding bad jobs	1.02 [.95, 1.09]	1.10*** [1.06, 1.14]	1.06*** [1.03, 1.09]
Approaching better jobs	.94 [.87, 1.01]	1.00 [.93, 1.06]	.97 [.91, 1.02]
Inverse mills ratio	.93 [.77, 1.12]	.92 [.79, 1.07]	.92 [.83, 1.01]
Control variables			
Spelling accuracy	1.01 [.98, 1.03]	1.05 [.95, 1.15]	1.03 [.98, 1.08]
Years of experience	.95 [.82, 1.10]	1.13*** [1.07, 1.20]	1.05 [.98, 1.13]
Prior district employment	.71*** [.67, .79]	1.01 [.97, 1.23]	.89** [.83, .96]
Prior work as a teacher	.78** [.67, .91]	.88 [.76, 1.02]	.83*** [.77, .89]
Advanced degree	.97 [.86, 1.10]	1.23*** [1.16, 1.31]	1.10*** [1.05, 1.15]
Employment gap	1.03 [.97, 1.10]	.93* [.86, 1.00]	.98 [.95, 1.00]
Controlled for application year and position type	Yes	Yes	Yes
Observations	2,225	2,225	2,225

Note. Coefficients are hazard ratios. Confidence intervals in brackets.
* $p < .05$. ** $p < .01$. *** $p < .001$.

we recommend the 100 applicants who applied to that position in that year who our model predicts to have the highest levels of performance.

Adverse impact. Adverse impact is often operationalized by studying the ratio of hiring rate of minority group and hiring rate of majority group. The closer this number is to one, the lower the risk of adverse impact (Aguinis & Smith, 2007). In the result section, we report this ratio for applicants who are female versus male and people of color versus white. As an alternative standard, we compared the power of the demographic variables to predict hiring under the observed selection system and recommended hiring under our model using two simple Probit models. If our model lowers the risk of adverse impact relative to the district's real hiring decisions during the period studied, the demographic variables will have less explanatory power for our recommendations than for the observed hiring decisions. Results are reported in Table 8. In the first model, the outcome is whether the applicant actually was hired and in the other models it is whether the applicant is recommended for hire by each of our models. Predictors include gender and race dummy variables, age and age-squared, and control variables for application year and position type.

We found that for the female to male selection ratio, the conventional hiring rate for the school district has an adverse impact ratio of 0.99, which is essentially equivalent to the adverse impact ratio for our model of 0.99. For the non-White to White selection ratio, the conventional hiring rate for the school district would have an adverse impact ratio of 0.93, whereas hiring based on our model would have an adverse impact ration of 1.03. Therefore, the risk of adverse impact is relatively low for our overall model and is somewhat superior to the existing hiring method. Also, as shown in the last column of Table 8, gender, race, and age are each strong predictors of hire in the district's actual selection system ($\beta_{female} = 0.06, p < .05$; $\beta_{white} = 0.11, p < .01$; $\beta_{age} = 0.48, p < .001$; $\beta^2_{age} = -0.13, p < .001$). However, across our models of all posthire outcomes, following the model's recommendation would imply selection decisions where gender and race are not significant predictors of hire. For example, selecting on predicted composite performance yields no association of hiring with gender or race ($\beta_{female} = -0.02$, *ns*; $\beta_{white} = 0.02$, *ns*). Similar results are obtained across all models. Age is still a predictor of selection in our models, but its effect size is smaller than that in the district's observed selection model, (e.g., selecting on composite performance model we have: $\beta_{age} = 0.35, p < .001$; $\beta^2_{age} = -0.13, p < .001$).

Selection quality. To evaluate the effectiveness of our proposed model in terms of selecting high performers against the observed selection system, we trained each outcome model on a 90% subsample. We then predicted outcome scores for all observations in the remaining 10% hold-out sample, both hires and nonhires. Preserving the proportion of applicants hired and based on a ranking of predicted scores within position, we split the hold-out sample into recommended and not-recommended. Among actual hires in the hold-out sample, we compared actual scores between the recommended and not-recommended hires. We iterated this process 200 times.

Table 9 reports the average actual score difference between the recommended versus not-recommended subsets of hires in the hold-out sample. For example, the cell in the first row and first column shows that, when selecting to maximize the predicted performance composite, the average observed performance composite score of the recommended group is significantly higher than that of the not-recommended group (difference = 0.4, 95% CI [0.38,0.43]). Selecting to maximize the performance composite, also generates a positive difference between the recommended and not-recommend groups' average performances as measured by student evaluations (difference = 0.10, 95% CI [0.07,0.13]), expert evaluation (difference = 0.35, 95% CI [0.33,0.37]), value added (difference = 0.31, 95% CI [0.27,0.35]) and expected years of retention (difference = 0.0.65, 95% CI [0.11,0.17]). The top row of Table 9 communicates these results. The following four rows repeat this exercise but, instead of making recommendations to maximize predicted composite performance, recommendations are made to maximize predicted student evaluations, predicted expert evaluations, predicted value added, and predicted years of retention, respectively, and results describe how these induced recommendations affect the difference in each outcome.

Next, we compared the school district's actual hires with our models' recommended hires. The shares of actual hires that the model recommended for hire ranged between 11% and 29%, depending on which outcome we select on and which outcome we compare. These comparisons suggest better applicants were available than those hired in the vast majority of cases. To quantify the level of agreement that would be created by hiring at random given

Table 8
Probit Models Comparing the Change in the Risk of Adverse Impact

Variable	Recommended based on performance composite	Recommended based on student evaluation	Recommended based on expert observation	Recommended based on value-added	Recommended based on turnover	Actual hires
Female	−.02 (.03)	−.00 (.03)	−.01 (.03)	−.01 (.03)	−.01 (.03)	.06* (.03)
White	.02 (.04)	.02 (.04)	−.01 (.03)	.01 (.04)	−.09* (.04)	.11** (.04)
Age	.35*** (.02)	−.10*** (.02)	.26*** (.02)	.58*** (.02)	.59*** (.02)	.48*** (.02)
Age^2	−.13*** (.01)	−.05*** (.01)	−.11*** (.01)	−.15*** (.01)	−.08*** (.01)	−.13*** (.01)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Constant	−.82*** (.09)	−.96*** (.09)	−.83*** (.09)	−1.00*** (.09)	−1.05*** (.09)	−1.44*** (.09)
Observations	16,071	16,071	16,071	16,071	16,071	16,071

Note. Standard errors in parentheses, n = 16,071. Standard errors adjusted for seven clusters in application year. Controlled for application year and position type.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 9
Comparison Between Outcomes of the Recommended and Not-Recommended Groups Among Hires in the Hold-Out Sample

Select on . . .	Actual scores				
	Performance composite	Student evaluation	Expert observation	Value-added	Retention
Performance composite	.40 [.38, .43]	.10 [.07, .13]	.35 [.33, .37]	.31 [.27, .35]	.65 [.59, .70]
Student evaluation	.14 [.11, .17]	.26 [.22, .29]	.16 [.13, .19]	.03 [−.01, .07]	.23 [.16, .30]
Expert observation	.40 [.37, .42]	.12 [.09, .14]	.37 [.35, .40]	.30 [.26, .34]	.09 [.04, .15]
Value-added	.32 [.29, .34]	.08 [.06, .11]	.27 [.25, .29]	.22 [.18, .25]	.46 [.4, .51]
Retention	−.01 [−.04, .02]	−.08 [−.11, −.04]	−.13 [−.16, −.10]	.14 [.09, .19]	3.53 [3.49, 3.58]

Note. The columns show the mean difference between recommended and not-recommended among hires in the hold-out sample. Numbers in brackets show the 95% confidence interval around the average value over 200 iterations.

the shares of applicants to openings across years and positions, we simulated a process of hiring at random. Averaging across iterations for each model, both the school district's actual hires and our models' recommended hires agreed with random hires between 13% and 14%.

Overall, the results of the effectiveness evaluation confirm that in general our models outperform the district selection system (differences are generally positive and the confidence intervals mostly do not include zero). If our models were just picking up noise, the two groups would be the same in expectation. Instead, our models are successfully sorting hires into groups that differ significantly and substantially in observed performance.

Discussion

The need for tools that facilitate the rapid collection and analysis of application information is constantly growing. On one extreme, human resource practitioners have taken to very rapidly scanning through a large number of applications by keywords, relying on either a very small number of cues or heuristics for rejecting candidates. Alternatively, purely empirical algorithmic methods for using large data sets are calibrated exclusively to a specific applicant pool and selection context, yielding prediction models that are unlikely to generalize to future occasions, and poorly integrated with the substantial body of knowledge already present in the field of selection. The method evaluated in this study proposes a middle way, that combines machine learning techniques that are directed to find and analyze themes that correspond to established selection techniques. Through this process, we are able to score applicant quality using relevant work experiences, tenure history, and reasons for leaving previous jobs.

Conceptual Implications

One of the key advantages of using a grounded approach to machine learning is the possibility to systematically evaluate whether predictions will generalize and be useful in other contexts. The matching model for O*Net, based on a well-defined and research validated competency model, is relevant for a wide variety of jobs and industries. Because the content of O*Net is publicly available, and most organizations require employees to specify both the duration and title for their previous jobs, there would be little difficulty in using this approach to quantify relevant work experience for any given job. This approach is focused on a consistent and transparent measure of relevant task exposure, which is associated with greater acceptability and perceived job for organizational leaders, job applicants, and the legal system (e.g., Hausknecht et al., 2004; Klehe, 2004; Ryan & Ployhart, 2000).

It is difficult to ascertain how the distribution of applicants' past occupations compares with that of applicants in other positions, particularly those requiring at least a bachelor's degree. But, even if pathways into teaching include an unusually uniform set of education and licensing experiences, work-experience relevance may be more important as a differentiator in other fields. In jobs where there is more variety in previous types of experience (i.e., less restriction of range on the predictors), correlations with performance can potentially be greater. In sum, although the exact model we proposed here will likely differ from job to job and organization to organization, the general methodology of machine learning presented here can be used to quickly, efficiently, and rigorously produce interpretable evaluations of KSAO and motivation-related constructs from large numbers of applicants.

The conceptualization of work experience relevance developed through machine learning focuses on specific job tasks across different occupations rather than job titles by themselves. There is no question that the predictive power of work experience is maximized by using task details (Dokko et al., 2009; Tesluk & Jacobs, 1998). By pairing job titles with job analyst ratings of task requirements in O*NET, we are able to use verifiable information, rather than relying on the unique words applicants use to describe previous job tasks in a resume. The machine learning system also makes it possible to predict the likelihood of one having used KSAOs in a more refined and continuous manner than alternative methods of matching titles across fields. Moreover, while previous work has mostly used this information on experience for the prediction of job performance, ours is the first study of which we are aware that has used task-specific job experiences to also predict voluntary turnover. The link between relevant work experience and turnover suggests that theories of person-job match in the turnover and job attitudes literature can be integrated more fully with other work on job performance.

The results related to tenure history and turnover are largely in line with prior work related to the "hobo syndrome" (Judge & Watanabe, 1995; Munasinghe & Sigman, 2004). In addition to previous findings, we also found that a history of short tenure in past jobs can be linked to lower performance on the job. The reason for these linkages is not entirely clear based on our data, but it does raise intriguing questions that might be examined in the future. One possibility is that individuals who switch jobs often have short time horizons for their work and, therefore, do not invest time or energy to become proficient

(Rusbult & Farrell, 1983). In a sense, this is a rational response, mirroring models of organizational commitment that show investment of time and energy into an organization are proportional to the expected duration of the relationship. On the other hand, individuals with a history of short tenure might not be good employees for reasons not measured in our study or other prior research, and their poor performance fuels leaving jobs quickly in a somewhat futile search to find a better fit.

We found that those who left a previous job to avoid a bad job were worse performers and were more likely to turnover. This does mirror our expectations based on the demonstrated consistency of negative job attitudes across employers. We go further than this evidence of job attitude stability, because we show these stable tendencies can be related to downstream measures of performance and turnover in future jobs. Because employment motivation data are collected prior to the point at which performance on this job can even exist, the direction of the relationship lends itself to causal inference compared with prior work. Considering approach motivation as indicated by leaving prior jobs to seek a better job also shows that motivation may carry over from the job search process to employment. Individuals who leave a job because they wish to do something more personally meaningful are shown to be superior workers.

Practical Implications

Organizations more than ever have access to large amount of text data from job applicants including applicants' responses to the online application forms, their cover letters, and their resumes. Our study helps organizations utilize these data to improve work outcomes while lowering the risk of adverse impact. Also, our method can help job applicants and organizations alike by making the selection process more systematic by reducing the likelihood of recruiters' biases or applicants' influence tactics to deviate the selection process.

Relevant experience has many positive features in practice beyond verifiability. In particular, work experience is seen as highly relevant and acceptable for selection purposes by organizational leaders and job applicants (Hausknecht et al., 2004). One key standard for legal defensibility is the use of job analysis information in the selection procedure (Borden & Sharf, 2007)—using O*Net job characteristics linked to prior work history is perfectly matched to this legal requirement. There are also concerns regarding personality or integrity tests because most applicants have some sense of how to "fake good" on such measures (Birkeland, Manson, Kisamore, Brannick, & Smith, 2006), and many applicants believe these questions are not job relevant (Hausknecht et al., 2004). Evidence shows that the likelihood of applicants engaging in dishonest impression management tactics in the verifiable parts of their application form such as education or work history is lower compared with that in other parts of job application (Cole, Rubin, Feild, & Giles, 2007; Ployhart, 2012; Waung, McAuslan, & DiMambro, 2016).

In the area of education, our study offers several contributions. The existing literature supports the contention that improving the teacher selection process, especially in public schools in disadvantaged areas, can help improve the quality of education, which leads to narrowing the achievement gap (e.g., Aaronson, Barrow, & Sander, 2007; Adnot, Dee, Katz, & Wyckoff, 2016; Chetty, Friedman, & Rockoff, 2014; Hanushek, Kain, O'Brien, & Rivkin, 2005). The ability to hire individuals who will perform well and will also remain on the job over time has the same potential to improve performance in other organizational contexts (Barrick & Zimmerman, 2009). Schools spend about 80% of their budget on labor, but their hiring practices are ineffective and inconsistent. Schools hire essentially at random (Goldhaber et al., 2014), wait up to 3 years to act on the measures of effectiveness, and decide whether or not to dismiss ineffective teachers. This process subjects many children to years of ineffective teaching, and expends resources through frequent hiring and firing. Improved selection might reduce our need to learn about teacher performance on the backs of children (Staiger & Rockoff, 2010). Finally, most teachers in public schools are unionized, and decisions about their compensation, job design, or termination are mandatory subjects of collective bargaining. However, management has greater flexibility to innovate in the selection of potential employees than other HRM areas. Factors like work history are legally acceptable predictors of work outcomes, because work history is considered a legitimate job-related criterion by the Equal Employment Opportunity Commission (1978, Section 14, B.3; Barrick & Zimmerman, 2005). Even in nonunionized organizations, the perceived fairness of selection measures increases with greater transparency and consistency of selection tools. Our machine learning method that starts from clearly stated and easily understood selection measures excels in these domains.

Finally, improving the quality of teacher selection has a substantial impact on nation's economy, welfare, and human capital at relatively minimal cost. According to the Bureau of Labor Statistics, about 4 million teachers were engaged in classroom instruction in 2016. This number accounts for 3% of the U.S. workforce. Teachers also contribute to the quality of human capital by educating the future workforce. Evidence suggests that teaching that exceeds mean performance by one standard deviation increases students' success in adult life and produces, conservatively, over $200,000 in net present social value for each teacher per year (Chetty et al., 2014; Hanushek, 2011). The results in Table 9 suggest that hiring guided by this model might raise average hires' effectiveness by a fifth to a third of a standard deviation, adding $50,000 to $67,000 of net present value of student learning annually per hire. Our estimates in Table 9 quantify the tradeoffs in screening to maximize different outcomes. Gaining these benefits would require some fixed costs to integrate the screening model into an organization's workflow. However, the marginal costs of operation would be very low because the machine learning approach we outline here operates off routine application data, rather than requiring an ongoing use of additional staff or applicant time for surveys or interviews. Alternative policies to raise student achievement and teacher retention, such as class size reductions, are quite expensive in contrast (Chalfin et al., 2016).

Limitations and Future Directions

Our study has several limitations. We do not have actual demographic data for 37% of our sample and imputed the missing values. This increases the risk of error in assessing adverse impact. Second, this study only includes one public school district in the U.S. It would be helpful to expand to other contexts. Although our study may be generalizable to other workers such as nurses, doctors, social workers, or other service jobs similar to teachers for which aspects like approach motivation, interest, or specific individual characteristics are important, it would be informative to examine the predictive validity of the proposed variables in this study in jobs of different nature too. While these results are based on well-defined and broadly applicable concepts, there is a need for further validation of the method in other occupational and organizational contexts. Third, in this study we only show the direct relationships between the predictors and outcomes. Future studies can investigate underlying mechanisms. For example, we show that those who expressed that they left a previous job to seek a better job are more likely to be high performers and stay longer in the organization. Further study should attempt to explain why.

One of the main limitations of machine learning is that algorithms from biased training data will reproduce structural biases. If one population of individuals is systematically discriminated against in performance evaluations, the machine learning system will incorporate such bias, perpetuating discrimination. To mitigate the risk of these biases linked to task-irrelevant characteristics, our theory driven, job-relevant predictors derived from job analyses and motivation theories are more easily identified. However, we recognize that any algorithmic model can be prone to structural biases. For example, it is possible that certain groups of employees are involuntarily terminated due to managers' prejudice and discrimination. In this study we can only show evidence that, compared with the current selection system in the district, our model decreases the risk of adverse impact. A fruitful future direction would evaluate the risk of different types of biases in the predictors introduced in this study.

References

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics, 25,* 95–135. http://dx.doi.org/10.1086/508733

Adnot, M., Dee, T., Katz, V., & Wyckoff, J. (2016). Teacher turnover, teacher quality, and student achievement in DCPS. *Center for Education Policy Analysis* (Working Paper No. 16–03). http://dx.doi.org/10.3102/0162373716663646

Aguinis, H., & Smith, M. A. (2007). Understanding the impact of test validity and bias on selection errors and adverse impact in human resource selection. *Personnel Psychology, 60,* 165–199. http://dx.doi.org/10.1111/j.1744-6570.2007.00069.x

Aiolli, F., De Filippo, M., & Sperduti, A. (2009). Application of the preference learning model to a human resources selection task. *2009 IEEE Symposium on Computational Intelligence and Data Mining* (pp. 203–210). Nashville, TN: IEEE. http://dx.doi.org/10.1109/CIDM.2009.4938650

Bao, Y., & Datta, A. (2014). Simultaneously discovering and quantifying risk types from textual risk disclosures. *Management Science, 60,* 1371–1391. http://dx.doi.org/10.1287/mnsc.2014.1930

Barrick, M. R., Mount, M. K., & Strauss, J. P. (1994). Antecedents of involuntary turnover due to a reduction in force. *Personnel Psychology, 47,* 515–535. http://dx.doi.org/10.1111/j.1744-6570.1994.tb01735.x

Barrick, M. R., & Zimmerman, R. D. (2005). Reducing voluntary, avoidable turnover through selection. *Journal of Applied Psychology, 90,* 159–166. http://dx.doi.org/10.1037/0021-9010.90.1.159

Barrick, M. R., & Zimmerman, R. D. (2009). Hiring for retention and performance. *Human Resource Management, 48,* 183–206. http://dx.doi.org/10.1002/hrm.20275

Baum, J. R., & Locke, E. A. (2004). The relationship of entrepreneurial traits, skill, and motivation to subsequent venture growth. *Journal of Applied Psychology, 89,* 587–598. http://dx.doi.org/10.1037/0021-9010.89.4.587

Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment, 14,* 317–335. http://dx.doi.org/10.1111/j.1468-2389.2006.00354.x

Blatt, R., & Ashford, S. J. (2006). *Making meaning and taking action in knowledge and creative work: Lessons from independent workers*. Ann Arbor: University of Michigan Press.

Blume, B. D., Ford, J. K., Baldwin, T. T., & Huang, J. L. (2010). Transfer of training: A meta-analytic review. *Journal of Management, 36,* 1065–1105. http://dx.doi.org/10.1177/0149206309352880

Bonneville-Roussy, A., Lavigne, G., & Vallerand, R. J. (2011). When passion leads to excellence: The case of musicians. *Psychology of Music, 39,* 123–138. http://dx.doi.org/10.1177/0305735609352441

Borden, L. W., & Sharf, J. C. (2007). Developing legally defensible content valid selection procedures. In D. L. Whetzel & G. R. Wheaton (Eds.), *Applied measurement: Industrial psychology in human resources management* (pp. 385–401). New York, NY: Taylor and Francis.

Brown, B., & Campion, M. (1994). Biodata phenomenology: Recruiters' perceptions and use of biographical information in resume screening. *Journal of Applied Psychology, 79,* 897–908. http://dx.doi.org/10.1037/0021-9010.79.6.897

Burns, M. O., & Seligman, M. E. P. (1989). Explanatory style across the life span: Evidence for stability over 52 years. *Journal of Personality and Social Psychology, 56,* 471–477. http://dx.doi.org/10.1037/0022-3514.56.3.471

Campion, M. C., Campion, M. A., Campion, E. D., & Reider, M. H. (2016). Initial investigation into computer scoring of candidate essays for personnel selection. *Journal of Applied Psychology, 101,* 958–975. http://dx.doi.org/10.1037/apl0000108

Chalfin, A., Danieli, O., Hillis, A., Jelveh, Z., Luca, M., Ludwig, J., & Mullainathan, S. (2016). Productivity and selection of human capital with machine learning. *The American Economic Review, 106,* 124–127. http://dx.doi.org/10.1257/aer.p20161029

Chamorro-Premuzic, T., Winsborough, D., Sherman, R. A., & Hogan, R. (2016). New talent signals: Shiny new objects or a brave new world? *Industrial and Organizational Psychology: Perspectives on Science and Practice, 9,* 621–640. http://dx.doi.org/10.1017/iop.2016.6

Chetty, B. R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers: Evaluating bias in teacher value-added estimates. *American Economic Review, 104,* 2593–2632.

Cleves, M. A., Gould, W., & Marchenko, Y. V. (2016). *An introduction to survival analysis using Stata* (revised 3rd ed.). College Station, TX: Stata Press.

Clougherty, J. A., Duso, T., & Muck, J. (2016). Correcting for self-selection based endogeneity in management research. *Organizational Research Methods, 19,* 286–347. http://dx.doi.org/10.1177/1094428115619013

Cole, M., Rubin, R., Feild, H., & Giles, W. (2007). Recruiters' perceptions and use of applicant résumé information: Screening the recent graduate. *Applied Psychology, 56,* 319–343. http://dx.doi.org/10.1111/j.1464-0597.2007.00288.x

Converse, P., Oswald, F., Gillespie, M., Field, K. A., & Bizot, E. B. (2004). Matching individuals to occupations using abilities and the O* NET: Issues and an application in career guidance. *Personnel Psychology, 57,* 451–487. http://dx.doi.org/10.1111/j.1744-6570.2004.tb02497.x

Danielson, C. (2007). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: Association for Supervision & Curriculum Development.

Davis, P. R., Trevor, C. O., & Feng, J. (2015). Creating a more quit-friendly national workforce? Individual layoff history and voluntary turnover. *Journal of Applied Psychology, 100,* 1434–1455.

de Montjoye, Y. -A., Quoidbach, J., Robic, F., & Pentland, A. (2013).

Predicting personality using novel mobile phone-based metrics. In A. M. Greenberg, W. G. Kennedy, & N. D. Bos (Eds.), *Social computing, behavioral-cultural modeling and prediction: SBP 2013, Lecture notes in computer science* (Vol. 7812). Berlin, Heidelberg: Springer. http://dx.doi.org/10.1007/978-3-642-37210-0_6

Dickter, D. N., Roznowski, M., & Harrison, D. A. (1996). Temporal tempering: An event history analysis of the process of voluntary turnover. *Journal of Applied Psychology, 81,* 705–716. http://dx.doi.org/10.1037/0021-9010.81.6.705

Diefendorff, J. M., & Mehta, K. (2007). The relations of motivational traits with workplace deviance. *Journal of Applied Psychology, 92,* 967–977. http://dx.doi.org/10.1037/0021-9010.92.4.967

Dokko, G., Wilk, S. L., & Rothbard, N. P. (2009). Unpacking prior experience: How career history affects job performance. *Organization Science, 20,* 51–68. http://dx.doi.org/10.1287/orsc.1080.0357

Doyle, G., Goldberg, A., Srivastava, S., & Frank, M. (2017). Alignment at work: Using language to distinguish the internalization and self-regulation components of cultural fit in organizations. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL)*. Retrieved from https://www.gsb.stanford.edu/faculty-research/publications/alignment-work-using-language-distinguish-internalization-self

Edwards, J. R., & Harrison, R. V. (1993). Job demands and worker health: Three-dimensional reexamination of the relationship between person-environment fit and strain. *Journal of Applied Psychology, 78,* 628–648. http://dx.doi.org/10.1037/0021-9010.78.4.628

Elliot, A. J., & Harackiewicz, J. M. (1996). Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis. *Journal of Personality and Social Psychology, 70,* 461–475. http://dx.doi.org/10.1037/0022-3514.70.3.461

Elliot, A. J., & Sheldon, K. M. (1997). Avoidance achievement motivation: A personal goals analysis. *Journal of Personality and Social Psychology, 73,* 171–185. http://dx.doi.org/10.1037/0022-3514.73.1.171

Elliot, A. J., & Thrash, T. M. (2002). Approach-avoidance motivation in personality: Approach and avoidance temperaments and goals. *Journal of Personality and Social Psychology, 82,* 804–818. http://dx.doi.org/10.1037/0022-3514.82.5.804

Elliot, A. J., & Thrash, T. M. (2010). Approach and avoidance temperament as basic dimensions of personality. *Journal of Personality, 78,* 865–906. http://dx.doi.org/10.1111/j.1467-6494.2010.00636.x

Equal Employment Opportunity Commission. (1978). Adoption by four agencies of uniform guidelines on employee selection procedures. Federal Register, *43,* 38290–38315.

Feldman, R., & Sanger, J. (2007). *The text mining handbook: Advanced approaches in analyzing unstructured data*. Cambridge, UK: Cambridge University Press.

Ferris, D. L., Rosen, C. R., Johnson, R. E., Brown, D. J., Risavy, S. D., & Heller, D. (2011). Approach or avoidance (or both?): Integrating core self-evaluations within an approach/avoidance framework. *Personnel Psychology, 64,* 137–161. http://dx.doi.org/10.1111/j.1744-6570.2010.01204.x

Flandez, R. (2009, May 11). Resume overload: New software helps small companies keep up with swarm of applicants. *The Wall Street Journal*.

Förster, J., Higgins, E. T., & Idson, L. C. (1998). Approach and avoidance strength during goal attainment: Regulatory focus and the "goal looms larger" effect. *Journal of Personality and Social Psychology, 75,* 1115–1131. http://dx.doi.org/10.1037/0022-3514.75.5.1115

Gale, S. F. (2017). RPO customers: We want more of everything. *Workforce, January–February,* 44–45.

Goldhaber, D., Grout, C., & Huntington-Klein, N. (2014). Screen twice, cut once: Assessing the predictive validity of teacher selection tools. *Education Finance and Policy, 12,* 197–223. http://dx.doi.org/10.1162/EDFP_a_00200

Grambsch, P. M., & Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika, 81,* 515–526. http://dx.doi.org/10.1093/biomet/81.3.515

Grensing-Pophal, L. (2017). *Recruiting: Why traditional methods still matter*. Retrieved from www.shrm.org

Griffeth, R., Hom, P. P. W., & Gaertner, S. (2000). A Meta-analysis of Antecedents and Correlates of Employee Turnover: Update, Moderator Tests, and Research Implications for the Next Millennium. *Journal of Management, 26,* 463–488.

Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of Education Review, 30,* 466–479. http://dx.doi.org/10.1016/j.econedurev.2010.12.006

Hanushek, E. A., Kain, J. F., O'Brien, D. M., & Rivkin, S. G. (2005). The market for teacher quality. *NBER Working Papers, 94305,* 1–52.

Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology, 57,* 639–683. http://dx.doi.org/10.1111/j.1744-6570.2004.00003.x

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica, 47,* 153–161. http://dx.doi.org/10.2307/1912352

Johnson, J. W., Steel, P., Scherbaum, C. A., Hoffman, C. C., Jeanneret, P. R., & Foster, J. (2010). Validation is like motor oil: Synthetic is better. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 3,* 305–328. http://dx.doi.org/10.1017/S1754942600002479

Jovanovic, B. (1984). Matching, turnover, and unemployment. *Journal of Political Economy, 92,* 108–122. http://dx.doi.org/10.1086/261210

Judge, T. A., & Watanabe, S. (1995). Is the past prologue?: A test of Ghiselli's hobo syndrome. *Journal of Management, 21,* 211–229. http://dx.doi.org/10.1177/014920639502100203

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2012). Have we identified effective teachers? Validating measures of effective teaching using random assignment. *Bill & Melinda Gates Foundation Technical Report*.

Klehe, U. C. (2004). Choosing how to choose: Institutional pressures affecting the adoption of personnel selection procedures. *International Journal of Selection and Assessment, 12,* 327–342. http://dx.doi.org/10.1111/j.0965-075X.2004.00288.x

Lanaj, K., Chang, C. H., & Johnson, R. E. (2012). Regulatory focus and work-related outcomes: A review and meta-analysis. *Psychological Bulletin, 138,* 998–1034. http://dx.doi.org/10.1037/a0027723

Lee, F., & Peterson, C. (1997). Content analysis of archival data. *Journal of Consulting and Clinical Psychology, 65,* 959–969. http://dx.doi.org/10.1037/0022-006X.65.6.959

Lee, T. W., Mitchell, T. R., Holtom, B. C., McDaneil, L. S., & Hill, J. W. (1999). The unfolding model of voluntary turnover: A replication and extension. *Academy of Management Journal, 42,* 450–462.

Maertz, C. P., & Campion, M. A. (2004). Profiles in quitting: Integrating process and content turnover theory. *Academy of Management Journal, 47,* 566–582.

Maier, S. F., & Seligman, M. E. P. (2016). Learned helplessness at fifty: Insights from neuroscience. *Psychological Review, 123,* 349–367. http://dx.doi.org/10.1037/rev0000033

Maner, J. K., & Gerend, M. A. (2007). Motivationally selective risk judgments: Do fear and curiosity boost the boons or the banes? *Organizational Behavior and Human Decision Processes, 103,* 256–267. http://dx.doi.org/10.1016/j.obhdp.2006.08.002

Manning, C., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.

Mathieu, J., & Zajac, D. (1990). A review and meta-analysis of the antecedents, correlates, and consequences of organizational commitment. *Psychological Bulletin, 108,* 171–194. http://dx.doi.org/10.1037/0033-2909.108.2.171

McKenny, A. F., Aguinis, H., Short, J. C., & Anglin, A. H. (2018). What doesn't get measured does exist: Improving the accuracy of computeraided text analysis. *Journal of Management, 44,* 2909–2933. http://dx.doi.org/10.1177/0149206316657594

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2012). *Foundations of machine learning*. Cambridge, MA: The MIT Press.

Munasinghe, L., & Sigman, K. (2004). A hobo syndrome? Mobility, wages, and job turnover. *Labour Economics, 11,* 191–218. http://dx.doi.org/10.1016/j.labeco.2003.05.001

Newton, T., & Keenan, T. (1991). Further analyses of the dispositional argument in organizational behavior. *Journal of Applied Psychology, 76,* 781–787. http://dx.doi.org/10.1037/0021-9010.76.6.781

O*NET. (n.d.). Retrieved from onetonline.org/help/onet

Owens, W. A., & Schoenfeldt, L. F. (1979). Toward a classification of persons. *Journal of Applied Psychology, 64,* 569–607. http://dx.doi.org/10.1037/0021-9010.64.5.569

Papay, J. P., & Kraft, M. A. (2015). Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics, 130,* 105–119. http://dx.doi.org/10.1016/j.jpubeco.2015.02.008

Peres, S., & Garcia, J. R. (1962). Validity and dimensions of descriptive adjectives used in reference letters for engineering applicants. *Personnel Psychology, 15,* 279–286. http://dx.doi.org/10.1111/j.1744-6570.1962.tb01622.x

Peterson, N. G., Mumford, M. D., Borman, W. C., Jeanneret, P. R., & Fleishman, E. A. (Eds.). (1999). *An occupational information system for the 21st century: The development of O*NET*. Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/10313-000

Ployhart, R. E. (2012). Personnel selection: Ensuring sustainable organizational effectiveness through the acquisition of human capita. In S. W. J. Kozlowski (Ed.), *The Oxford handbook of organizational psychology* (pp. 221–246). Oxford, UK: Oxford University Press. http://dx.doi.org/10.1093/oxfordhb/9780199928309.013.0008

Quiñones, M., Ford, J., & Teachout, M. (1995). The relationship between work experience and job performance: A conceptual and meta-analytic review. *Personnel Psychology, 48,* 887–910. http://dx.doi.org/10.1111/j.1744-6570.1995.tb01785.x

Rusbult, C. E., & Farrell, D. (1983). A longitudinal test of the investment model: The impact on job satisfaction, job commitment, and turnover of variations in rewards, costs, alternatives, and investments. *Journal of Applied Psychology, 68,* 429–438. http://dx.doi.org/10.1037/0021-9010.68.3.429

Ryan, A. M., & Ployhart, R. E. (2000). Applicants' perceptions of selection procedures and decisions: A critical review and agenda for the future. *Journal of Management, 26,* 565–606. http://dx.doi.org/10.1177/014920630002600308

Ryan, A. M., & Ployhart, R. E. (2014). A century of selection. *Annual Review of Psychology, 65,* 693–717. http://dx.doi.org/10.1146/annurev-psych-010213-115134

Ryan, A. M., Sacco, J. M., McFarland, L. A., & Kriska, S. D. (2000). Applicant self-selection: Correlates of withdrawal from a multiple hurdle process. *Journal of Applied Psychology, 85,* 163–179. http://dx.doi.org/10.1037/0021-9010.85.2.163

Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology, 85,* 112–118. http://dx.doi.org/10.1037/0021-9010.85.1.112

Saks, A. M., & Belcourt, M. (2006). An investigation of training activities and transfer of training in organizations. *Human Resource Management, 45,* 629–648. http://dx.doi.org/10.1002/hrm.20135

Schleicher, D. J., Hansen, S. D., & Fox, K. E. (2011). Job attitudes and work values. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology* (Vol. 3, pp. 137–189). Washington, DC: American Psychological Association.

Singer, J., & Willett, J. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195152968.001.0001

Spangler, W. D. (1992). Validity of questionnaire and TAT measures of need for achievement: Two meta-analyses. *Psychological Bulletin, 112,* 140–154. http://dx.doi.org/10.1037/0033-2909.112.1.140

Staiger, D. O., & Rockoff, J. E. (2010). Searching for effective teachers with imperfect information. *The Journal of Economic Perspectives, 24,* 97–118. http://dx.doi.org/10.1257/jep.24.3.97

Staw, B., McKechnie, P., & Puffer, S. (1983). The justification of organizational performance. *Administrative Science Quarterly, 28,* 582–600. http://dx.doi.org/10.2307/2393010

Staw, B., & Ross, J. (1985). Stability in the midst of change. A dispositional approach to job attitudes. *Journal of Applied Psychology, 70,* 469–480. http://dx.doi.org/10.1037/0021-9010.70.3.469

Steel, P., & Kammeyer-Mueller, J. D. (2009). Using a meta-analytic perspective to enhance job component validation. *Personnel Psychology, 62,* 533–552. http://dx.doi.org/10.1111/j.1744-6570.2009.01147.x

Tesluk, P., & Jacobs, R. (1998). Toward an integrated model of work experience. *Personnel Psychology, 51,* 321–355. http://dx.doi.org/10.1111/j.1744-6570.1998.tb00728.x

Trevor, C. O. (2001). Interactive effects among actual ease of movement determinants and job satisfaction in the prediction of voluntary turnover. *Academy of Management Journal, 44,* 621–638.

Tsai, W.-C., Huang, T.-C., Wu, C.-Y., & Lo, I.-H. (2010). Disentangling the effects of applicant defensive impression management tactics in job interviews. *International Journal of Selection and Assessment, 18,* 131–140. http://dx.doi.org/10.1111/j.1468-2389.2010.00495.x

United States Department of Labor, Bureau of Labor Statistics. (n.d.). Retrieved October 4, 2017, from https://www.bls.gov/news.release/tenure.t06.htm

Vallerand, R., Mageau, G., Elliot, A., Dumais, A., Demers, M-A., & Rousseau, F. (2008). Passion and performance attainment in sport. *Psychology of Sport and Exercise, 9,* 373–392. http://dx.doi.org/10.1016/j.psychsport.2007.05.003

Value-Added Research Center & Minneapolis Public Schools. (2013). *Minneapolis value-added model* (Technical report). Madison: Wisconsin Center for Education Research at University of Wisconsin-Madison.

Wang, Q., Bowling, N. A., & Eschleman, K. J. (2010). A meta-analytic examination of work and general locus of control. *Journal of Applied Psychology, 95,* 761–768. http://dx.doi.org/10.1037/a0017707

Waung, M., McAuslan, P., & DiMambro, J. (2016). Impression management use in résumés and cover letters. *Journal of Business and Psychology, 32,* 1–20.

Wernimont, P. F., & Campbell, J. P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology, 52,* 372–376. http://dx.doi.org/10.1037/h0026244

Wilk, S. L., Desmarais, L. B. L., & Sackett, P. (1995). Gravitation to jobs commensurate with ability: Longitudinal and cross-sectional tests. *Journal of Applied Psychology, 80,* 79–85. http://dx.doi.org/10.1037/0021-9010.80.1.79

Wiswall, M. (2013). The dynamics of teacher quality. *Journal of Public Economics, 100,* 61–78. http://dx.doi.org/10.1016/j.jpubeco.2013.01.006

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.

Wrzesniewski, A., Dutton, J. E., & Debebe, G. (2003). Interpersonal sensemaking and the meaning of work. *Research in Organizational Behavior, 25,* 93–135. http://dx.doi.org/10.1016/S0191-3085(03)25003-6

(*Appendix follows*)

Appendix

Technical Details

Naive Bayes Classification

In this document classification method, we first convert each document (self-reported job description) to a feature vector, $\mathbf{d} = (w1, w2, \ldots)$, so that each meaningful word is represented in a column by the number of times each word occurs in the document. This representation is called the "*bag of words*" representation in which the order of the words is not represented (Feldman & Sanger, 2007). For instance, assume we have the following two documents in our data, reflecting part of an applicant's job responsibilities:

1. Work with schools to improve their diversity practices.
2. Developed a diversity initiative in the district.

The document-term matrix that represents these documents would be as below. The rows reflect each of the two documents.

Work Schools Improve Their Diversity Practices Developed Initiative District

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}$$

Note that the algorithm ignores the common words, or "stop words," such as "to," "the," or "with."

In the naive Bayes approach, we define the probability that the document d belongs to class c using Bayes theorem as follows:

$$P(c|d) = \frac{P(d|c)P(c)}{P(d)} \tag{4}$$

We need to choose a priori bag of words that gives information regarding each class based on what we have in the training set (Manning & Schütze, 1999). In this study, we use O*NET standardized job descriptions and job titles as the training set in classifying self-reported job title and descriptions into O*NET standardized occupations. We use a manually trained data set for the reasons for leaving classification.

The marginal probability $P(d)$ is constant for all classes and can be dropped. The assumption of naive Bayes method to calculate $P(d|c)$ is that all features in the document vector $\mathbf{d} = (w_1, w_2, \ldots, w_n)$ are independent:

$$\mathrm{P(d|c)} = \prod_i P(w_i|c) \tag{5}$$

So, the classifier function would be:

$$P(c_i|d) = \prod_i P(w_i|c_i)P(c_i) \tag{6}$$

Using the a priori class information in the training set, the *Bayes' classifier* chooses the class with the highest posterior probability; that is, it assigns class C_m *to* a document if

$$\mathrm{P(C_m|\mathbf{d})} = \max_i P(c_i|\boldsymbol{d}) \tag{7}$$

Received April 2, 2018
Revision received January 26, 2019
Accepted February 9, 2019 ■