The Science behind Seedlink

14 July 2019

Abstract

The aim of this paper is to explain how Seedlink can help organizations improve their employee selection process by tackling challenges in human bias and siloed data. Seedlink proposes a method in which candidates are assessed for a value fit or specific job function through their language. More specifically, a predictive model computes scores for required key traits and competencies from the candidate's answers to open-ended questions. Reliability and bias of the predictive models are studied on a large number of different client cases. Reliability is evaluated by comparing it (1) to client feedback of candidates, and (2) by comparing it to post-hire performance of candidates. Bias of the predictive models is evaluated by visualizing the score distributions of candidates with different gender, major, or native language. The results show that there is a clear relation between Seedlink score and client feedback, i.e. candidates with higher Seedlink scores perform better in structured interviews and progress to later interview rounds than candidates with lower Seedlink scores. Furthermore, gender, major, and native language were shown to have no significant or very moderate impact on Seedlink scores. Seedlink assessments remove human bias and show reliable performance across different languages, industries and job functions.

1. Challenges in employee selection

The difficulty of predicting fit and success

The employee selection process is one of the most important business processes in organizations, but it is a process with a very high failure rate. Across different job levels approximately 50 percent of new hires fail within eighteen months (Sullivan, 2017). Harvard Business Review reports that nearly half of the leaders hired from outside fail within the first eighteen months (Martin, 2014).

Retention curves of new employees in New Zealand across different age groups are shown in Figure 1. Eighteen months after starting with an employer only 30 to 40 percent of new employees are still employed. The retention rates in New Zealand are not a special case; across different industries and countries, retention rates of new employees after eighteen months of 40 to 50 percent are considered typical. Employee attrition is a complex and multi-faceted problem but could partly be explained by the lack of person-organization fit (Vancouver and Schmitt, 1991).

The main reasons for high failure rates and lack of person-organization fit of employees are related to poorly designed selection processes that are based on past practices or intuition rather than on data-driven or science-based insights (Sullivan, 2017). In most organizations the change towards a more data-driven selection process is not straightforward but, if successful, can be very rewarding since each failed hire is associated with significant costs in terms of negative business impact and lost productivity.

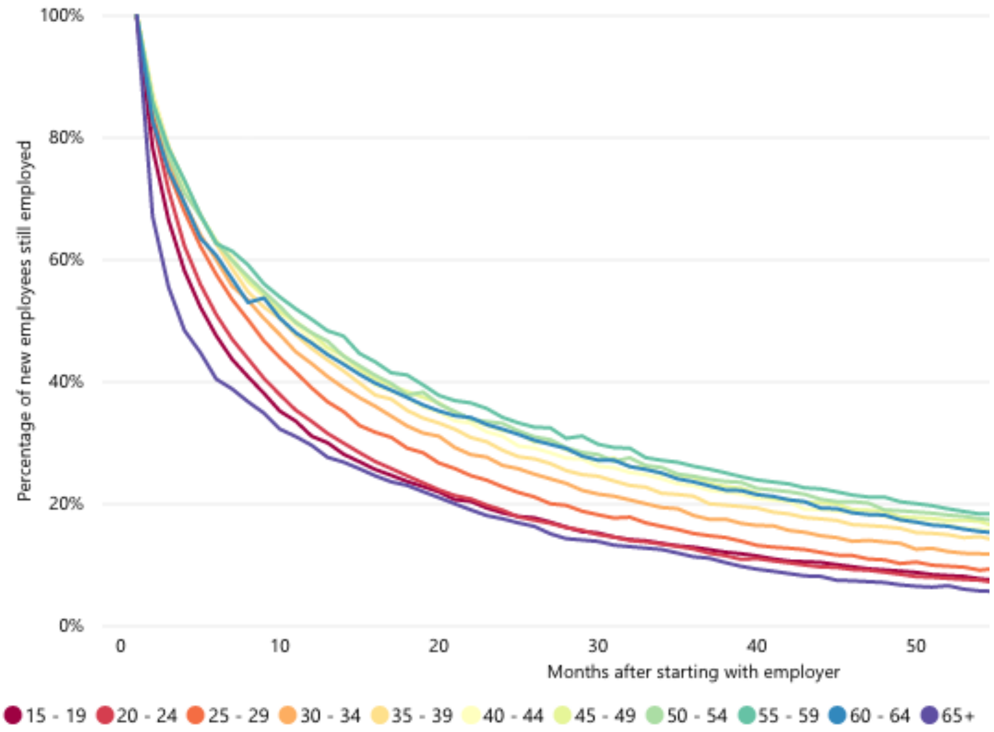


Figure 1. Employee retention across different age groups in New Zealand. Image courtesy of "http://www.sweetanalytics.co.nz/2-general/39-employee-retention-by-age".

The challenge of human bias

An important challenge to overcome in revising selection processes is the reduction of human bias. Based on past experiences and intuition, HR practitioners are tempted to generate preferred profiles for candidates in terms of gender, age, personality traits, competencies, educational background, work experience or nationality. Throughout the selection process such profiling is wittingly and unwittingly used, without evaluating the predictive validity of such preferred profiles for value fit or job performance. The challenge with human bias is to become aware of it and take measures to reduce it (Kandola, 2009).

The challenge of siloed data

A data-driven approach to generate preferred profiles for new hires can be achieved by leveraging the data of the current (and past) employees in an organization. If HR practitioners are able to identify the set of traits and competencies that make the current employees successful (or unsuccessful) in their job or organization, this information may directly be leveraged to improve the employee selection processes. The challenge here is to continuously establish links between the data of employees and candidates.

Relying on individual HR practitioners to continuously monitor employee performance and using this data to generate preferred hiring profiles may work in small organizations, but such attempts are not scalable to large organizations with thousands of employees and many different job functions. In large organizations, a more systematic approach is required to make full use of all available data, for example a software solution that self-learns the preferred profile for a specific job function from the current employees in that role.

2. Methodology

Overcoming challenges in employee selection

Seedlink has developed technology that helps organizations improve their selection process by overcoming the challenges of human bias and siloed data. The technology relies heavily on the consensus that personality traits and competencies are important predictors for job performance (e.g. Barrick and Mount, 1991; Hurtz and Donovan, 2000; Jackson and Rothstein 1991), and person-organization fit (e.g. O'Reilly III et al., 1991; Kristof-Brown et al., 2005; and Gardner, et al. 2012).

Traditionally, psychometric personality questionnaires would be administered to candidates to evaluate their personality traits. Obviously, when a job is at stake, the responses to such questionnaires may not be entirely truthful, because responses may reflect a candidate's perception of the ideal candidate rather than themselves (Furnham, 1990).

To make personality assessments more robust to dishonest answers, Seedlink's technology aims to predict these personality traits and competencies from answers to open-ended questions. Recent advances in natural language processing and machine learning have enabled efficient and reliable estimation of relevant traits and competencies from the subconscious use of linguistic markers in texts.

Personality trait prediction from text

Pennebaker and King (1999) were the first to investigate correlations between frequencies of word categories (e.g. positive emotion words, negative emotion words, pronouns) and personality traits. Using multiple writing samples of several hundred college students they found modest correlations to self-reports of Big Five personality dimensions. Their approach of using word categories to analyze texts became known as Linguistic Inquiry and Word Count (LIWC) and has since become a popular approach to study associations between personality and language use in different contexts, including directed writing assignments (Hirsh and Peterson, 2009), recording of day-to-day speech (Mehl et al., 2006), structured interviews (Fast and Funder, 2008), and online blogs (Yarkoni, 2010).

The technique of word categories provides insight into associations between personality and language use, but the reported correlations are typically too low to reliably infer author personality from text. Nowson and Oberlander (2006) found that using *n*-grams (sequences of *n* items typically used to capture word collocations) resulted in more accurate predictions of personality and gender from online blogs than LIWC. Schwartz et al. (2013) showed that an open-vocabulary approach on a large corpus containing 700 million words, phrases, and topic instances collected from Facebook messages of 75,000 volunteers, provided insights and accuracies that could not be obtained with closed-vocabulary word-category analyses such as LIWC.

Apart from predictive ability, another consideration when training language-based predictive models is their susceptibility to deception. It may be undesirable if an introvert could be classified as an extravert by deliberately using words that are mainly used by extraverts, e.g. *party* and *beach* (Schwartz et al., 2013). An approach to mitigate such straightforward deception attempts is by focusing on *how* someone writes, rather than *what* he writes. This method is known as computational stylometry and involves feature types such as simple character *n*-grams, punctuation, token *n*-grams, semantic and syntactic class distributions and patterns, parse trees, complexity and vocabulary richness measures, and even discourse features (Daelemans, 2013). Stylometric features have been used to predict personality traits from student essays (Luyckx and Daelemans, 2008), transcribed video blogs (Verhoeven and Daelemans, 2014), and twitter messages (Verhoeven et al., 2016).

More recently, deep learning techniques have enabled computers to efficiently learn semantic vector representation of words, sentences and paragraphs from large corpora (Mikolov et al., 2013; Pennington et al., 2014, and Le and Mikolov, 2014). By representing words, sentences or paragraphs as (sequences of) dense *N*-dimensional vectors, significant performance gains have been reported in various natural language processing problems including sentiment classification, machine translation, and question-answer systems (Young et al., 2018).

Not surprisingly, Majumder et al. (2017) report that a neural network using these word-level vector embeddings outperforms traditional approaches (e.g. *n*-grams, closed-vocabulary and open-vocabulary approaches) in terms of accuracy for Big Five personality traits. IBM personality insights, a commercial service to extract personality characteristics from text, is no longer using a LIWC-based model for predictions but is currently using a machine learning algorithm operating on word-level vector embeddings (IBM Personality Insights, 2019).

Seedlink's technology also exploits recent deep learning techniques to infer personality traits from natural language. Using a proprietary unsupervised learning technique on a large answer corpus, we learn dense *N*-dimensional vector embeddings that capture the stylistic as well as semantic characteristics of answers to open-ended questions. In turn, these vector embeddings in combination with supervised machine learning techniques enable accurate personality trait prediction models to be learnt from relatively small training datasets.

Building predictive models

In Figure 2 an overview is presented of the different steps in building an initial predictive model. In the first step, the client selects the key traits and competencies required in the organization or for the specific job function and selects a representative sample of at least 50 employees. This sample should include top performers, average performers, and low performers. Managers, HR professional, peers, and even customers can rate these employees on key traits and competencies according to their *perceptions*. The employees in the sample are also invited to respond to a small number of open-ended questions.

After all the trait and competency scores, as well as the language from the sample employees have been collected, an unsupervised machine learning algorithm preprocesses the answers to the open-ended questions and converts each of them to a dense *N*-dimensional vector. These vector representations capture semantic properties as well as consciously and subconsciously used stylistic characteristics of an answer to make the system more robust to deception attempts. Next, the vector representations of the answers and the human scores for competencies and behaviors are used as inputs by a supervised machine learning algorithm to generate a predictive model.

Seedlink uses *k*-fold cross-validation to validate the model internally, but in the validation step the client also has the opportunity to validate the model on an additional sample of employees not previously seen by Seedlink. Data from the validation phase are used to update the model, and once this step is completed the model is production-ready.

During deployment of the model, candidates are invited to answer the same open-ended questions as answered by the sample of employees. Based on the predictions for each of the key traits and competencies, a Seedlink score is computed that expresses the degree of fit for a specific role in the range from 0 to 100. Model development does not stop after the initial model has been deployed. By periodically adding language data and trait scores of new hires to the employee sample, the predictive model is able to increase its accuracy over time.

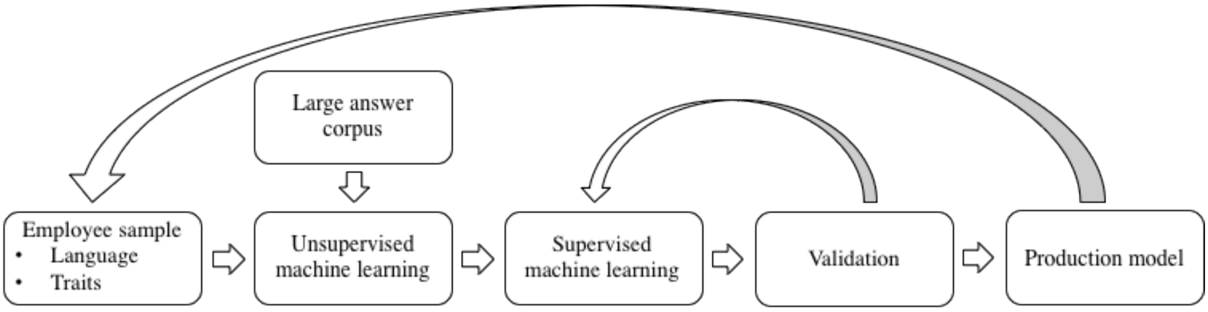


Figure 2. Overview of Seedlink methodology for custom-built predictive models

3. Reliability

The reliability of Seedlink technology is evaluated in two different ways; (1) by comparing it to client feedback of candidates, and (2) by comparing it to post-hire performance and retention.

Client feedback of candidates

If the client assesses candidates with reliable instruments (e.g. structured interviews), it is expected that candidates with higher Seedlink scores receive more favorable scores and reach later phases of the interview process. This implies that the mean Seedlink score increases, and the standard deviation decreases, for later phases in the interview progress. Below, reliability is evaluated for one case with blind structured interviews and three cases of interview progression in different clients.

Case: Blind structured interviews in large FMCG client

In 2019, a large FMCG client hired a renowned consultancy firm to review their global use of AI technology, including the Seedlink software used in their employee selection processes. In one of their experiments they evaluated a predictive model based on French responses to open-ended questions by conducting blind structured interviews. In detail, four candidates with high Seedlink scores (i.e. higher than 70) and four candidates with low Seedlink scores (i.e. lower than 65) were randomly selected from a large pool of applicants to complete a structured interview in an assessment center. The structured interview consisted of three exercises related to product presentation, digital projects, and creativity. Two recruiters evaluated each of the eight candidates on the same set of predetermined criteria, and both recruiters did not know the candidate CV and Seedlink score.

The results of this experiment are shown in Table 1, in which the names are fictious to protect the privacy of the individuals. The two candidates with the highest Seedlink scores (Alice and Bob) received 'Go'-evaluations, the two candidates with the next highest Seedlink scores (Caroline and Daniel) received 'Medium'-evaluations, and the four candidates with the lowest Seedlink scores all received 'No go'-evaluations.

Table 1. Blind assessments of eight candidates by two recruiters. The names have been changed to protect the privacy of the individuals.

Candidates	Seedlink score	Recruiters evaluation		Projection
		Isabelle	Jasper	
Alice	80	4/5	4/5	Go
Bob	80	4/5	3/5	Go
Caroline	77	3/5	2/5	Medium
Daniel	71	2/5	2/5	Medium
Edward	64	2/5	1/5	No go
Francois	60	1/5	1/5	No go
Gerald	59	2/5	2/5	No go
Helena	58	2/5	1/5	No go

Case: Shop management traineeship in clothing and accessories retailer

In 2017, Seedlink assisted an American worldwide clothing and accessories retailer with recruitment for shop management traineeships. More than 8000 candidates applied and were scored using a Seedlink predictive model based on Chinese responses to open-ended questions.

The interview progression of these candidates is shown in Table 2 and in Figure 3, i.e. 1029 candidates progressed to the first structured interview round, of whom 292 progressed to the second unstructured interview round, and of whom 54 received an offer. Note that candidates who progressed further in the recruitment process, had on average higher Seedlink scores. Furthermore, the standard deviation of the Seedlink score decreased in later interview stages.

Table 2 Averages and standard deviations of Seedlink scores from candidates in different phases of the interview.

Stage	N	Mean	S.d.
All candidates	8394	58.0	13.0
First unstructured interview	1029	63.8	11.4
Second unstructured interview	292	64.9	10.0
Offer	54	67.1	9.2

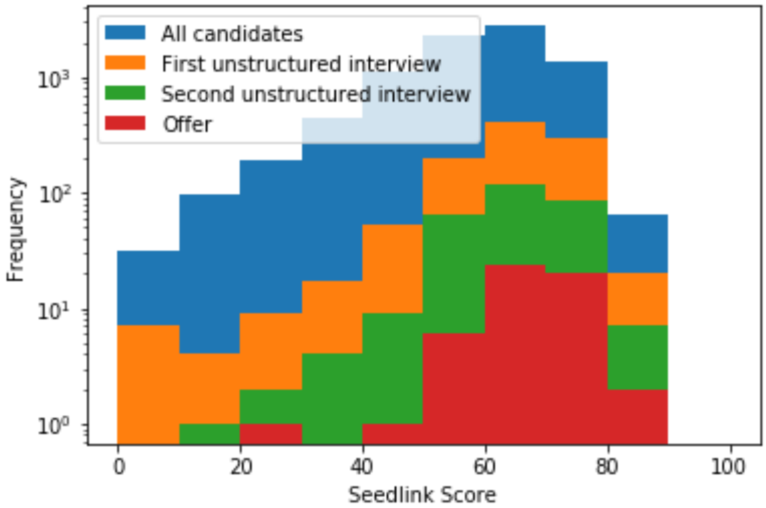


Figure 3. Histogram of interview progression for shop management traineeships. Note that the y-axis uses a logarithmic scale for visualization purposes.

Case: Management traineeship in food-products corporation

In 2017, a French food-products corporation used a Seedlink model for recruiting young professionals for management traineeships in China. More than 8000 candidates responded in Chinese to open-ended questions, and their answers were scored using a Seedlink model.

The interview progression of these candidates is shown in Table 3 and in Figure 4, i.e. 1302 candidates progressed to the first structured interview round, of whom 518 passed the phone interview, of whom 143 passed an unstructured interview, of whom 32 were invited to an assessment center, and of whom 17 received an offer. It can be seen that the mean Seedlink score gradually increases for candidates reaching later interview stages. Furthermore, the standard deviation of the Seedlink score decreases from 11.4 for all candidates to 7.9 for candidates who received an offer.

Table 3. Averages and standard deviations of Seedlink scores from candidates in different phases of the interview.

Stage	N	Mean	S.d.
All candidates	8208	51.0	11.4
Passed first selection	1302	56.9	9.6
Passed phone interview	518	57.2	9.9
Passed unstructured interview	143	58.2	10.5
Assessment center	32	58.7	9.2
Hiring decision	17	59.4	7.9

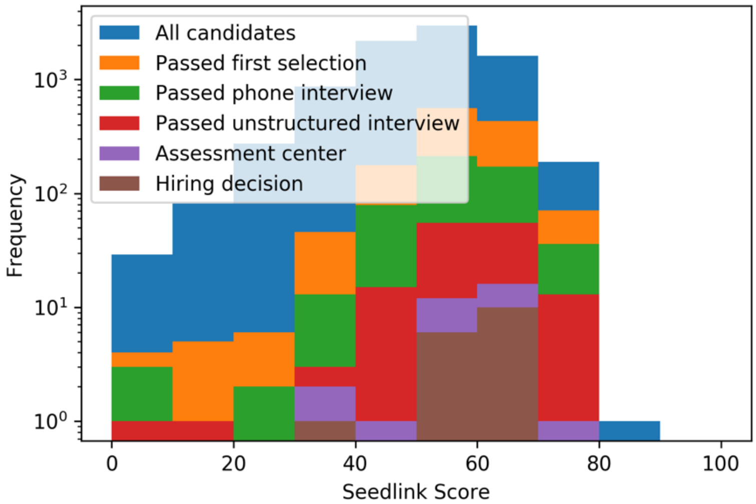


Figure 4. Interview progression for management traineeship. Note that the y-axis uses a logarithmic scale for visualization purposes.

Case: Campus recruitment for industrial gases company

In 2017, Seedlink assisted an American industrial gases company with their campus recruitment process in China. More than 1300 candidates responded in Chinese to open-ended questions, and their answers were scored using a Seedlink model.

The interview progression of these candidates is shown in Table 4 and in Figure 5, i.e. 49 candidates passed the third round interview of whom 21 passed the fourth round interview, of whom 9 received an offer. It can be seen that the mean Seedlink score gradually increases for candidates reaching later interview stages. The standard deviation of the Seedlink score decreases from 11.8 for all candidates to 7.0 for candidates who received an offer.

Table 4. Averages and standard deviations of Seedlink scores from candidates in different phases of the interview.

Stage	N	Mean	S.d.
All candidates	1313	62.6	11.8
Passed third round interview	49	65.4	8.6
Passed fourth round interview	21	66.8	8.0
Hiring decision	9	67.2	7.0

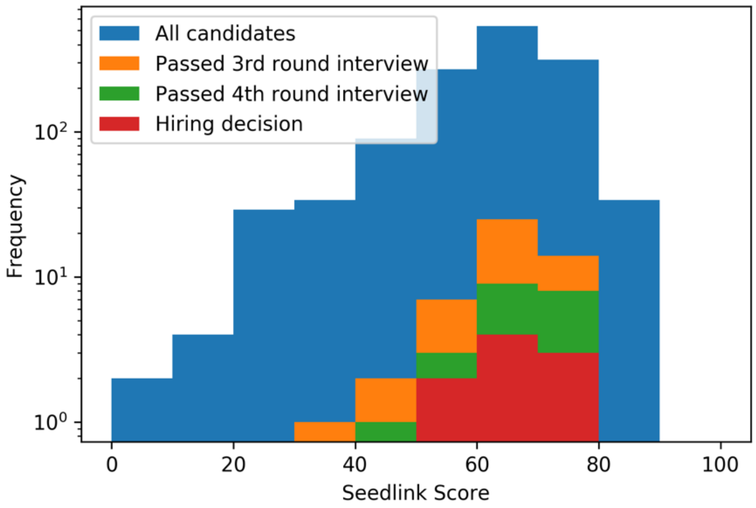


Figure 5. Interview progression for campus recruitment Note that the y-axis uses a logarithmic scale for visualization purposes.

Post-hire impact

Clients who use Seedlink models to hire candidates are expected to see organizational impact in terms of increased productivity and increased retention. Below the post-hire impact is evaluated for two cases.

Case: Retention and productivity in a China-based recruitment consultancy firm

A China-based recruitment consultancy firm was facing a challenge in identifying candidates with the potential of becoming productive recruiter consultants. Recruiter consultants generate the revenue of the firm by prospecting clients and filling client vacancies. The most productive consultants primarily excelled in soft skills, such as being *result driven* and *resilient*, for which traditional assessments (e.g. interviews or psychometric personality questionnaires) have limited predictive power.

In 2017, Seedlink developed a custom model that based on Chinese responses to open-ended questions, scored candidates on competencies relevant to becoming productive recruiter consultants. After March 2017, all new hires in the firm applied through Seedlink.

To evaluate the post-hire impact of using Seedlink on the organization, the average revenue of a post-Seedlink cohort (63 employees hired in the period March 2017 to March 2018), a pre-Seedlink cohort (39 employees hired in the period March 2016 to March 2017) and a pre-pre-Seedlink cohort (34 employees hired in the period March 2015 to March 2016) were compared. The results are shown in Figure 6. In each of the first six quarters after onboarding, post-Seedlink hires generated on average more revenue than previous cohorts. After 6 quarters, the accumulated revenue of post-Seedlink hires was approximately 50 percent higher than the accumulated revenue of previous cohorts.

Furthermore, the retention rate of the pre-Seedlink cohort was compared to the post-Seedlink cohort. The results are shown in Figure 7 in the form of normalized retention curves. A small improvement in retention rates was observed 365 days after the onboarding date of about approximately 6-8 percent point compared to the pre-Seedlink cohort and the pre-pre-Seedlink cohort (employees hired between March 2015 and March 2016).

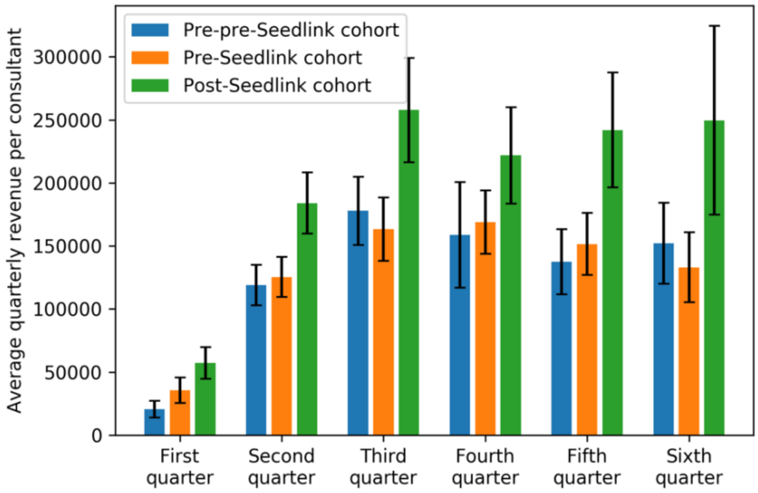


Figure 6. The average quarterly revenue of employees in the first size quarters after onboarding. Error bars denote standard deviation of the mean.

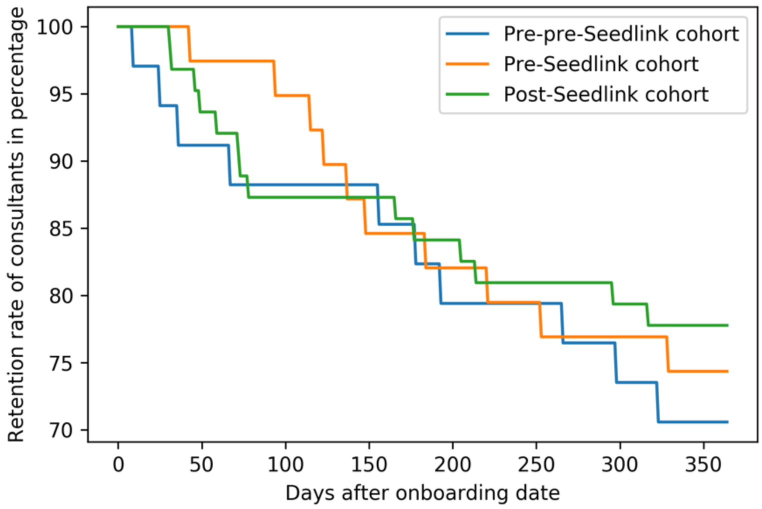


Figure 7. Retention curves of a post-Seedlink cohort, a pre-Seedlink cohort and a pre-pre-Seedlink cohort.

Case: Interns in a global FMCG client

In 2018, a global FMCG client recruited 84 interns for various positions in the United Kingdom. Using a Seedlink predictive model, interns with diverse backgrounds were hired. For example, 28 interns were categorized as diversity hires because of their gender or ethnicity, 22 interns were hired from universities that were not historically preferred, and 10 interns were studying majors that were unrelated to the position they applied for.

At least six months after onboarding, all interns were rated on the key traits and competencies. Specifically, we compare diversity hires versus traditional hires, interns from historically preferred universities to interns from other universities, and interns with position-related majors to those with unrelated majors. The results are shown in Figure 8.

It can be observed that there is no difference in average scores of diversity hires versus non-diversity hires. Hires from historically preferred universities received lower scores than those from other universities, but the difference is not statistically significant (p-value > 0.05).

Interns with majors unrelated to the position for which they were hired received on average slightly lower scores, but also this difference is not statistically significant (p-value > 0.05).

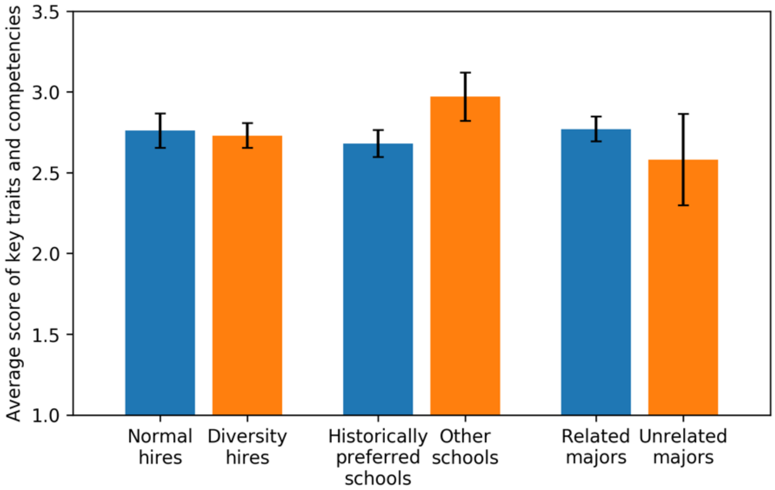


Figure 8. Average score for key competencies and traits of interns at least 6 months after their onboarding date. Error bars denote standard deviation of the mean.

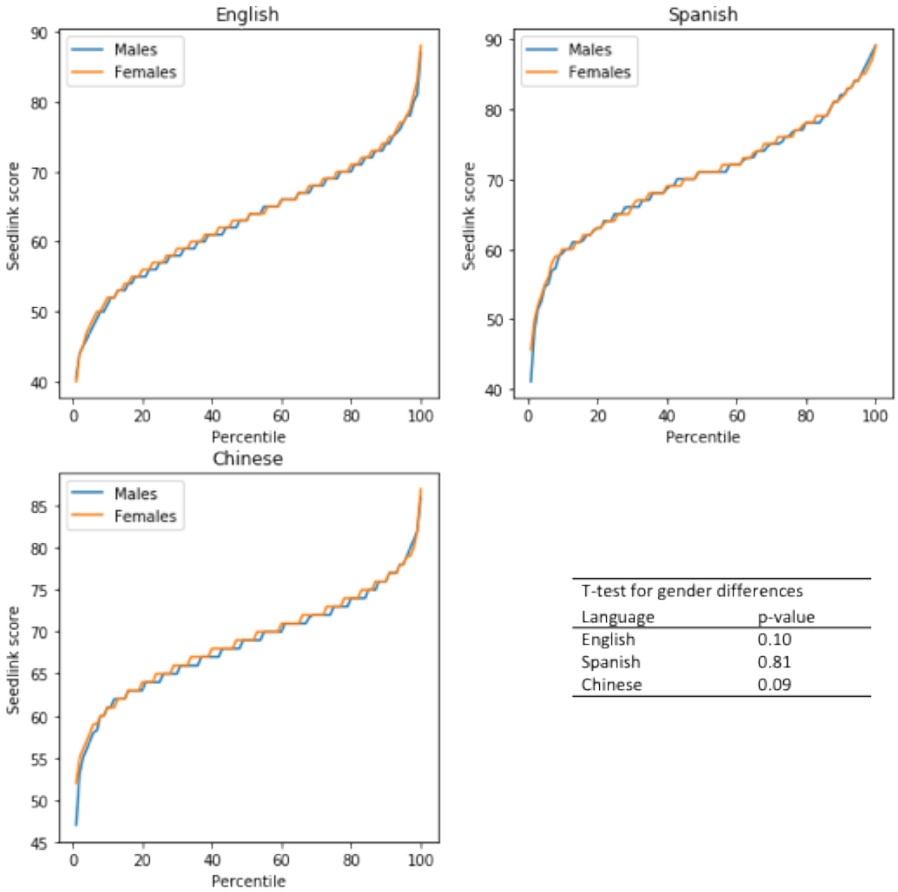
4. Bias

Seedlink aims to help clients reduce human bias in the hiring process and increase diversity in hiring decisions. Particularly for predictive models that assess person-organization fit, it is essential that no unfair advantages are given to candidates with a certain gender, school, major, work experience, first language, age or ethnicity. Below, potential biases in predictive models are evaluated by visualizing the distributions from different groups, and by conducting t-tests between group averages.

Case: Gender bias in global FMCG client

In the past years, Seedlink assisted a global FMCG client with a gender imbalanced workforce (primarily female) with their recruitment practices in various countries, including China, Spain and the United Kingdom. Based on gender-imbalanced training datasets, Seedlink built predictive models for each country using employee-responses to open-ended questions in the local language (i.e. Chinese, Spanish, and English respectively). Next, these models were used to score large numbers of applicants and the dependence of Seedlink score on gender was evaluated by plotting the percentile curves for males and females.

The results are shown in Figure 9. Percentile curves for Seedlink score for males and females appear very similar, indicating that the model does not give an unfair advantage to a certain gender. T-tests for gender differences all reported p-values above 0.05, confirming that there is no significant difference in Seedlink score between male and female candidates.



T-test for gender differences	
Language	p-value
English	0.10
Spanish	0.81
Chinese	0.09

Figure 9. Percentile curves for males and females for Seedlink models in three different languages.

Case: Bias by university major

In 2018, a global FMCG company used Seedlink to recruit candidates in the United Kingdom for various positions. More than 10,000 applications with English answers to open-ended questions were scored using a custom-built model for person-organization fit. The majors of all applicants were known to Seedlink and were used to determine the influence of a particular major on Seedlink score.

The score distributions of management and non-management majors, as well as the score distributions of the 25 most common majors are visualized with boxplots in Figure 10. There is no significant difference between management and non-management majors ($p > 0.05$). Furthermore, the differences between the 25 most common majors are relatively small. Compare these results also to Figure 8, where interns with and without relevant university majors did not receive statistically different performance scores.

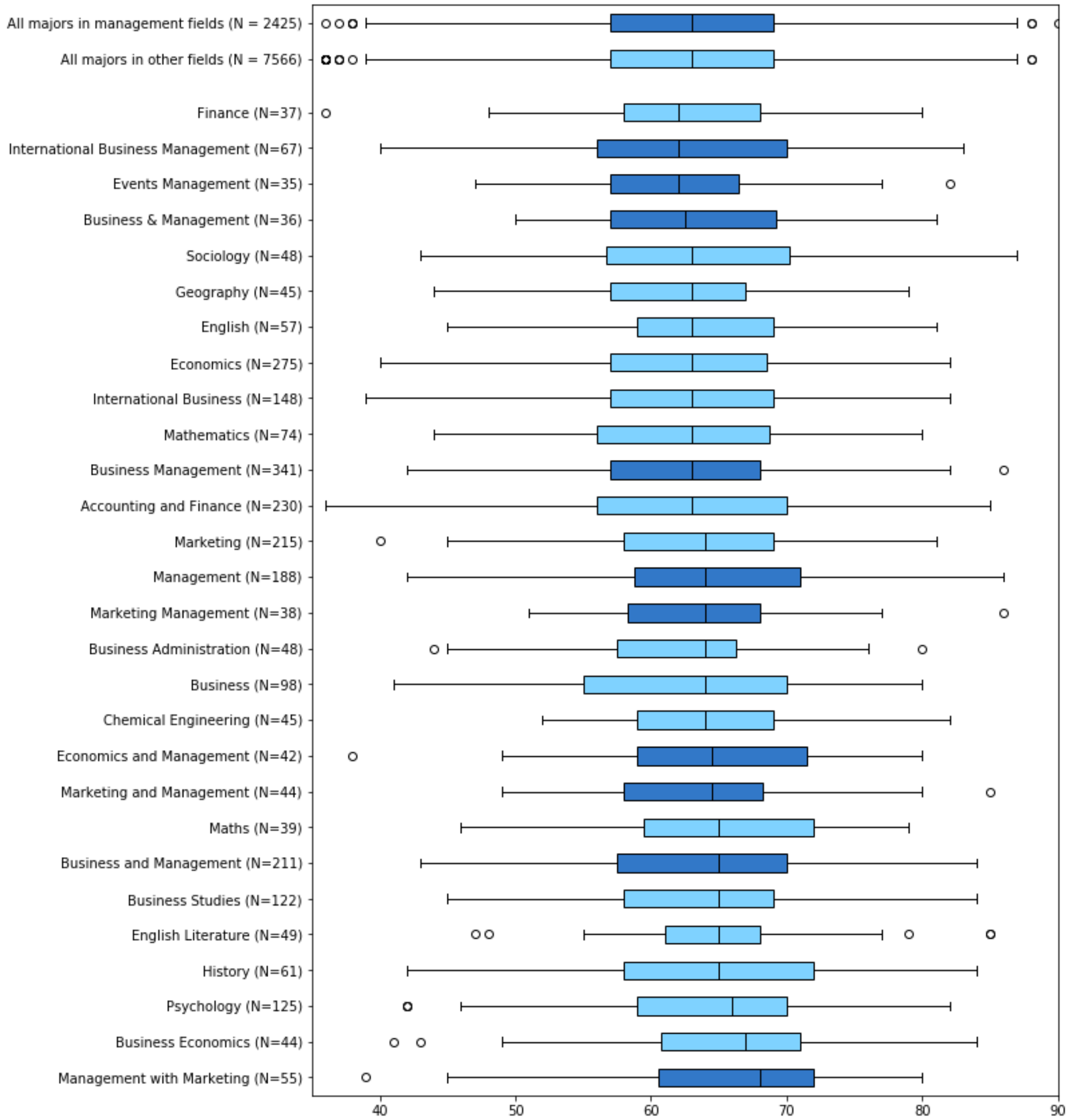


Figure 10. Boxplots showing the score distributions of candidates with different majors. Boxplots from management majors are shown in dark-blue, boxplots from remaining majors are shown in light-blue. The first and second boxplots aggregate over all candidates from management and non-management majors respectively.

Case: Bias by native language

In 2016, a Seedlink predictive model was used to recruit young professionals for a leadership program about sustainability and impact creation. More than 1700 candidates from more than 100 different countries responded in English to four open-ended questions, and Seedlink scores for each candidate were computed. Because the native language of candidates was not known, the official languages of the country of origin were used as an alternative means to determine native and non-native English speakers (Wikipedia, 2019).

Boxplots of score distributions for native and non-native English speakers are shown in Figure 11, as well as individual boxplots for all countries with 15 or more applications. Native English speakers on average scored slightly higher than non-native speakers ($p < 0.05$), particularly native English speakers from first-world countries like United States, United Kingdom, Canada or Australia.

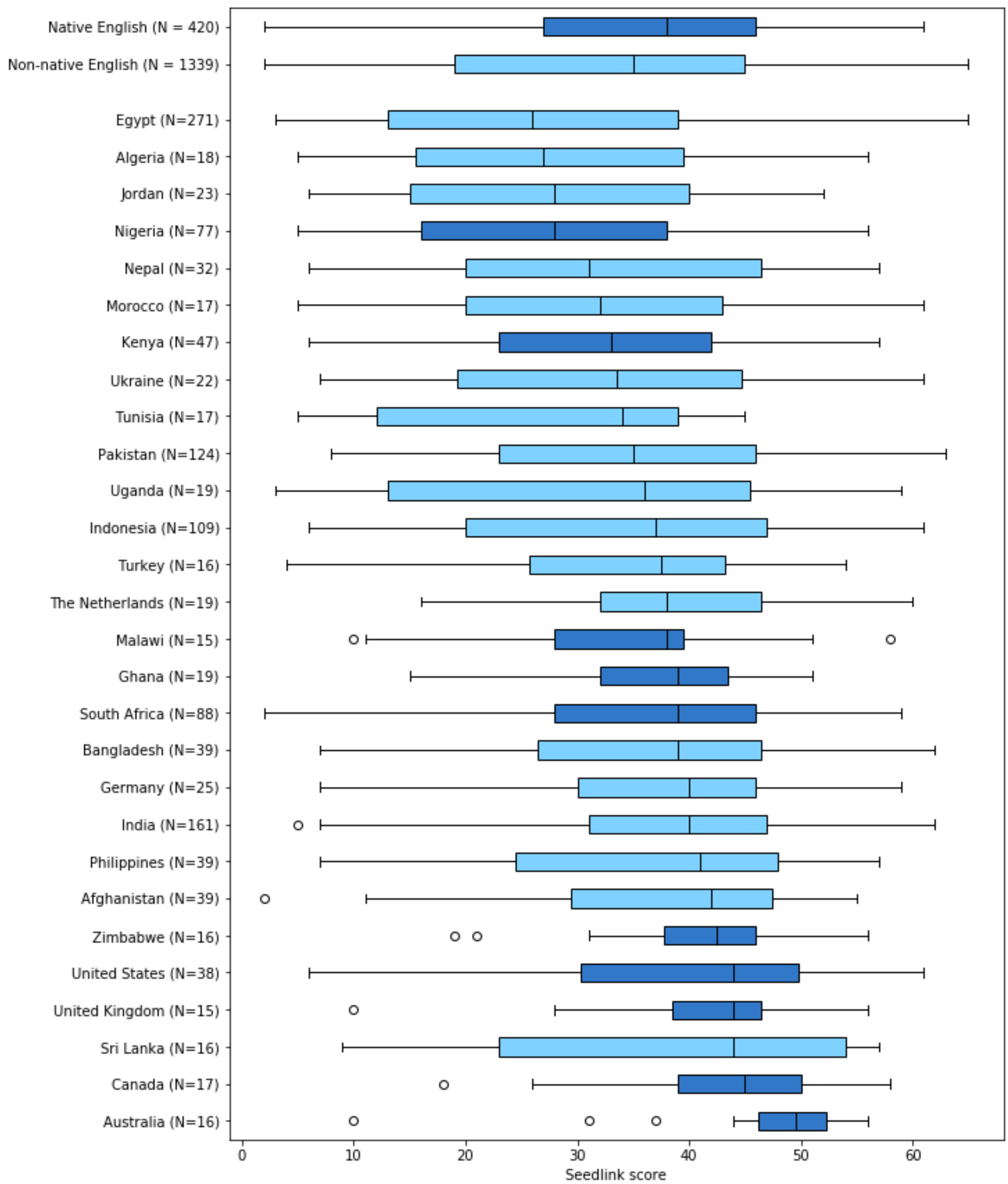


Figure 11. Boxplots showing the score distributions of candidates from different countries. Boxplots from native English-speaking countries are shown in blue, boxplots from remaining countries are shown in orange. The first and second boxplots aggregate over all candidates from non-native and native English-speaking countries respectively. No boxplots are shown for countries with fewer than 15 applications.

5. Discussion and conclusion

Reliability

Interview feedback

An almost identical ranking was observed for assessments based on Seedlink score and assessments based on blind structured interviews. These results are even more noteworthy, when it is realized that the Seedlink assessments required only minutes to complete, whereas the structured interviews required hours. This demonstrates the potential of Seedlink's predictive models to deliver assessments that combine accuracy and time-efficiency. A limitation of this experiment is of course the small sample size (N=8), future work will focus on repeating such experiments with larger sample sizes.

Seedlink scores also appeared to be predictive for interview progression, because in all studied cases, scores gradually increased towards later interview rounds. However, caution is required with drawing strong conclusions here. The recruiters from the client were not blind to the Seedlink score of candidates and are likely to have used it as a selection or progression criterion (which is of course what the Seedlink score is intended for). The increasing Seedlink scores are therefore a combination of two effects, namely the predictive power of the score and the effect of recruiters using the Seedlink score as criterion for selection or progression.

Post-hire impact

The post-hire impact was evaluated on a client case where objective performance measures were available, namely quarterly revenue figures of pre-Seedlink and post-Seedlink cohorts. The post-Seedlink cohort generated quarterly revenues that were approximately 50 percent higher than those of previous cohorts. Although not shown in this white paper, we have rigorously investigated alternative hypotheses to explain the outperformance, e.g. hypotheses related to favorable market conditions after March 2017 or post-Seedlink hires primarily joining teams operating in high-revenue markets, and we conclude that none of the alternative hypotheses explain the outperformance. We will continue to monitor the revenue figures of new hires, and future work will study if outperformance is maintained after 6 quarters.

The second case on which post-hire impact was evaluated showed that candidates with unrelated majors and candidates from non-historically-preferred universities did not underperform compared to their peers. Some clients traditionally excluded such candidates in early stages of the selection process, but as demonstrated by this case, this may need to be reconsidered.

Removing human bias

In all the cases presented in this paper, gender, major, and native language were shown to have no significant or very moderate impact on Seedlink scores. These results support the thesis that Seedlink's methodology is successful in removing bias from the employee selection process. This is primarily explained by the used methodology:

Firstly, competency scores of employees are a relatively bias-free source of training labels compared to other (more easily available) labels, such as recruiter perceptions of applicants or career progression of employees. Recruiter perceptions are mostly dependent on generally desirable characteristics such as articulateness, positive personal appearance, and good general communication skills, rather than more unique characteristics that better predict person-organization fit (Rynes et al., 1993; and Kristof-Brown et al., 2002). The risk of using career progression as a training label is its sensitivity to both historical and current (organizational) biases, which may result in discriminative behavior of the algorithm against females and ethnic minority groups, as was the case with an AI recruitment tool in Amazon (Reuters, 2018).

Secondly, unlike resumes or candidate videos, answers to open-ended questions do not directly provide information of gender, school, major, age or ethnicity. Demographics such as gender, age and race were shown to have no predictive value for person-organization fit (Cable and Judge, 1996). Therefore, restricting direct access of the scoring algorithm to demographics is not expected to impact accuracy of predictive models but does drastically reduce the risk of introducing unfair biases in the Seedlink score.

Conclusion

Seedlink assessments remove human bias and show reliable performance across different languages, industries and job functions.

References

- Argamon, S., Dhawle, S., Koppel, M., & Pennebaker, J. W. (2005, June). Lexical predictors of personality type. In *Proceedings of the 2005 Joint Annual Meeting of the Interface and the Classification Society of North America* (pp. 1-16).
- Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: a meta-analysis. *Personnel psychology*, *44*(1), 1-26.
- Cable, D. M., & Judge, T. A. (1996). Person–organization fit, job choice decisions, and organizational entry. *Organizational behavior and human decision processes*, *67*(3), 294-311.
- Daelemans, W. (2013, March). Explanation in computational stylometry. In *International Conference on Intelligent Text Processing and Computational Linguistics* (pp. 451-462). Springer, Berlin, Heidelberg.
- Fast, L. A., & Funder, D. C. (2008). Personality as manifest in word use: correlations with self-report, acquaintance report, and behavior. *Journal of personality and social psychology*, *94*(2), 334.
- Furnham, A. (1990). Faking personality questionnaires: Fabricating different profiles for different purposes. *Current psychology*, *9*(1), 46-55.
- Gardner, W. L., Reithel, B. J., Cogliser, C. C., Walumbwa, F. O., & Foley, R. T. (2012). Matching personality and organizational culture: Effects of recruitment strategy and the Five-Factor Model on subjective person–organization fit. *Management Communication Quarterly*, *26*(4), 585-622.
- Gill, A. J., Nowson, S., & Oberlander, J. (2009, March). What are they blogging about? Personality, topic and motivation in blogs. In *Third International AAAI Conference on Weblogs and Social Media*.
- Hirsh, J. B., & Peterson, J. B. (2009). Personality and language use in self-narratives. *Journal of research in personality*, *43*(3), 524-527.
- Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *Journal of applied psychology*, *85*(6), 869.
- IBM Personality Insights (2019), The science behind the service. Retrieved from https://cloud.ibm.com/docs/services/personality-insights?topic=personality-insights-science.
- Kandola, B. (2009). *The value of difference: Eliminating bias in organisations*. BookBaby.
- Kristof-Brown, A., Barrick, M. R., & Franke, M. (2002). Applicant impression management: Dispositional influences and consequences for recruiter perceptions of fit and similarity. *Journal of Management*, *28*(1), 27-46.
- Kristof-Brown, A. L., Zimmerman, R. D., & Johnson, E. C. (2005). Consequences of individual's fit at work: a meta-analysis of person-job, person-organization, person-group, and person-supervisor fit. *Personnel psychology*, *58*(2), 281-342.
- Le, Q., & Mikolov, T. (2014, January). Distributed representations of sentences and documents. In *International conference on machine learning* (pp. 1188-1196).
- Lee, C. H., Kim, K., Seo, Y. S., & Chung, C. K. (2007). The relations between personality and language use. *The Journal of general psychology*, *134*(4), 405-413.
- Luyckx, K., & Daelemans, W. (2008). Using syntactic features to predict author personality from text. *Proceedings of Digital Humanities*, *2008*, 146-9.
- Majumder, N., Poria, S., Gelbukh, A., & Cambria, E. (2017). Deep learning-based document modeling for personality detection from text. *IEEE Intelligent Systems*, *32*(2), 74-79.
- Martin, J. (2014). For Senior Leaders, Fit Matters More than Skill. *Harvard Business Review*, Retrieved from https://hbr.org/2014/01/for-senior-leaders-fit-matters-more-than-skill.
- Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of personality and social psychology*, *90*(5), 862.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- Nowson, S., & Oberlander, J. (2006, March). The Identity of Bloggers: Openness and Gender in Personal Weblogs. In *AAAI spring symposium: Computational approaches to analyzing weblogs* (pp. 163-167).

- Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532-1543).
- Wikipedia (2019), List of official languages by country and territory. Retrieved from https://en.wikipedia.org/wiki/List_of_official_languages_by_country_and_territory.
- O'Reilly III, C. A., Chatman, J., & Caldwell, D. F. (1991). People and organizational culture: A profile comparison approach to assessing person-organization fit. *Academy of management journal*, *34*(3), 487-516.
- Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of personality and social psychology*, *77*(6), 1296.
- Reuters (2018). Amazon scraps secret AI recruiting tool that showed bias against women. Retrieved from https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.
- Rynes, S. L., & Gerhart, B. (1993). Recruiter perceptions of applicant fit: Implications for individual career preparation and job search behavior. *Journal of Vocational behavior*, *43*(3), 310-327.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124(2), 262-274.
- Sullivan J. (2017). Ouch, 50% Of New Hires Fail! 6 Ugly Numbers Revealing Recruiting's Dirty Little Secret. Retreived from https://www.ere.net/ouch-50-of-new-hires-fail-6-ugly-numbers-revealing-recruitings-dirty-little-secret/.
- Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: a meta-analytic review. *Personnel psychology*, *44*(4), 703-742.
- Vancouver, J. B., & Schmitt, N. W. (1991). An exploratory examination of person-organization fit: Organizational goal congruence. *Personnel psychology*, *44*(2), 333-352.
- Verhoeven, B., & Daelemans, W. (2014, November). Evaluating content-independent features for personality recognition. In *Proceedings of the 2014 ACM Multi Media on Workshop on Computational Personality Recognition* (pp. 7-10). ACM.
- Verhoeven, B., Daelemans, W., & Plank, B. (2016). Twisty: a multilingual twitter stylometry corpus for gender and personality profiling. In *Proceedings of the 10th Annual Conference on Language Resources and Evaluation (LREC 2016)/Calzolari, Nicoletta [edit.]; et al.* (pp. 1-6).
- Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of research in personality*, *44*(3), 363-373.
- Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based natural language processing. *ieee Computational intelligenCe magazine*, *13*(3), 55-75.