**COMPANY MEMO**

**BLUE ORIGIN**

To: Tom Weiss

From: Julian Williams

Date: Autumn 2023

Subject: 180-Day Quality of Hire: Validation Study

CC: HR Team

References:

1. **EXECUTIVE SUMMARY**
2. **Purpose**

Quality of hire was introduced as a key performance indicator to ensure that recruiters were bringing in the right kind of talent to Blue Origin in April of 2023. The goal was to create a scalable and consistent measure of new hires that can assist with maximizing the prediction of employee performance and attrition. The first iteration of the Quality of Hire survey assessed employee performance after 90 days of tenure at the company. This intervention point can serve as early detection system to identify what skills, qualities, and technical proficiency can begin to emerge within employees from the perspective of managers ([*2023 Quality of Hire: Validation Study Report*](https://blueorigins-my.sharepoint.us/:w:/g/personal/jwilliams15_blueorigin_com/EY1m3ceXU6ZAgBGlnVBmHd8BY1l1uykdwEkUJtolMkS5RA?e=oR38Nj)). Establishing a baseline for what makes an acceptable quality of score required additional data. The purpose of this memo is to provide that additional information by:

1. Exploring the results of a quality of hire survey administered 180-days after an employee has been hired.
2. Determining the utility of the 90-day and/or 180-day Quality of hire survey in Talent Acquisition for the next fiscal year.

Having managers assess their employees after 180 days provides new hires with more time to acquire and develop the necessary skills and knowledge required for their role. Furthermore, the duration enables the identification of individuals who continue to excel and contribute positively beyond the initial onboarding phase, which can be valuable for talent retention and succession planning. Cultural fit is an important factor in employee success and job satisfaction. Assessing after 180 days can create a more holistic picture of new hires alignment with the Blue Origin Leadership principles and build relationships with colleagues and managers. Assessing the quality of hire data at two different time points, specifically at the 90-day and 180-day marks, contributes to a process-oriented approach within the HR business function. Comparing the results of both surveys can provide insight on how employees can evolve in their roles overtime, where managers can utilize resources to support employees, improve the efficiency of current reports, and provide metrics to assist with decision-making on survey distribution.

## 3.0 Methodology & Survey Composition

For the purpose of evaluating the quality of hire assessment scores between two different time points, the study sources from the [*2023 Quality of Hire: Validation Study Report*](https://blueorigins-my.sharepoint.us/:w:/g/personal/jwilliams15_blueorigin_com/EY1m3ceXU6ZAgBGlnVBmHd8BY1l1uykdwEkUJtolMkS5RA?e=oR38Nj) to define Quality of hire as representing an employee’s overall potential to raise the bar at Blue Origin. Managers will assess new employees across the following quality of hire dimensions:

**The extent to which employees:**

*Fits Blue Origin’s culture*

*(Leadership Principles)*

***Have the skills and knowledge for the role* the**

***An employee meets the bar in 180 days* the**

***Employees are assigned the correct level and role* the**

The four dimensions of the 180-day quality of hire assessment is made up of 11 questions, two being removed in the previous iteration due to item redundancy. There are also two demographic questions to help determine the employee type and how long the employee has been a manager’s direct report. Managers rated their employees on a 5-point Likert-type scale ranging from *(1) strongly disagree* to (*5) strongly agree*.

|  |  |  |
| --- | --- | --- |
| Dimension | Example Questions | Response Scale |
| Has Required Level of Knowledge and Skills | *This employee has the technical background to succeed in this role.*  *This employee has the skills to succeed in this role.*  *This employee has produced quality work.* | *5. Strongly Agree*  *4. Agree*  *3. Neither Agree or Disagree*  *2. Disagree*  *1. Strongly Disagree* |
| Fits with the Leadership Principles | *This employee works well with others.*  *This employee is motivated to do well.*  *This employee demonstrates our Leadership Principles.* |
| Fits with Level and Role | *This employee was placed at the correct job level (for example, Engineer III versus Engineer II).*  *~~This employee's current role is a good fit for them.~~*  *This employee will succeed in the current role.* |
| Raises the Bar | *I would hire this employee again.*  *This employee raises the bar for this role.*  *~~This employee was a good hire.~~*  *Based on my observations so far, this employee has the potential to be a top performer (i.e., in the top 5%).* |

\*\* ~~Strikethroughs~~ indicate questions that were excluded from analyses due to item redundancy

These dimensions were maintained to assess the external validity of the measure and utilize the existing automations embedded in Qualtrics for survey distribution ([Quality of Hire Automation Report 2023.docx](https://blueorigins-my.sharepoint.us/:w:/g/personal/jwilliams15_blueorigin_com/ESvGjoAK7UZBkubLpD3Caw0BxKd5nG8MSH0PS1O3_RMJ6g?e=6s9hE8)). The main alteration in the 180-day survey was the use of a response matrices instead of multiple-choice questions to improve user experience while completing the assessment (See Appendix A.)

# 4.0 180-Day Quality of Hire Survey Results

**4.1 Response Rates and Sample Size:** The research team prioritized real-time data to drive this study, and identified 110 employees that had been at Blue Origin for 180 days as of March 15th, 2023. In this quality of hire study, the researcher initiated participant engagement through an email distributed on October 11th, 2023, to a sample of 110 individuals. The survey had an average completion period of approximately 2-3 minutes. Outliers were excluded from average to complete if they were less than 1.5 minutes and more than 15 minutes. The survey was distributed to all 110 participants on October 11th, 2023, with subsequent reminders sent on October 23rd and October 31st.

Initially, 81 participants (representing a 74% response rate) engaged with the study. During the data cleaning process three main incidences constituted the removal of cases from the sample:

1. Duplicate completion of the survey by a manager assessing the same employee
2. Distributions emailed as tests for working links and survey logic
3. Incomplete survey responses

The exclusion of responses deemed unqualified for the study's criteria led to five tests, four duplicate responses, and seven incomplete responses being removed (15 total). The effective response rate was adjusted to 65 out of the original 110, equating to a 59% valid response rate.

**TABLE: RESPONSE RATES OF SAMPLE**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sample (N) | Survey Responses | Duplicate Reponses | Test  Responses | Incomplete  Responses | Effective Responses |
| 110 | 81 | 4 | 5 | 7 | 65 |

Smaller sample sizes may not accurately represent trends within a large population, but they can still be used to give insight on research questions regarding specific subgroups (smaller demographics). Furthermore, the findings of this initial study can help guide the design of future research with more representative samples.

**4.2 Quality of Hire Raw Scores**

The 180-day Quality of Hire survey raw scores provide an in-depth look into the performance and adaptability of new hires across various job families, supervisory organizations, and job levels. These scores are critical in depicting a comprehensive visual of how segments within the organization are integrating and performing.

Table: Average Scores Across All Supervisory ORGANIZATIONS

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | N | mean | std | min | 25% | 50% | 75% | max |
| **Technical Skills** | 58 | 14.0 | 1.2 | 12 | 13 | 14.5 | 15 | 15 |
| **Leadership Principles** | 60 | 14.3 | 1.1 | 12 | 14 | 15 | 15 | 15 |
| **Level Fit** | 59 | 9.3 | 0.8 | 8 | 9 | 10 | 10 | 10 |
| **Overall Performance** | 47 | 14.0 | 1.6 | 8 | 13 | 15 | 15 | 15 |
| **Quality of Hire Score** | 44 | 52.7 | 2.9 | 45 | 51 | 54 | 55 | 55 |

**TABLE: QUALITY OF HIRE SCORE BY SUPERVISORY ORGANIZATION**

|  |  |  |  |
| --- | --- | --- | --- |
| Supervisory Organization | **N** | **Average Quality of Hire score** | **Standard Deviation** |
| Blue Engines | 2 | 90 | 11.14 |
| Engineering and Program Management | 2 | 90 | 1.59 |
| In-Space Systems | 1 | 93 | -- |
| Lunar Transportation | 3 | 98 | 2.25 |
| New Glenn | 7 | 86 | 10.77 |
| New Shepard | 4 | 87 | 12.86 |
| Operations | 25 | 82 | 33.05 |
| Safety and Mission Assurance | 1 | 93 | -- |
| Space Systems Development | 1 | 96 | -- |
| Test and Flight Operations | 2 | 89 | 15.91 |

Several of the supervisory organizations have a representative sample of less than three. The average scores are not an estimate of overall quality of hire because for each organization because of the lack of sample size. Another way to garner insight from small samples is comparing similar subgroups, for example: Blue Engines, Engineering and Program Management, Test and Flight Ops, New Shepard, and Lunar Transportation have a similar sample size of respondents, but the variance of manager ratings for new hires within the business unit vary greatly.

**FIGURE: SUPERVISORY ORGANIZATIONS WITH SIMILAR SAMPLE SIZE-QUALITY OF HIRE SCORES COMPARED**

**Chart, scatter chart

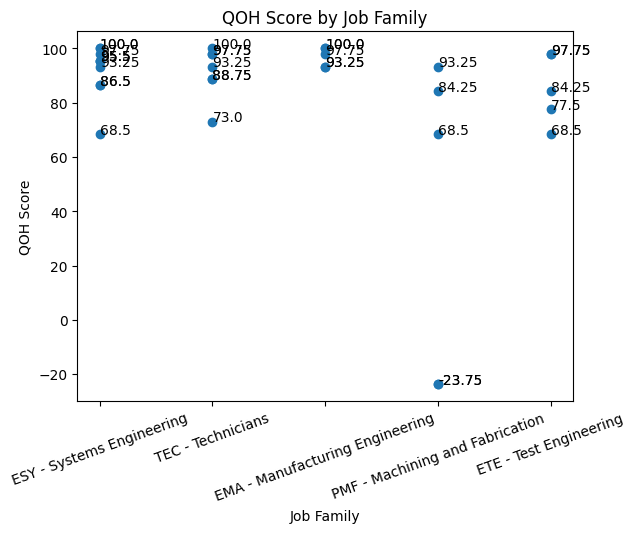
Description automatically generated**

**TABLE: QUALITY OF HIRE SCORE BY JOB FAMILY**

|  |  |  |  |
| --- | --- | --- | --- |
| Job family | N | Average Quality of Hire Score | Standard Deviation |
| EAH - Avionics Engineering | 1 | 82 | -- |
| EFS - Flight Sciences Engineering | 1 | 93 | -- |
| EHS - EHS Operations | 1 | 93 |  |
| EMA - Manufacturing Engineering | 5 | 97 | 3 |
| EME - Mechanical Engineering | 1 | 75 | -- |
| EMP - Materials & Processes​ Engineering | 1 | 89 | -- |
| EPR - Propulsion Engineering | 1 | 98 | -- |
| ESY - Systems Engineering | 10 | 92 | 9 |
| ETE - Test Engineering | 5 | 85 | 13 |
| PMF - Machining and Fabrication | 5 | 40 | 59 |
| PPC - Production Planning | 1 | 78 | -- |
| PPM - Program Management | 2 | 93 | 3 |
| QAI - Quality Assurance | 4 | 92 | 9 |
| SCL - Logistics | 1 | 89 | -- |
| SWD - Software Development | 1 | 96 | -- |
| TEC - Technicians | 7 | 91 | 9 |
| TSP - Technical Specialties | 1 | 100 | -- |

A majority of the job families have a sample size of less than four respondents. But the variability between each of the data points in job families in larger groups show notable differences. Below are the quality scores of the systems engineering, technician, manufacturing engineering, test engineering, and machining & fabrication job families.

**FIGURE: JOB FAMILIES WITH SIMILAR SAMPLE SIZE-QUALITY OF HIRE SCORES COMPARED**



**4.3 Quality of Hire 180-Day Reliability Scores**

Researchers conducted this analysis to asses the extent to which quality of hire scores are psychometrically reliable. Reliability coefficients, such as Cronbach's alpha, quantify the level of error in these measurements, ranging from 0.00 to 1.00. The psychometric target is typically a reliability coefficient (r) of 0.70 or higher (error variance of 30%). The results from the current analysis demonstrated:

**180-DAY RELIABILITY ANALYSIS SCORES**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Overall Quality of Hire Scale | Technical Skills Factor | Blue Leadership Principles Factor | Fit with Role Factor | Overall Performance Factor |
| 0.94 | 0.84 | 0.57 | 0.66 | 0.83 |

**r =**

Larger sample sizes generally provide more stable and reliable Cronbach's Alpha estimates. The overall measure indicates a high degree of consistency in the responses. However, the blue leadership and fit with role subfactors miss the psychometric target by 13% and 4%. Conclusions drawn about these subfactors should be considered tentative, because the inconsistency may indicate that managers are interpreting items differently or lack of cohesiveness among items in the subfactor.

**4.3 180-day Quality of Hire Correlations**

The correlation coefficient measures the relationship between two factors. By examining the correlation coefficients, businesses can gain insights into how different factors are related and make informed decisions based on these relationships (See Appendix B).

**TABLE: 180-DAY CORRELATIONS BETWEEN QUESTIONS**



Correlations calculated from small sample sizes may be more prone to fluctuation. This instability can make it challenging to generalize the results to the broader population. Based on the correlation analysis, there is a potential for several significant associations between different factors:

1. Perceived **quality of work** is strongly associated with the likelihood of **raising the bar**.
2. Willingness to **hire someone again** is strongly linked to perceiving that person as someone who **raises the bar**.
3. **Belief in an individual's future success** is strongly linked with their perceived ability to **raise the bar.**
4. **Having skills** is moderately to strongly associated with **demonstrating leadership principles** and **producing quality work**.
5. **Being motivated** correlates moderately with **demonstrating leadership principles and raising the bar**.
6. **Working well with others** generally has lower correlations with other factors, suggesting that it might be perceived as a somewhat distinct attribute.

These findings can provide insights can provide insight for deeper dives into analyses with larger samples, helping to understand the relationships between different factors and make informed decisions based on which to target as well as ways to leverage manager’s relationships with their new hires.

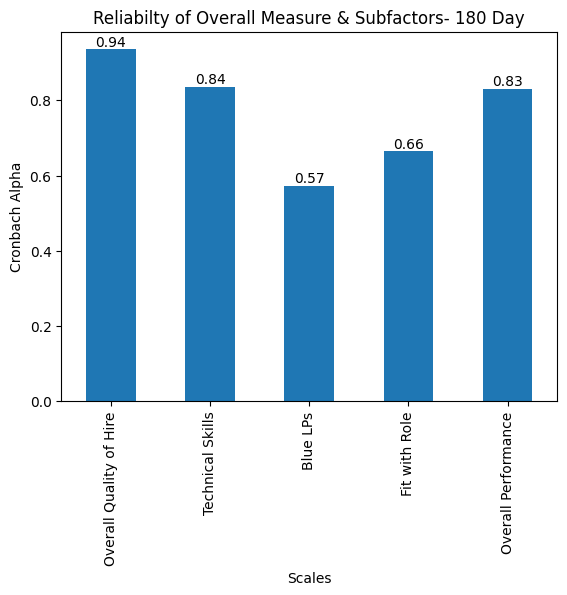
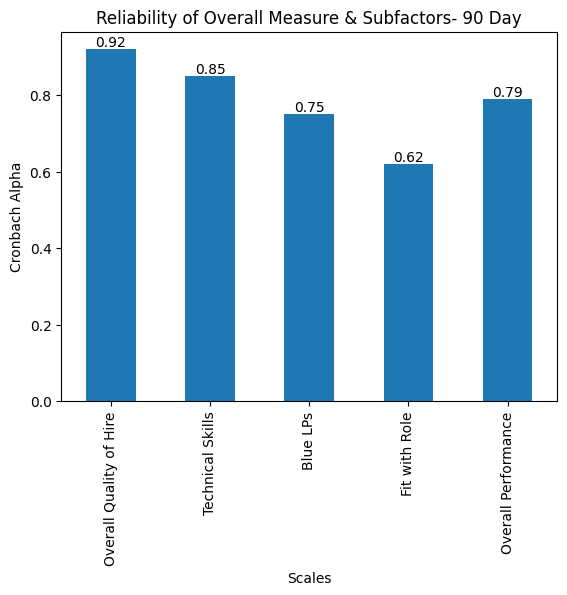
# 5.0 180-Day and 90-Day Quality of Hire Comparison

A comparison between the two surveys was conducted to investigate the need to integrate both or remove one of the measures for the quality of hire assessment process. The comparison highlighted the dynamic nature of employee development and the evolving perceptions of managers on the performance of their employees. The research team prioritized real-time data to drive this study and identified 245 employees (N=245) that had been at Blue Origin for 90-days as of August 17th, 2023.

**5.1 Reliability**

The output from the 90-day and 180-day reliability analyses provide Cronbach's Alpha scores for the overall scale and various factor items within two different versions (points of intervention) of the scale.

**FIGURE: RELIABILITY SCORES**

**Interpretation and Considerations:** The analysis reveals a notable increase in the reliability of the overall scale and several factors at the 180-day mark compared to the 90-day mark, despite issues with sample characteristics. This enhancement suggests a maturation in responses over time, with respondents likely providing more informed and consistent answers as they become more accustomed to their roles and the survey context. The variances in reliability scores between the two time points underscore the dynamic nature of respondent perceptions and attitudes over time.

**Influence of Sample Characteristics and Response Maturation:** The observed differences in reliability, despite the smaller sample size in the 180-day survey, can be attributed to several key factors. First, the sample characteristics at the 180-day mark might represent a more engaged or reflective subgroup, leading to more consistent responses. Second, respondents' learning and adaptation over time likely contribute to more stable and accurate reflections on their experiences, thus enhancing the reliability of their responses. Finally, the increased familiarity with the survey's constructs and a more mature understanding of their roles contribute to the increased reliability observed in the 180-day survey.

**5.2 180-Day & 90-Day Correlations Comparison**

The comparison between the two surveys reflect the complex interplay between skills, background, motivation, teamwork, and learning potentials in influencing various aspects of work performance and perceptions over different time frames. The 90-day survey had a representative sample with 178 more effective responses than the later iteration. Correlation coefficients in the 180-day may be inflated due to the reduced sample size. Insights on consistency across the two time points despite sample restrictions can be identified (See Appendix C).

**TABLE: 180-DAY CORRELATIONS BETWEEN QUESTIONS**



The correlation between **has skills** and **produced quality** remains strong in both surveys but shows a slight decrease in the 180-day survey, indicating a subtle shift in how these factors interact over time.

The strong correlation between **hire again** and **raises bar** in the 180-day survey, which is slightly higher than in the 90-day survey. This indicates that the perception of an employee’s ability to raise performance standards becomes more pronounced over time, influencing rehire decisions.

The correlations related to **demonstrates leadership principles** and factors like **has skills** **and is motivated** are notably stronger in the 180-day survey despite sample size. This suggests the potential that the ability to demonstrate leadership principles becomes more apparent and relevant with longer tenure, aligning with a deeper understanding of an employee's capabilities.

The 90-day survey shows moderate correlations in **motivation**, which are consistent but not as pronounced as in the 180-day survey. This consistency over time highlights the importance of these factors in early and ongoing assessments of employee performance and fit within the organization.

# 6.0 Recommendations Future Implications

6.1.1 Administering the 180-Day Quality of Hire Survey with a Larger Sample:

The 180-Day Quality of Hire Survey is a critical tool for gaining insights into the long-term effectiveness of new hires. However, to enhance the validity and generalizability of its findings, it is imperative to administer it with a larger sample size. Gaining support from Human Resource Business Partners is crucial in emphasizing the importance of this initiative. A larger sample size will enable more robust data analysis, ensuring that the survey results are reflective of the diverse workforce. It will also help in identifying trends and patterns that may not be visible with a smaller sample. This, in turn, aids in making informed decisions about recruitment strategies and talent management. The involvement of HR Business Partners in advocating for this initiative is key to its success, as their support can drive organizational commitment and resource allocation toward this effort.

**6.1.2 Pros and Cons of the 90-Day Survey; Emphasizing its Necessity:**

While the 90-Day Quality of Hire Survey offers valuable early insights into new hires' initial adaptation and performance, it has its limitations. The primary advantage is the early detection of potential issues and successes, allowing for timely interventions. However, it may not fully capture the long-term potential and cultural integration of the employees. Despite these cons, the 90-Day survey remains the best and safest option currently, especially until the 180-Day survey is more robustly implemented. Researchers often value 180-day assessments due to their more holistic understanding of managers' perceptions of their employees, but until such time, the 90-Day survey provides a necessary, albeit limited, perspective.

**6.1.3. Expansion of Items to Capture Culture through Blue Leadership Principles:**

Despite the smaller sample size, there is a need to expand the survey items that capture organizational culture, particularly through the lens of Blue Leadership Principles. This expansion is critical to comprehensively assess how new hires align with and contribute to the company's cultural ethos. However, it is important to note that the reliability score for these items has not been consistently high across the two time points. This inconsistency calls for a careful review and potential revision of these items to ensure they accurately and reliably reflect cultural fit and alignment with leadership principles.

**6.1.4 Understanding Low Reliability Scores for Fit with Role:**

The low reliability scores for the 'fit with role' subfactor in both the 90-Day and 180-Day surveys raise concerns about the clarity and structure of how managers assess job level and role fit. This could indicate that 90 days may not be sufficient for managers to make a definitive judgment on an employee's long-term success in their role. It suggests a need for more structured guidelines or criteria for managers to effectively assess fit with role. This insight also underlines the importance of the 180-Day survey, which can provide a more extended period for evaluation and a more accurate assessment of long-term fit and performance.

**7.0 APPENDICES**

**APPENDIX A: QUALITY OF HIRE SURVEY**

This short survey contains questions about a new employee that you hired six months ago. Quality of Hire represents an employee's overall potential to raise the bar at Blue Origin. You will be asked about the new hire's level of knowledge and skill, their fit with Blue Origin's Leadership Principles, their fit with the role, and their overall performance. These responses are confidential and will only be shared at the aggregate Please enter your first name, last name, and email address here:

Your First Name:

Your Last Name:

Your Blue email address:

For questions or concerns, please contact your HRBP. Thank you!

**Here is the new hire you will be rating:\_\_\_\_\_\_. This is an employee on your team that was hired 180+ days ago. Does this person still work for you?**

* Yes, I am this person's manager
* No, I am not this person's manager

**Demographic Questions**

**Which category best describes this new employee?**

* External Hire: No prior work history with Blue
* Contractor Conversion: They were performing this role, or a similar role
* Intern Conversion: They were performing this role, or a similar role
* Internal Transfer: They were already a FT Employee at Blue
* Former Blue Origin Employee: They had previously separated from Blue and have now returned
* I Don't Know

**How long have you managed this employee?**

* One month or less
* More than 1 month, but less than or equal to 3 months
* More than 3 months, but less than or equal to 6 months
* More than 6 months
* I am not this person's manager

**QOH Measure**

**New Hire's Knowledge and Skills**  
The following questions ask about your employee's technical ability in their role. Please respond based on your observations of the employee in the last 180 Days.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Strongly Disagree** | **Disagree** | **Neither Agree or Disagree** | **Agree** | **Strongly Agree** |
| **This employee has the technical background to succeed in this role.** |  |  |  |  |  |
| **This employee has the skills to succeed in this role.** |  |  |  |  |  |
| **This employee has produced quality work.** |  |  |  |  |  |

**New Hire's Fit with Blue**  
The following questions ask about how well your employee embodies the Leadership Principles at Blue. Please respond based on your observations of the employee in the last 180 Days.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Strongly Disagree** | **Disagree** | **Neither Agree or Disagree** | **Agree** | **Strongly Agree** |
| **This employee works well with others.** |  |  |  |  |  |
| **This employee is motivated to do well.** |  |  |  |  |  |
| **This employee demonstrates our Leadership Principles.** |  |  |  |  |  |

**New Hire's Level and Role Assignment**  
The following questions ask about your new employee's fit with the role you have assigned them. Please respond based on your observations of the employee in the last 180 Days.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Strongly Disagree** | **Disagree** | **Neither Agree or Disagree** | **Agree** | **Strongly Agree** |
| **This employee was placed at the correct job level (for example, Engineer III versus Engineer II).** |  |  |  |  |  |
| **This employee will succeed in the current role.** |  |  |  |  |  |
| **This employee's current role is a good fit for them.** |  |  |  |  |  |

**Overall New Hire's Evaluation**  
The following questions ask about your perspective on the new employee. Please respond based on your observations of the employee in the last 180 Days.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Strongly Disagree** | **Disagree** | **Neither Agree or Disagree** | **Agree** | **Strongly Agree** |
| **I would hire this employee again.** |  |  |  |  |  |
| **This employee raises the bar for this role.** |  |  |  |  |  |
| **This employee was a good hire.** |  |  |  |  |  |
| **Based on my observations so far, this employee has the potential to be a top performer (i.e., in the top 5%).** |  |  |  |  |  |

## Appendix B: Reading Correlation Coefficients

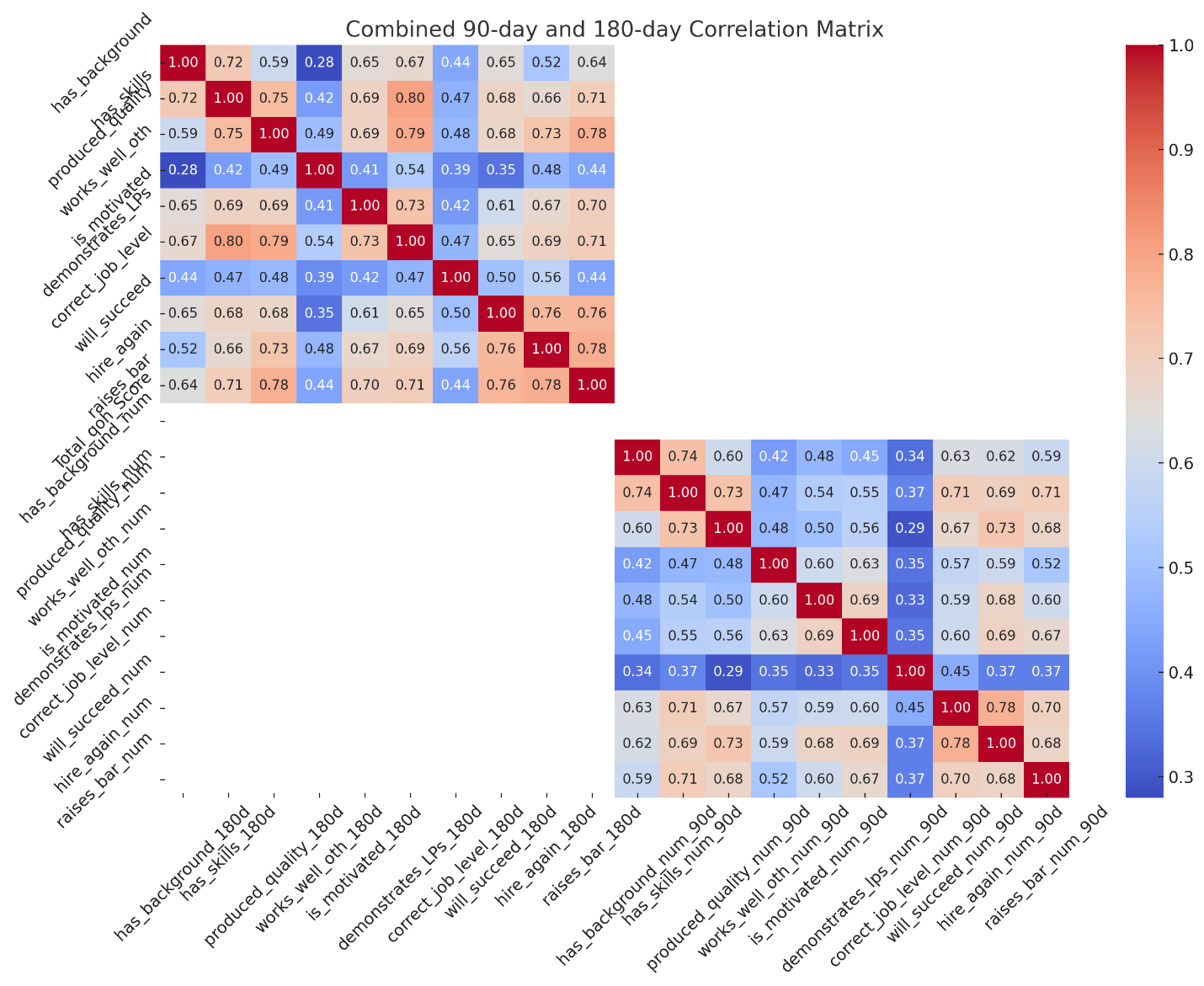
A value close to 1 indicates a strong positive correlation, meaning that as one factor increases, the other factor tends to increase as well. Similarly, a value close to -1 indicates a strong negative correlation, meaning that as one factor increases, the other tends to decrease. On the other hand, a value around 0 suggests no correlation between the factors, meaning that changes in one factor don't have a consistent impact on the other.

**Small Correlation:** A correlation coefficient between 0.1 and 0.3 is typically considered small.

**Medium Correlation:** A correlation coefficient between 0.3 and 0.5 is generally considered medium.

**Large Correlation:** A correlation coefficient above 0.5 is often considered large.

APPENDIX C: Heat Map of 180-Day & 90-day Correlations



NOTE. The heatmap above visually represents the combined 90-day and 180-day correlation matrices. In this visualization, you can compare the correlations for each attribute over the two different time frames. The color gradient, ranging from blue (low correlation) to red (high correlation), makes it easier to identify patterns and differences between the two periods.