



Phenom Fit Score:

2025 Audit Report

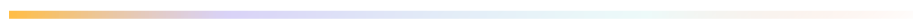


Table of Contents

Executive Summary	1
Fit Score Overview	2
Validity and Reliability Motivation and Methods	9
Validity and Reliability Results	13
Bias Audit Motivation and Methods.....	18
Bias Audit Results	22
Resources and References	26

Executive Summary

Phenom Fit Score is a productivity tool that helps recruiters decide whether or not to move a candidate forward in a selection process. Fit Score alone does not make hiring decisions.

As in prior years, Phenom has examined the potential adverse impact of Fit Score using an industry-standard, legally defensible, and science-based approach. New for 2025, we have extended the audit to examine the criterion validity of Fit Score.

Statistical tests show that the use of Fit Score is reliable and predicts relevant hiring outcomes. Tests also show that Fit Score is unlikely to introduce or worsen adverse impact during the interview decision stage of an employee selection process:

- **Fit Score is a stable method to assess candidate fit.**
- **Candidates with a higher Fit Score were found to be more likely to progress in the hiring process than those with a lower Fit Score.**
- **Fit Score recommended moving candidates forward to the interview phase at rates that were comparable across gender groups, racial/ethnic groups, and intersectional groups of gender and race/ethnicity.**

These findings show that Fit Score contributes to customers' fair hiring processes, establishing the tool's relevance to hiring.

This report is intended to give Phenom customers insight into the Phenom Fit Score model. The information provided herein does not and is not intended to constitute legal advice.

Fit Score Overview

Introduction to Phenom Fit Score

What is Phenom Fit Score?

Fit Score is an AI model that powers the Phenom Intelligent Talent Experience platform. This model helps recruiters identify candidates to interview, bringing speed and objectivity to talent acquisition. **Fit Score on its own does not make hiring decisions.** Fit Score is designed to be a helpful tool in the beginning stage of a hiring process, where it should be overseen by humans.

Fit Score is built using traditional machine learning technologies; it is not a Large Language Model. For further technical information on Fit Score and other Phenom AI features, please refer to [Ethical AI Development at Phenom](#).

Why did Phenom develop Fit Score?

Phenom developed Fit Score to help recruiters handle large flows of job seeker data and empower them with objective and fair decision support. The reality of modern recruiting is that the volume of job seekers can be overwhelming, such as when a Phenom career site attracts many more applicants than a traditional system. Further, recruiters may be unfamiliar with some applicants' highly specialized skills. These challenges of large volumes of data and résumé complexity can be exacerbated by time pressures from hiring managers and job seeker expectations. Under pressure, human decision-making can introduce bias or make bias harder to detect.

Fit Score rises to these challenges by streamlining the hiring process and saves time and resources by prioritizing candidate job fit in terms of objective, job-related characteristics. This allows recruiters to maintain oversight over a large number of job applications that need to be reviewed manually, ultimately leading to better outcomes for both the organization and job candidates.

How does Phenom Fit Score work?

Fit Score works with objective, job-related characteristics, such as relevant career paths and skills-based hiring, to compute a value that aims to predict how a given candidate profile fits the job match criteria. To a recruiter, this value is represented as a letter grade: A, B, C, and No Fit.

In brief, the Fit Score model computes the letter grade by comparing the candidate profile, which is typically based on a résumé, to the job description. The AI behind Fit Score includes AI language modeling of similarities between candidate information and job-related criteria, such as the match between a candidate's skills and the skills listed in a job description. Fit Score is applied to all candidates equally, regardless of their demographic or other protected characteristics. The only inputs to Fit Score are job-related; there are no inputs of demographic or biometric information.

What are appropriate ways of using Fit Score?

Fit Score helps recruiters decide whether or not to move a candidate forward in a selection process. In general, the higher the letter grade, the stronger the match between a candidate profile and the job match criteria.

One way to think about Fit Score is that it is an approximate, at-a-glance summary of how well the candidate matches a job. Without AI, a recruiter could sort and filter candidates only by structured fields or Boolean keyword queries, such as the name of a specific prior employer or an exact matching job title. Without AI, a search for a "computer programmer" would miss all those who had the title "software engineer," and vice versa. With Fit Score, it is possible to sort and filter candidates in terms of their attributes relative to job requirements, such as whether the résumé lists a skill that is a close semantic match to the job criteria. The recruiter can manage the inputs to Fit Score for both the candidate and the job, the latter affecting all candidates. In this way, Fit Score can represent the job-relevant attributes that are important to the recruiter.

Recruiters and employers have established workflows prior to becoming Phenom customers, and these workflows can incorporate Fit Score in different ways. Consider these examples:

Calibrating the matching criteria

A recruiter posts a position and receives a number of applications, each of which is processed by Fit Score. The recruiter reviews several applications that have Fit Score grades of “A” and “No Fit” to confirm that Fit Score is calibrated in accordance with the recruiter’s expectations. The recruiter can then use Fit Score to have an at-a-glance summary of the application strengths in terms of the job matching criteria.

Adjusting the job description

A recruiter posts a position and receives a number of applications, each of which is processed by Fit Score. Fit Score gives these applications grades “C” and “No Fit”; no applications receive grades “A” or “B”. The recruiter uses this distribution of grades in the applicant pool to determine that the job posting is not reaching its desired audience and revises the posting and its marketing. After revising the posting, all of the original applications are automatically re-scored to enable a more comprehensive review of the applicant pool.

Emphasizing objective, job-relevant attributes

A recruiter posts a position and receives a large number of applications, each of which is processed by Fit Score. In the past, to save time and resources, the recruiter’s go-to strategy without Fit Score was to only closely examine the applicants who were employed in a similar role by prominent competitors. This approach screens candidates and imparts a condition that may not necessarily be job-relevant. There may be a risk of overlooking diverse talent. With Fit Score, the recruiter has a better alternative. Fit Score emphasizes decision-making based on job-relevant attributes, reducing the need for screening with potentially irrelevant criteria. As Fit Score is computed for all applications, all candidates can be considered fairly. The recruiter can now prioritize applicants who are especially relevant to multiple job-relevant variables at once. For example, applications that receive grades “A” or “B” should be reviewed manually for additional factors before scheduling initial interviews.

Fine-tuning to hiring manager preference

A recruiter posts a position and receives a number of applications, each of which is processed by Fit Score. On paper, the position has two hiring managers, but the recruiter knows which of the two will make the ultimate decision. The recruiter adjusts the job matching criteria to express that manager’s priorities especially closely.

By contrast, it is inappropriate to use Fit Score as the sole basis or most important factor in choosing how to act on applications, or to use Fit Score to overrule human decision-making.

Is it possible to restrict or disable Fit Score?

Fit Score is an optional feature of the Phenom platform. Phenom can disable Fit Score for all of a customer's jobs, or only for jobs in specific geographic regions.

The configuration of Fit Score for applicants is independent of the configuration of the AI Discovery feature. AI Discovery is a sourcing function, not a tool to facilitate a selection process. Either of these features can be enabled or disabled independently.

Is Fit Score an assessment instrument under the Uniform Guidelines?

Fit Score involves evaluating candidate qualifications and some organizations may choose to use Fit Score as an assessment instrument. At the same time, Fit Score does not have a fixed set of questions or "items", as a traditional assessment would. By contrast, its application is tailored to each specific job posting, with the matching criteria adjustable by an employer to align with different hiring managers' priorities and job requirements.

How can employers use Fit Score while complying with New York City Local Law 144?

New York City Local Law 144 (LL 144) "prohibits employers and employment agencies from using an automated employment decision tool unless" certain conditions are met to audit the tool and notify the public.

One way for an employer to comply with LL 144 is to use Fit Score in a manner other than as an Automated Employment Decision Tool (AEDT). As designed, Phenom Fit Score is not intended to be an automated employment decision tool. It is inappropriate to use Fit Score as (i) the sole basis or (ii) most important factor in choosing how to act on applications, or (iii) to use Fit Score to overrule factors such as human decision-making. The manner in which Fit Score is used is the choice of an employer and it is not controlled by Phenom.

A second way to comply with LL 144 is to satisfy the requirements of LL 144 to audit the tool and notify the public. Phenom retains an independent, third-party auditor to perform a regular audit of Fit Score, which includes the data of multiple customers. The outcome of Phenom's audit is available [here](#). Under LL 144, an employer “may rely on a bias audit of an AEDT” under one of two conditions ([§ 5-302 Data Requirements](#)):

- **The employer provided their historical data to the auditor conducting the bias audit, OR**
- **The employer has never used the vendor's AEDT before**

Therefore, employers who wish to comply with LL 144 by relying on Phenom's audit should contribute their data to Phenom's bias audit. To do so, employers should reach out to their Phenom point of contact, typically the Customer Value Manager or Account Manager. Employers who do not need to comply with LL 144 are also welcome to contribute their data to ensure that their talent pools are included for a maximally comprehensive, generalizable, and reliable audit.

Alternatively, a Phenom customer may perform their own audit of Fit Score exclusive to their jobs and applications; Phenom can support a customer's auditor by providing the outputs of Fit Score for the customer's job applications, information on the operation of Fit Score, and our understanding of best practices in auditing Fit Score.

The customer can distribute the outcome of either Phenom's audit or the customer's audit on the customer's career site. For example, customers can create a page in their Phenom CMS to host the audit summary and link to the audit summary page from all job detail page templates.

How can employers use Fit Score while complying with other regulations?

The regulatory landscape for Fit Score is evolving, with potentially relevant regulations arising in jurisdictions in the USA, the EU, and elsewhere. Phenom tracks these regulatory efforts to determine whether a given regulation applies to Phenom software, whether it creates new obligations for Phenom, or whether Phenom can support customers in complying with the law. We then publish guidance to customers on these issues. Phenom customers can refer to <https://community.phenom.com> for regulation-specific guidance.

Has an AI risk assessment been conducted for Fit Score?

Yes. Phenom conducts risk assessments following its Governance Policy for AI Technologies. This policy is based on the **Model Artificial Intelligence Governance Framework** originally developed by the government of Singapore for the World Economic Forum, and it is similar to other AI governance approaches, such as the **Google AI Principles**.

The Governance Policy states that a risk assessment should consider the probability that the AI may lead to harm, the potential severity of harm, and strategies for mitigation.

According to the risk assessment, Fit Score, when used according to best practices, has a low probability of harm and a low severity of harm. The risk assessment suggests monitoring for adverse impact as one approach to harm mitigation. This audit report implements this suggested harm mitigation approach.

Validity and Reliability Motivation and Methods

What is validity?

Validity refers to the degree to which an assessment or tool measures what it claims to measure, and whether the measurement is appropriate for the intended purpose. In the context of Fit Score, validity ensures that the system aligns with job-related criteria and supports fair, defensible decisions.

Under the guidelines of the Equal Employment Opportunity Commission (EEOC), validity can demonstrate that a tool is job-related and consistent with business necessity.

Different forms of validity ensure that tools such as Fit Score are accurate and useful.

Criterion-related validity evaluates whether the tool accurately predicts its intended employment outcomes, making it both the most essential and the most challenging form of validity to establish.

What are validation studies, and why are they necessary?

Validation studies are formal investigations that assess whether a hiring tool is effective. They establish a scientific basis for using the tool and demonstrate its alignment with job-related criteria. For Fit Score, validation studies are necessary to:

- **Support compliance with employment standards, such as the EEOC guidelines**
- **Provide evidence that the tool supports fair hiring practices**
- **Maintain transparency and accountability in talent acquisition**

Validation studies also help organizations monitor the ongoing performance of AI systems, ensuring they adapt to changing job requirements and workforce trends.

How was the validation study conducted?

Phenom's validation strategy for Fit Score centers on criterion validation, which evaluates the relationship between Fit Score as a predictor and hiring progression as the outcome.

Specifically, the Hiring Progress Indicator provides the basis for the criterion variable. It is a version of the hiring status from an Applicant Tracking System (ATS), further standardized by a dedicated machine learning model. The dichotomous criterion variable is defined as whether a job application progressed from recruiter review versus did not progress.

This definition of the criterion reflects the fact that Fit Score operates at scale, evaluating job applications across diverse job roles.

Conventional validity studies may use job performance or employee retention as criteria, but validation of Fit Score requires a more robust approach. Our validation methods account for the broad application of Fit Score and the rich data that are available on the Phenom platform:

Traditional criteria only apply to a small set of jobs, while Fit Score applies to practically all job families in the labor market that use résumés and natural-language job descriptions. Our validation approach scales to all Fit Score outputs that can be compared to a hiring status, regardless of job family.

Traditional criteria represent events that occur long after the use of Fit Score in the hiring process. It is preferable to use events that are more closely linked to the use of an assessment instrument, which is possible using Phenom's modern digital platform.

Traditional criteria fail to measure how an assessment instrument performs on all of the job applications where it is used. For example, the criterion of employee retention can validate an assessment instrument on people who are hired and retained vs. hired and not retained. By contrast, Fit Score is used early in the selection process, including for applications that do not lead to a hire or even an interview. It is more comprehensive to validate Fit Score on all applications where a hiring status can be used to validate, rather than only on those applications that lead to a hire.

A key challenge in validating Fit Score is criterion contamination. As recruiters can view Fit Score Grades and may factor them into hiring decisions, validity coefficients may be inflated. To mitigate this, the analysis includes a statistical control for recruiter usage of Fit Score Grade, isolating the relationship between Fit Score Value and hiring progression. Logistic regression modeling is used to quantify Fit Score's predictive validity. This method estimates the probability of hiring progression as a function of Fit Score Value, returning validity coefficients in the form of odds ratios. Odds ratios above 1.00 indicate positive prediction, with values from 1.25–1.75 reflecting small to medium effect sizes typical for psychological predictors, and values above 2.00 indicating strong prediction.

What is reliability?

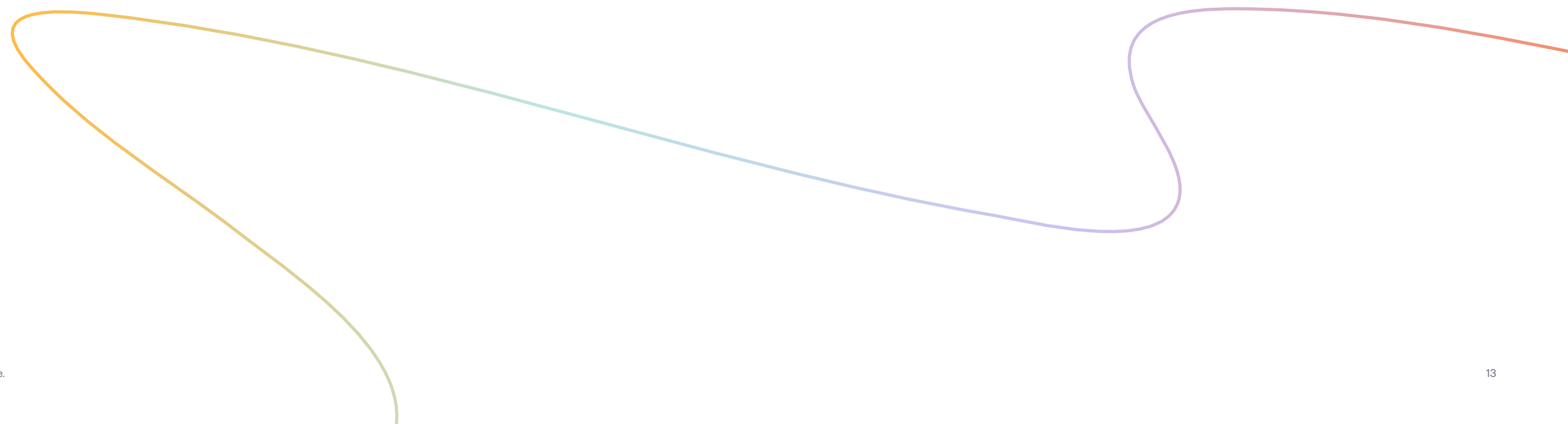
Reliability is a foundational element of any assessment tool, as it underpins confidence in validity metrics. Unlike traditional assessments, Fit Score adapts dynamically to the specific characteristics of each job, posing unique challenges for measuring its reliability using traditional methods. Standard measures (e.g., Cronbach's alpha, test-retest, and alternative forms) are not directly applicable to an AI-driven, process-oriented model like Fit Score, which generates scores tailored to each job.

How was the reliability study conducted?

As an alternative to traditional reliability tests, we used a wave analysis approach to assess the stability of Fit Score across different applicant cohorts during a job's hiring period. This method examines whether Fit Score maintains consistent computation of applicant-job alignment. Specifically, this analysis reveals the degree of differences in mean Fit Score among early, middle, and late applications after a job is open to candidates.

Applicant waves were defined by categorizing applicants as early (first 33% of applicants), middle (second 33% of applicants), or late (remaining applicants) based on the order in which they applied to a job.

Validity and Reliability Results



What data sets were used in the validity and reliability analysis?

We used candidate application data collected from January 1, 2024 to December 31, 2024 for this analysis. Data was sourced from 57 customers who were actively using Fit Score in their hiring process.

What were the reliability results?

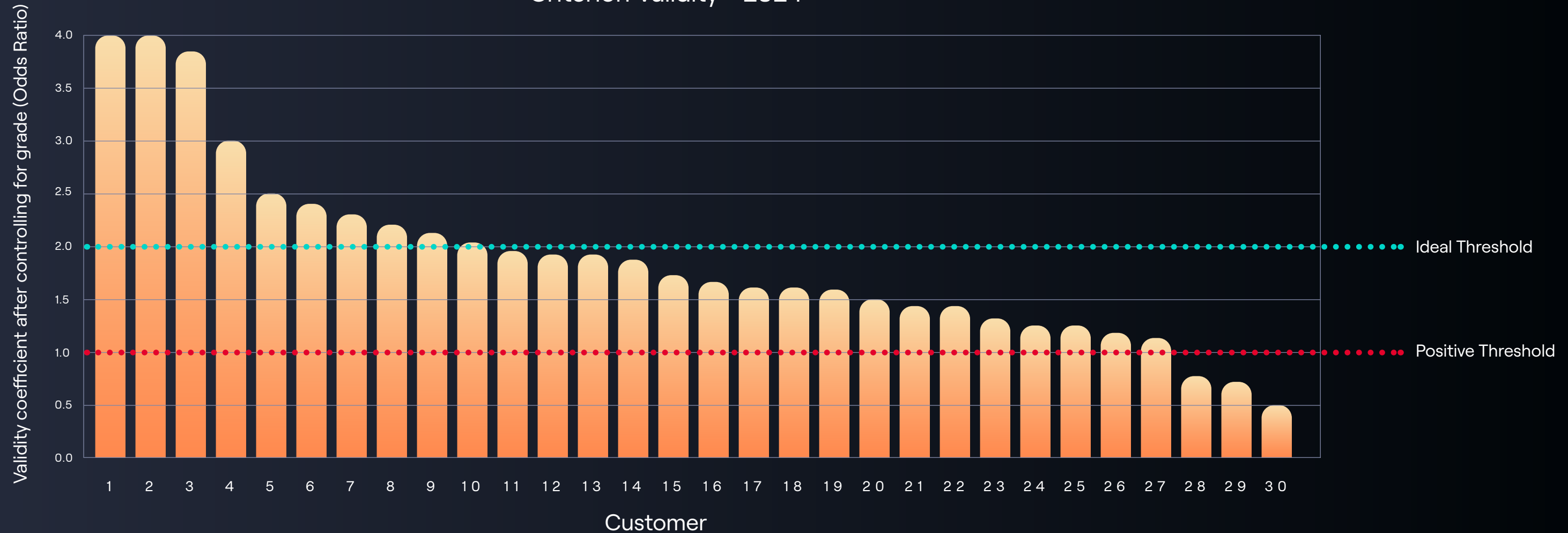


Mean difference tests using linear regression models were run to identify large differences in Fit Score between early, middle, and late application waves. Due to the large data volume, the effect size of mean differences was given more weight for drawing conclusions than meeting significance thresholds (p-value < .05). We used Eta squared as a measure of effect size and considered values below .01 to be well enough below what would be considered a “small effect”, i.e., the mean difference is not meaningful, indicative of Fit Score stability over time.

Fit Score was found to be reliable across several customers, showcasing the stability of Fit Score over time. In all but one case, mean differences in Fit Score between early, middle, and late waves were found to not be meaningful. As such, subsequent validity analyses can be interpreted with confidence given that the reliability of Fit Score has been established.

What were the validity results?

Criterion Validity - 2024



Among customers whose data produced statistically significant model results and for whom we were able to produce a Hiring Progression Indicator, Fit Score was found to positively predict the hiring progress criteria for 27 out of 30 customers. On average, applications with higher Fit Score values were found to be 2.2 times more likely to progress than those with lower Fit Score values after accounting for differences between letter grades.

That is, this finding is not merely that applications with a grade of “A” were more likely to progress than applications with a grade of “B”. Rather, this finding was true within letter grades (e.g., applications with a higher B score were more likely to progress than those with a lower B score).

This result indicates that Fit Score has strong criterion validity for its use case.

Analysis Conclusion

Fit Score was found to be a reliable and valid instrument to use in the hiring process. The above statistical analyses speak to Fit Score's strong psychometric properties. Establishing an instrument's fairness requires further examination, such as through adverse impact analysis, as discussed below.

Bias Audit Motivation and Methods

What are bias and adverse impact?

Adverse impact refers to the tendency of certain policies, practices, or procedures to disproportionately and negatively affect certain groups of people, based on their protected class. The legal definition of adverse impact is a “**substantially different rate of selection in hiring, promotion, or other employment decision which works to the disadvantage of members of a race, sex, or ethnic group**.” Adverse impact reports are commonly used in legal inquiries on employee selection.

Adverse impact can occur even when the policies or practices in question are not intended to discriminate. Adverse impact examines the consequences of policies and practices, not their intent.

The term “**bias**” has **many definitions**, including in cognitive psychology, statistics, computer science, and industrial/organizational psychology. For example, the American Psychological Association (APA) Principles for the Validation and Use of Personnel Selection Procedures defines bias as “systematic error in a test score that differentially affects the performance of different groups of test takers”. In this definition, bias relates to the tool and score itself. From the perspective of computer science and software engineering, Phenom Fit Score development addresses bias by only using job-relevant variables and developing our model on broadly representative data.

Scientifically, bias and adverse impact are distinct concepts, but adverse impact analyses provide a practical way to think about bias. For example, selection rates and impact ratios are the foundation of the “4/5ths rule”, which is one way to measure adverse impact. Selection rates, scoring rates, and impact ratios are also how New York City Local Law 144 defines its “bias audit” formula.

What is a bias audit, and why is it necessary?

A bias audit is a process of examining and evaluating data, algorithms, or systems for potential biases that could negatively impact certain groups or individuals. The goal of a bias audit is to identify unfair or discriminatory practices in a particular system or process.

Bias audits are necessary because they help ensure that systems and processes are fair for all users, regardless of their demographic characteristics. Without a bias audit, biases may go unnoticed. After an audit, corrective actions may be taken to mitigate bias.

If Fit Score does not use protected class information in its design, why is an audit desired?

By conducting a bias audit, we hope to demonstrate Phenom's commitment to fairness and to help build trust and confidence in our systems. Given Fit Score's design, the probability of bias is very slim. Still, an adverse impact analysis provides concrete evidence.

In addition, by conducting a bias audit, we aim to help our customers comply with regulatory requirements. Many hiring practices in U.S. organizations are subject to regulation at the federal level (i.e., EEOC) or from state or local governments. These regulations call for identifying and mitigating bias in hiring decisions.

Does this audit satisfy the requirements for an independent auditor?

Yes. This audit was performed by Conductor AI, a third-party auditor that is independent of Phenom.

What data sets were used in preparing this audit report?

The report includes a subset of anonymized and de-identified data from Phenom customers who used the Phenom platform. It does not include data from all Phenom customers, or all jobs or applicants from any one customer. The sample represents Fit Score use cases appropriately, as the data covers multiple United States geographies, jobs, job families, employers, and talent pools. All data included in this report pertain only to applications in the United States between January 1, 2024 and December 31, 2024.

How were customer data and candidate data protected?

Confidentiality, data privacy, and integrity were central to this process. Customer information was never shared with third parties and all analyses and findings were determined based on non-identifiable applicant data from a broad pool of candidates. Our approach complies with Phenom policies on information security and data privacy.

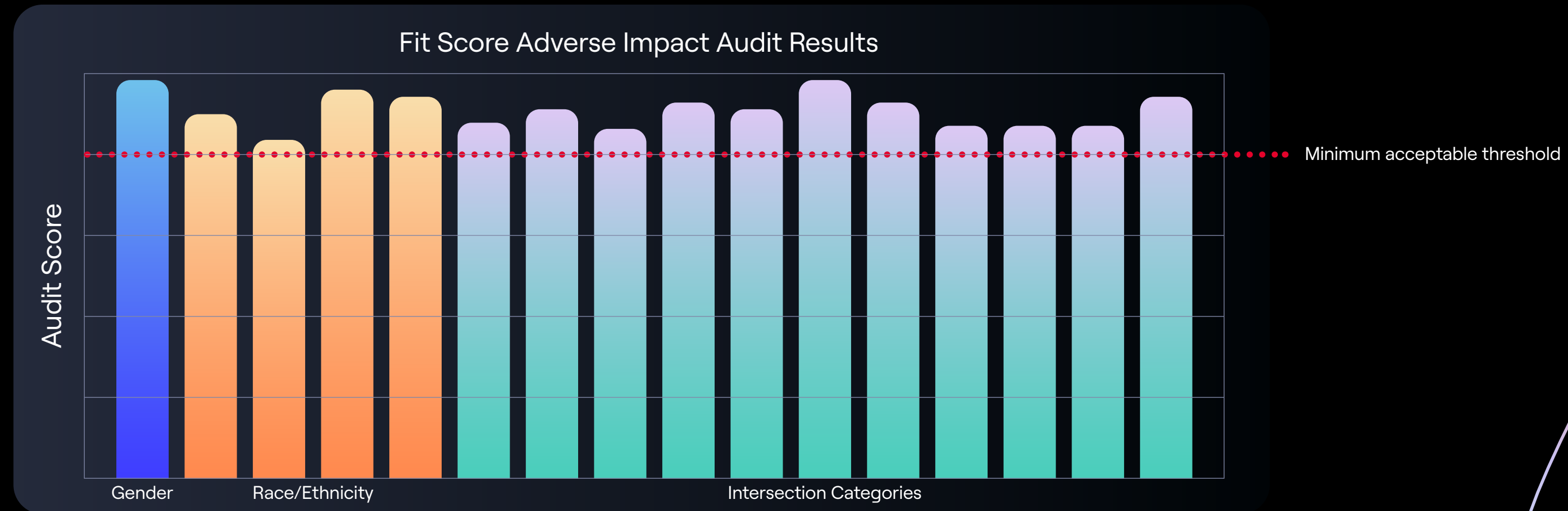
The auditor received information in a table having the following columns. Each row of data represented one job application, but the table had no application identifier.

Employer	Unique identifier for an employer (anonymized)
JOB_ID	Unique identifier for a job (anonymized)
Gender	Standardized gender variable based on self-identified gender on a job application
Race_Ethnicity	Standardized race/ethnicity variable based on self-identified race/ethnicity on a job application
Fit_Score_Grade	Fit Score letter grade
Fit_Score_Value	Fit Score numeric value

Bias Audit Results

What did the audit find?

The following discussion interpreting the audit findings is solely the responsibility of Phenom and not the third-party auditor. As is typical for LL 144 audits, the auditor's scope of work only included computing the relevant impact ratios, not interpreting them.



The results from a series of statistical tests comparing group scoring rates based on Fit Score of multiple protected classes of over nine million job applications showed no evidence of adverse impact. The audit found that Fit Score recommendations have no adverse impact on gender, no adverse impact on

race/ethnicity, and no adverse impact on gender and race/ethnicity intersectional demographic categories.

Based on this adverse impact analysis, we confirm that Fit Score is a fair selection tool for prioritizing candidates for interviews.

If I use Fit Score as part of my selection procedure, what is the risk of treating protected groups unfairly?

As Fit Score is based only on job-related characteristics, Phenom does not expect a direct influence of demographics on Fit Score. By design, all candidates, regardless of their demographic characteristics, are treated equally when their data is input into the Fit Score computation process. This adverse impact analysis confirms that the risk of treating protected groups unfairly by using Fit Score is low.

Can Fit Score help to reduce adverse impact in hiring?

Yes. Many conventional hiring practices can subject protected groups to unintentional unfair treatment. Fit Score promotes fair treatment by providing an objective evaluation of candidates based on job-related attributes such as skills and standardizing hiring criteria for a job for all candidates. This minimizes conscious and unconscious biases.

Do the findings from this bias audit generalize beyond LL 144?

Because different employers have their own recruiting and hiring workflows, Phenom customers differ in how they use Fit Score. A comprehensive adverse impact audit for an employer should be specific to the employer's jobs, candidates, hiring process, and role of Fit Score in that process.

That said, the findings in this report can be generalized. These findings provide useful insights as a starting point in understanding the use of Fit Score in many organizations and hiring processes. The applicability of these findings depends on various factors such as the size of an organization, types of jobs and job family classifications, the number of applicants and interviews in each job, the diversity of the applicant pool and workforce, and the selection processes aside from Fit Score.

What best practices should my company follow in using Fit Score to manage the risk of adverse impact in our hiring process?

Fit Score alone does not make hiring decisions. Phenom recommends that Fit Score should be used as one part of a hiring process, subject to human-in-the-loop control by a recruiter. (See [What are appropriate and inappropriate ways of using Fit Score?](#)) Here are additional best practices that an organization can adopt to further mitigate the risk of adverse impact:

Conduct a job analysis: Identify the essential functions of the job and the knowledge, skills, and abilities required to perform these functions. Use the “Edit Matching Criteria” function in the Phenom CRM to ensure that the matching criteria used by Fit Score are appropriate to the job.

Standardize selection processes: Evaluate candidates in a consistent way. Use the same interview procedures, questions, and scoring systems for all candidates for a given role.

Review job descriptions: Ensure they accurately reflect the essential functions of a job and its requirements. Avoid language that may discourage qualified individuals from applying. For example, [scientific research suggests](#) that the language of job descriptions can affect men and women differently, especially in majority-male occupations. Use the Phenom X+ job description generator to create tailored job descriptions and avoid biased language.

Use multiple selection criteria: Properly designed, multiple selection criteria and evaluation methods will converge on a holistic evaluation of a candidate. Depending on the type of occupation, helpful evaluation approaches may include general ability and personality tests, work sample tests, and structured interviews.

Train hiring process personnel: Train recruiters and hiring managers on the importance of using fair hiring practices, identifying and addressing potential bias, conducting interviews, and evaluating candidates.

Monitor your hiring practices for bias: Review your hiring practices to manage the risk of bias. This can include measuring adverse impact, reviewing selection criteria, and analyzing your interview questions.

Will Phenom produce audit reports in the future?

We release a large-scale audit report annually. We previously released bias audit reports in 2023 and 2024.

Resources and References

Empowering Recruiters with Phenom AI.

<https://community.phenom.com/kb/articles/11177-empowering-recruiters-with-phenom-ai>

Governance Policy for AI Technologies. Available on request from Phenom Product Knowledge Management.

Singapore Ministry for Communications and Information. 2020. Model artificial intelligence governance framework, 2nd ed. World Economic Forum. <https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai-sgmodelaigovframework2.pdf>

Society for Human Resource Management. (2017). What are disparate impact and disparate treatment? Retrieved from <https://www.shrm.org/resourcesandtools/tools-and-samples/hr-qa/pages/disparateimpactdisparatetreatment.aspx#:~:text=For%20example%2C%20testing%20all%20applicants,treatment%20is%20intentional%20employment%20discrimination.>

U.S. Equal Employment Opportunity Commission. (1978). Uniform guidelines on employee selection procedures. Federal Register, 43(166), 38290–38315. Retrieved from <https://www.govinfo.gov/content/pkg/CFR-2006-title29-vol4/xml/CFR-2006-title29-vol4-sec1607-16.xml>

American Psychological Association. (2018). Principles for the validation and use of personnel selection procedures (5th ed.). <https://www.apa.org/ed/accreditation/about/policies/personnel-selection-procedures.pdf>

Landers, R. N., & Behrend, T. S. (2023). Auditing the AI auditors: A framework for evaluating fairness and bias in high stakes AI predictive models. *American Psychologist*, 78(1), 36–49. <https://doi.org/10.1037/amp0000972>