HackerRank

Behind HackerRank's Certified Assessments:

The role-based technical tests you can trust

Table of Contents

Introduction 3
Identify the Role 4
Confirm Job-Related Skills 5
Develop Assessment Content 6
- **Key Components of Certified Assessments** 6
  - Skills 6
  - Items 7
  - Item Bundles 7
  - Test Sections 7
  - Tests 7
  - Scoring 7
- **Item Development and Scoring** 7
  - Coding Questions 8
  - Database Questions 8
  - Project Questions 9
  - Multiple-Choice Questions 9
- **Promote Fairness and Reduce Bias** 9
- **Independent Bias and Sensitivity Review** 9
- **Minimizing Adverse Impact** 10
- **Deploy Valid Assessments** 10
- **Protect Score Integrity** 11
- **Leaked Question Management** 12
- **Proctoring** 12
  - Copy/Paste Tracking 12
  - Tab Proctoring 13
  - Image Analysis 13
- **Plagiarism Detection** 13

Monitor Assessments 13
Conclusion 14
References 15

Behind HackerRank's Certified Assessments

About the author:

Jeff Facteau
Chief I/O Psychologist

Jeff Facteau received his Ph.D. in Industrial and Organizational Psychology from the University of Tennessee. He has over 25 years of experience advising organizations on the use of pre-employment tests in the hiring process. He is a fellow of the Society for Industrial and Organizational Psychology. At HackerRank, Jeff promotes assessment best practices with customers and in internal product development initiatives. He enjoys reading, music, getting outside, and having a good laugh with family and friends.

Introduction

Today, every company is either a tech company or is pursuing an aggressive transformation strategy to better compete in a digital marketplace. In either case, companies need a highly skilled developer workforce to out-innovate their competition. Skills have become the new currency in the labor market, and today's organizations understand the need for a robust strategy for acquiring and developing employees with the critical skills that are needed

Using technical assessments in the hiring process is an important component of this strategy.

Assessments help organizations efficiently screen large numbers of applicants based on their job-relevant skills, and they help eliminate the biases associated with resume-based screening. However, creating, standardizing, and maintaining technical skills assessments can be time-consuming and challenging. Most talent acquisition professionals are unfamiliar with the work that developers perform and the skills they need. These skills change and evolve quickly. There also can be a strong dependency on the goodwill and input of technical subject matter experts (SMEs) to help establish and maintain technical screening programs, which takes SMEs away from their day jobs helping drive the organization's innovation.

HackerRank's certified assessments make it easy for you to implement, standardize, and maintain technical testing programs. They are easy because HackerRank does much of the important but difficult work for you. We have done the research to identify the critical skills required for most technical roles. We have developed thousands of technical questions and challenges following a rigorous process that ensures they are valid and free from content that could introduce biases (e.g., gender, cultural, racial, ethnic, or socioeconomic).

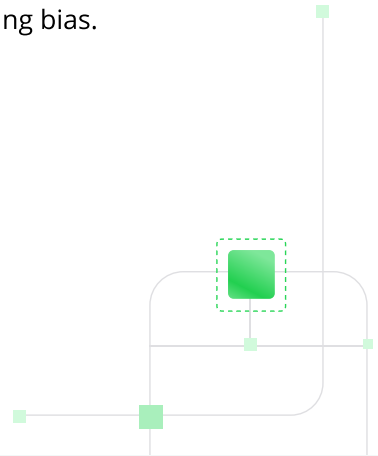
We also make it significantly easier for you to manage questions that get leaked or compromised—because we do it for you.

In this white paper, we demonstrate why you can trust HackerRank's certified assessments and how they make technical testing easier. This includes our work to create the most up-to-date and comprehensive skills taxonomy that provides the foundation for creating job-related technical tests. We discuss our rigorous approach to content development, scoring, and test creation to provide confidence in the validity and fairness of certified assessments.

We also discuss the many ways in which we help you maintain the integrity of technical assessment scores over time—not the least of which includes monitoring and replacing leaked questions on your behalf.

We have organized our presentation around the key tasks that must be addressed to create, implement, and maintain valid, legally defensible testing programs. These critical steps include:

1. Identifying the role for which an assessment is needed.
2. Confirming the job-related skill requirements for the role.
3. Developing assessment content.
4. Promoting fairness and mitigating bias.
5. Deploying valid assessments.
6. Protecting score integrity.
7. Monitoring assessment trends.

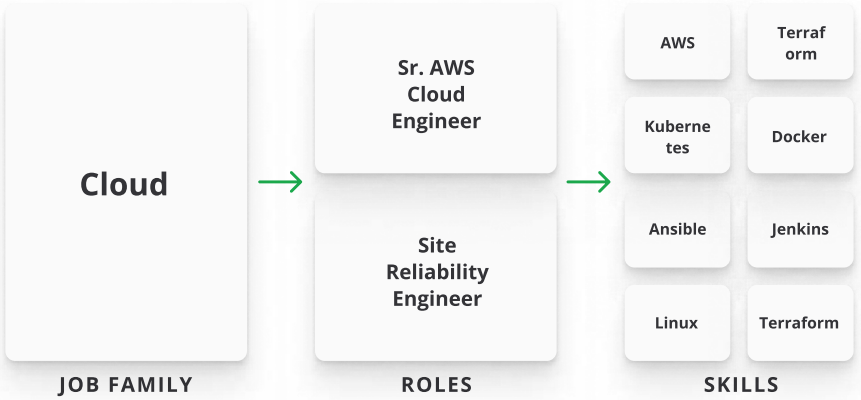


Identify the Role

HackerRank's certified assessments are based upon a comprehensive skills taxonomy that has been years in the making. The taxonomy incorporates three levels of analysis: job families, roles, and skills. It is an incredibly robust resource for identifying a specific job and the skills that are required to perform it.

Our skills taxonomy is based upon multiple inputs and data streams. These include guidance from our Skills Advisory Council, 10-plus years of skills data from our customers, and machine learning insights from 25,000 (and growing) technical job descriptions. These inputs have been assimilated by subject matter experts who have deep experience with technical roles. The result is a framework rooted in data and expert judgment that articulates the skills required for in-demand technical roles across nine job families, including:

Figure 1. Skills Taxonomy Structure



- Artificial Intelligence & Machine Learning
- Cloud
- Cybersecurity
- Data Engineering
- Data Science & Analytics
- Mobile
- Quality Assurance
- Software Engineering
- Web Development

Our skills taxonomy provides the foundation for identifying the skills required for nearly any technical role. By consulting it, you can easily identify the job that best matches your hiring needs and the technical skills that likely are required for success.

Confirm Job-Related Skills

Certified assessments are designed to measure the most important and foundational skills for any role in our skills taxonomy, making it easy for you to implement a standardized, job-related test in your hiring process. Next, we describe the way in which we developed our skills taxonomy to ensure you can trust that our certified assessments are measuring the right skills.

To identify the skills required for each role in our framework, we conducted a comprehensive and systematic analysis of publicly available job descriptions. This involved extracting skill information from job descriptions as well as combining the insights we gained with the judgment of experienced SMEs and the skills data we have collected over the past 10-plus years.

To begin, we trained and validated machine learning (ML) and natural language processing (NLP) models to extract skills data from job descriptions.

To begin, we trained and validated machine learning (ML) and natural language processing (NLP) models to extract skills data from job descriptions.

To ensure that our analysis was based on representative and relevant job descriptions, we sampled from across our customer base, our customers' competitors, and technology and tech-forward companies across industries.

We also employed a number of quality checks to ensure that the job descriptions we analyzed had the potential for contributing useful skills data to our analysis. In total, we analyzed over 25,000 unique job descriptions to build the latest version of our skills taxonomy.

Our models are trained to first differentiate technical from nontechnical jobs. They extract skill information from the sections of technical job descriptions that are most likely to convey skill requirements (e.g., duties, responsibilities, required qualifications, etc.). The models also have been trained to semantically differentiate skills from other information. This ensures that the resulting insights only reflect the skill expectations for a role and not degree or experience requirements.

Once skills data had been extracted from the individual job descriptions, the models conducted a mapping of the data to our existing skills taxonomy. This enabled us to identify technical skills that already were included in the taxonomy as well as those that were not. The latter reflect gap skills—opportunities to update our taxonomy to better represent the technical skills that employers are seeking today.

To identify skill requirements for a particular job (e.g., Full Stack Developer, Data Scientist), we clustered individual job descriptions by job title. To do this, we initially employed automated clustering techniques and then SMEs reviewed the clusters and job

descriptions comprising them to ensure their integrity, internal consistency, and interpretability.

Clustering job descriptions enabled us to aggregate the skill information extracted from the individual job descriptions within a cluster. For example, it allowed us to identify the technical skills most frequently sought by employers across job descriptions for the same role—and whether or not those skills were represented in our taxonomy.

To make updates, revisions, or additions to our skills taxonomy, we worked with external SMEs who assimilated all the data that was available to them: our existing taxonomy, data and insights from scraping tens of thousands of job descriptions, and 10-plus years of HackerRank skills data. For any given role, we worked with a number of experienced and carefully vetted SMEs. They combined these inputs with their deep knowledge and experience to help us define and prioritize the technical skills required for each job in our framework.

The result of all this work is an up-to-date and comprehensive framework that can be used with confidence for identifying and understanding role-based technical skill requirements. Our skills taxonomy provided the foundation for specifying the skills to measure in each of our certified assessments.

Develop Assessment Content

Validity refers to the degree to which inferences or decisions made on the basis of a test score are supported by available evidence (Binning & Barrett, 1989). For technical skills assessments, the inferences and decisions we make from test scores generally concern whether or not a particular applicant possesses the foundational skills that are needed to succeed in a particular job. The question of validity, then, concerns the degree to which we can have confidence that test scores are an accurate reflection of an applicant's job-related skills.

In this section, we discuss the steps that HackerRank takes to ensure that our certified assessments produce an accurate signal about an applicant's technical skills—and one that is free from irrelevant biases.

We begin with a brief description of the major components or building blocks of a certified assessment. Then, we describe how test items and challenges are developed and validated. Finally, we describe the logic and criteria that are applied to items when generating a role-based certified assessment for each customer based on our skills taxonomy.

Key Components of Certified Assessments

When a customer adds a certified assessment to their test library, a unique test is generated for them by drawing upon banks of items that have been identified in advance for measuring the skills associated with a role (based on our skills taxonomy). In this section, we briefly describe the components of certified assessments that allow us to generate them in this way.

Skills

These are the building blocks for certified assessments. A skill reflects the ability or proficiency to perform a task which is acquired either through learning or experience (c.f., Dunnette, 1976; Lubinski & Dawis, 1992). Example skills in the HackerRank taxonomy include SQL, RESTful API, HTML, and System Design, among many others. Each certified assessment is designed to measure multiple job-relevant skills.

Items

These are questions or challenges that are designed to measure a skill at a particular proficiency level. Some questions measure the underlying knowledge required to demonstrate a skill. Others require applicants to actually demonstrate the skill by successfully completing a task.

Item Bundles

An item bundle is a large bank, or pool, of items that measure a specific skill (e.g., REST API) at a particular level of proficiency (e.g., basic, intermediate, advanced). They are an integral component of certified assessments because they enable the generation of tests that include alternate, interchangeable items for measuring a particular skill. Item bundles are carefully curated to ensure that the items that comprise them are as similar as possible in terms of difficulty and the time required to complete them. For example, the items in a particular bundle are similar in terms of their level of difficulty, and items that take significantly more time to complete than others excluded. We also exclude extremely difficult items—those for which less than 2% of applicants achieve a full score.

Test Sections

When customers add a certified assessment to their account, a test containing multiple sections is added to their test library. A test section is a subset of 5-7 items that measure a particular skill and proficiency level. The items in the section are drawn from a corresponding item bundle using sophisticated logic that ensures they are as interchangeable as possible. This allows for items from within a section to be randomly administered to applicants in a way that ensures fairness.

Tests

A test consists of one or more sections, each of which is designed to measure a different skill or proficiency level. Each section includes multiple, interchangeable items for measuring the same skill and proficiency level.

Scoring

The items and challenges included in certified assessment are scored automatically and objectively. While the scoring algorithms for different types of items (e.g., coding, database, project) utilize different assets and resources (e.g., test cases, evaluation files, unit cases), in each case an applicant's response is evaluated in terms of whether or not it produces the outputs that are expected for a given set of inputs.

Item Development and Scoring

Certified assessments utilize skill-based items from HackerRank's proprietary library of content. The development of these items follows a rigorous process that is designed to ensure that each item provides a valid, unbiased, and fair signal about an applicant's proficiency level for a particular skill. The following process is used to develop and validate each item in HackerRank's library.

1. Subject matter experts who have deep experience and strong proficiency in a skill are utilized to create each item.
2. Item writers complete a comprehensive, standardized training program. They practice developing items, and receive feedback about their work. They are required to create sample questions which are reviewed and must be approved by experienced item writers and SMEs.
3. Trained item writers review the required specifications for a new item (i.e., skill, proficiency level, type of question) and then create draft questions. Drafts include a prompt, the assets applicants

need to answer the question (e.g., information, data tables, arrays, etc.), and test cases for scoring the item.

4. Item writers complete a comprehensive, standardized training program. They practice developing items, and receive feedback about their work. They are required to create sample questions which are reviewed and must be approved by experienced item writers and SMEs.
5. Trained item writers review the required specifications for a new item (i.e., skill, proficiency level, type of question) and then create draft questions. Drafts include a prompt, the assets applicants need to answer the question (e.g., information, data tables, arrays, etc.), and test cases for scoring the item.
6. An independent SME conducts a technical review of the draft item to ensure it is clear and easily understood, can be solved in the targeted amount of time, and measures the intended skill. Further development of the item only occurs when the independent reviewer agrees that the item measures the skill that is intended.
7. An independent reviewer with a background in computer science reviews the item to ensure it is clear, understandable, and grammatically correct according to American English language standards. They also ensure the item is free from content that could introduce bias (e.g., culturally specific references) or that could be construed as insensitive by some applicants.
8. For questions that permit applicants to choose which programming language they would like to use (e.g., Python, C/C++, Java, etc.), several SMEs attempt the item using different languages to ensure the item is language agnostic.
9. If changes or revisions are required at any step in the process, the item is revised and the review process recommences. Items that pass all steps in the process are deployed to HackerRank's content library where they are available for customers to include in their tests.
10. The performance of the item (e.g., difficulty, time) is monitored over time and any feedback from applicants or customers is reviewed. If needed, items are revised and the review process is need to answer the question (e.g., information, data tables, arrays, etc.), and test cases for scoring the item.

There are four types of items that can appear in certified assessments. A brief description of each type of item and how they are scored appears next.

Coding Questions

A HackerRank coding question requires an applicant to submit code in any of the available languages to solve a specific task. Coding questions are scored using test cases, which are designed to evaluate the correctness and efficiency of a candidate's code. Test cases are a set of specified inputs and logical, expected outputs with a given execution condition (e.g., maximum run-time or memory required), and they provide the basis for automated evaluation of a candidate's code. Once a candidate submits their code, it is run against all the test cases in a standard execution environment and condition. The output from the candidate's code is compared with the expected output to see whether the test case has passed or failed. The score on a coding question is the sum of the scores of the individual test cases.

Database Questions

Database questions require applicants to query and manipulate data from one or more databases. These questions are automatically scored by comparing the data retrieved by an applicant to the correct answer in the form of an evaluation file. If the applicant's data file matches the evaluation file, a full score is awarded to the applicant.

Project Questions

Project questions simulate the actual work environment and require applicants to debug and/or enhance an existing feature based on their understanding of the code base. Alternatively, they may ask applicants to create new features based on product specifications. For example, a project question could require an applicant to build a web-based tool for looking up the time and date of an event based on inputs provided by a user. Project questions are automatically scored using scripts, unit tests, evaluation files, or test cases that compare an applicant's output or results to what is expected given the projects' requirements and specifications.

Multiple-Choice Questions

Multiple-choice items are dichotomously scored. When a candidate chooses the correct response, they receive full points for the question. If they select an incorrect response, they receive zero points for the question. If a multiple-choice item has more than one correct response (i.e., select all that apply), a candidate may receive partial credit based on the total number of the correct responses selected. We should note that over time we intend to use multiple-choice questions as little as possible in certified assessments. This is because it can be difficult to detect when applicants have used assistance from artificial intelligence (such as ChatGPT) to answer these types of questions.

To summarize, we follow a systematic and rigorous approach to developing certified assessment items. The process ensures the validity of the items by confirming that they measure the skills that are intended. It also ensures that items are clear, easy to understand, and can be completed in the time allotted—using whatever programming language the applicant prefers. Finally, the process minimizes the possibility for unintentional biases to influence the evaluation process, enabling organizations to hire for the skills they need.

Promote Fairness and Reduce Bias

First, we engage with an independent consultancy that specializes in helping organizations create fair and inclusive environments by ensuring the content they produce is free from bias. They conduct a systematic review of certified assessment items using a standardized rubric. Three reviewers independently examine the items to ensure the following criteria are met:

For certified assessments, we take two important steps that go above and beyond our rigorous content development processes to promote fairness, reduce bias, and minimize adverse impact. These are described below.

Independent Bias and Sensitivity Review

First, we engage with an independent consultancy that specializes in helping organizations create fair and inclusive environments by ensuring the content they produce is free from bias. They conduct a systematic review of certified assessment items using a standardized rubric. Three reviewers independently examine the items to ensure the following criteria are met:

- Items should include language that has a common, global meaning for all applicants.

- Where proper nouns are used, they should refer to common, globally recognized names and places.
- Acronyms should be used appropriately (if at all).
- First names should be used that are gender neutral across cultures (e.g., Alex, Blake).
- Pronouns should be avoided in favor of role titles (e.g., supervisor, coworker
- Items should be free of content that makes generalizations or assumptions about individuals based on their membership in a particular group (i.e., stereotypes).

When the reviews are completed, HackerRank's subject matter experts review the feedback and implement changes to address it. These adjustments are determined on a case-by-case basis, depending upon the feedback. Usually they involve minor modifications to phrasing (e.g., avoiding verbs ending in "-ing,") and grammar, replacing pronouns with role titles (e.g., "the employee" vs. "they"), and replacing gendered first names with gender-neutral alternatives. In each case our primary goals are to address the feedback provided and maintain the underlying properties and characteristics of the item (i.e., skill measured, time required, difficulty).

Minimizing Adverse Impact

Before releasing each certified assessment for general use, we conduct an analysis to evaluate demographic differences in scores (e.g., based on gender). As data accumulates, we evaluate differences in item scores as well as overall test scores. We use this data to eliminate items that disproportionately contribute to group differences in overall test scores.

We also evaluate the impact of setting cut-off scores at increasingly stringent percentiles (e.g., 30th, 50th, 75th, 90th) on group differences in passing rates.

We provide this information to our customers so that they can make informed decisions about cut-off scores that balance their needs to be selective as well as minimize adverse impact on the groups they are aggressively recruiting to diversify their technical workforce.

Deploy Valid Assessments

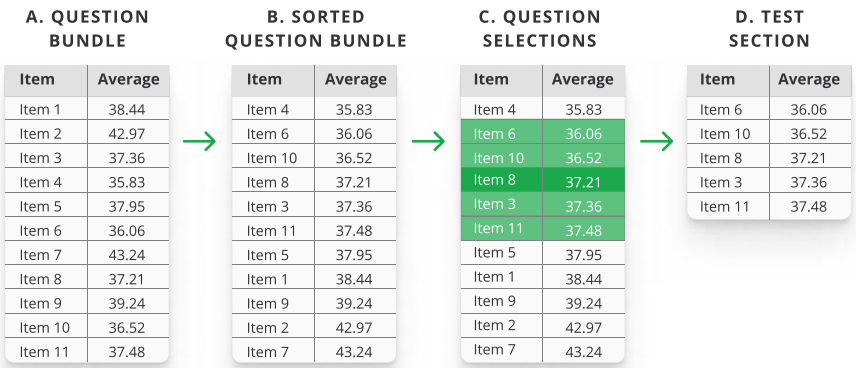
When a customer adds a certified assessment for a particular role to their HackerRank account, a test is dynamically generated using the item bundles mapped to the relevant skills from our skills taxonomy. Each certified assessment takes 60-100 minutes for applicants to complete, with longer times required for more senior jobs. Each test includes multiple sections that are each designed to measure one of the priority skills identified in the taxonomy (e.g., Intermediate SQL). Items within each section are selected from a corresponding item bundle. Once a certified assessment is created, it is ready to be administered to applicants, and test invitations can be initiated from within our platform or within an integrated applicant tracking system.

To illustrate the test generation process, when a customer adds a certified assessment for the Software Engineering Intern job to their test library, a test is created that includes sections for measuring Basic Problem Solving and Basic SQL, and the questions for each section are drawn from their respective item bundles. The process for ensuring that the items within each test section are interchangeable is depicted in Figure 2. First, the items in a bundle (A) are sorted by their difficulty (B). Next our item selection algorithm randomly chooses the first item for the test section.

In the figure, this is Item 8 in (C), which is highlighted in dark green. Next, the algorithm chooses the remaining items for the section that

that are closest in difficulty to the initial item. In the figure, these items are highlighted in light green and include Items 6, 10, 3, and 11. Based on this process, the test section that is generated (D) is comprised of items 6, 10, 8, 3, and 11. The same process is used to generate other sections of the test.

Figure 1. Skills Taxonomy Structure



There are a number of benefits that derive from this approach to certified assessment test creation. First, it ensures that the test measures the technical skills that are most important for the role as defined by our skills taxonomy. Second, it ensures that each test includes alternative items for measuring the same skill, so that questions measuring a particular skill can be randomized. This reduces the opportunity for applicants to collaborate on the same item while also maintaining fairness.

Third, our approach ensures that the certified assessments used by any two organizations will differ significantly in the specific items that are administered while measuring the same job-relevant skills and proficiency levels. Finally, it reduces the number of times a given question is administered across all applicants, which correspondingly reduces the probability that the question will be leaked.

To this point, we have described the rigorous process we follow to ensure the validity of certified assessments. There also is substantial peer-reviewed academic research that confirms the validity of tests like certified assessments that require applicants to demonstrate their job-related knowledge and ability to perform job-related tasks. These are known, respectively, as job-knowledge and work sample tests, and they are among the most predictive tools available for forecasting an applicant's success on the job. In a recent and definitive meta-analysis (a quantitative summary of individual research studies) conducted by Sackett, Zhang, Berry, and Lievens (2021), job knowledge and work sample tests were ranked second and fourth out of 24 different types of predictors of future job performance, in terms of their estimated validity. In addition, job applicants generally view work sample tests more positively than other types of assessments (e.g., Hattrup & Schmitt, 1990; Hausknecht, Day, & Thomas, 2004) because they represent the actual work to be performed. Taken together, these studies provide strong support for using HackerRank's certified assessments in your technical hiring programs.

Protect Score Integrity

Users of online technical tests are, unfortunately, well aware that some applicants will try to obtain inappropriate assistance when completing a test. Sometimes assistance comes in the form of answers to actual test questions that have been leaked on the internet. Applicants also may search the internet for guidance that helps them answer a question (e.g., example code). Finally, others may go so far as enlisting another person to sit with them and help while they take the test, or even take the test for them.

Regardless of how assistance is obtained, the result is that it calls into question the integrity of a person's test score.

Even worse, if stakeholders aren't confident that adequate precautions have been taken, it can erode their overall confidence in an assessment program. Once their confidence is lost, test adoption may decline, creating the temptation for hiring managers to implement idiosyncratic assessments that put the organization at risk. For these reasons, safeguards and preventative measures are often needed to significantly reduce applicant's opportunities to benefit from assistance. HackerRank provides a number of features for certified assessments (and other tests) that provide these safeguards and protect the integrity of your technical tests.

Leaked Question Management

Technical test questions will be leaked to the internet. There is little uncertainty about this, if any. Until now, organizations had to devote significant time to managing question leakage.

Technical SMEs would have to review leaked questions, identify alternatives in HackerRank's library, and manually replace them in multiple tests. Approval processes often required additional time from SMEs, IO Psychologists, and talent acquisition professionals. Simply put, managing leaked questions is a burdensome distraction that cannot be avoided with technical skills testing—but HackerRank manages this entire process for you.

For certified assessments, you no longer have to manage leaked questions. HackerRank takes on the burden of managing leaked questions, so you can stay focused on your mission critical priorities (like driving innovation) instead of tracking down and managing SMEs to help you find replacements for leaked questions.

Our process for managing leaked questions is simple. We actively scour the internet on a weekly basis for questions that have been leaked. When we find one, we remove it from your certified assessment and replace it with an interchangeable item using the criteria described earlier (i.e., similar difficulty and time to complete). At the same time, we pursue our rights under the Digital Millennium Copyright Act (DMCA) to protect our copyrighted material. Whenever our content is found on websites or within online communities, we issue notices to have it taken down. Once we "recover" an item in this way, we return it to our library where it once again becomes eligible to be included in a certified assessment.

In most cases, once a leaked item is detected we will replace it within 24 hours. In some instances (e.g,. more complex, project-based questions) this process may take longer, but usually no more than 15 to 30 days.

Proctoring

HackerRank offers a number of proctoring features that are designed to protect the integrity of technical test scores. They are designed to achieve two primary goals when it comes to test integrity. The first is to prevent dishonest test-taking behaviors by acting as a deterrent. Applicants who know proctoring is in place are less likely to engage in such activity. The second is to capture and record data points or signals that support the detection and communication of potential dishonesty to the consumers of test scores (e.g., recruiters, hiring managers) so that they can make well-informed decisions. Each of our proctoring features is described below.

Copy/Paste Tracking

If an applicant pastes code into their response that has been copied from another source, the pasted code can be viewed in their score report. In addition, you can view that pasted code in the candidate's test report. Also, if the candidate tries to paste the code from other sources, it will be captured in our plagiarism model. Finally, if desired, applicants can be notified

through a pop-up window that their copy/paste activity is being monitored.

Tab Proctoring

This option enables you to monitor if the applicant switches between browser tabs during the tests. When an applicant attempts to open another tab, a warning message is displayed informing them that the test is being proctored, and if they continue, it will be highlighted to the organization administering the test. This discourages them from utilizing information or resources from outside the HackerRank platform when completing their test.

Image Analysis

Enabling this feature helps you verify the identity of an applicant as well as monitor whether the same person is present throughout the test's duration. This is done by taking periodic snapshots of the applicant via their webcam. These images can then be manually reviewed to verify the applicant's identity and confirm that only they were present during the test session.

Plagiarism Detection

In addition to proctoring tools, the integrity of an assessment also relies on detecting when applicants have used material from other sources. The current industry standard for plagiarism detection relies heavily on MOSS code similarity. Not only does this approach often lead to higher false positives rates, but it is also unable to reliably detect plagiarism originating from conversational AI or large language models. This is because conversational AI can produce original code, which circumvents similarity tests.

Instead of relying on MOSS, HackerRank uses a machine-learning based plagiarism detection model to characterize coding patterns and check for plagiarism based on a number of signals. The model also uses self-learning to analyze past data points and continuously improve its confidence levels. The result is an ML-based plagiarism detection system that is three times more accurate at detecting plagiarism than traditional code similarity approaches—and can detect the use of external tools such as conversational AI. This dramatically reduces the number of false positives and better promotes a fair and equitable testing experience for all applicants.

Monitor Assessments

An important activity for ensuring the effectiveness of any pre-employment testing program is to monitor a variety of process and outcome metrics over time and use the insights to make improvements. As more applicants complete a test, these indicators reveal opportunities to make adjustments to the test itself, how it is used, and potentially even the broader recruiting and selection process.

HackerRank monitors a variety of indicators and metrics to ensure our certified assessments perform as intended over time. These metrics provide insight into how effective certified assessments are at differentiating among applicants based on their proficiency levels and in a way that is fair regardless of their demographic background. They also provide us with a mechanism to evaluate the quality of the applicant's experience and identify opportunities for improvement. Some of the specific metrics we monitor include the following:

- Invitation to test attempt rates
- Test completion rates

- Test abandonment patterns
- Time to complete tests
- Applicant satisfaction ratings
- Qualitative feedback from applicants
- Item and test statistics (average score, standard deviation, percentiles, etc.)
- Group difference in item and test scores

Based on our analyses, and the insights we gain, we make adjustments to certified assessments to improve their validity, usefulness, and fairness. In addition, working with our Customer Success team, you can track many of these metrics for your specific assessment program, so you can make refinements and adjustments to ensure your technical hiring goals and objectives are being met.

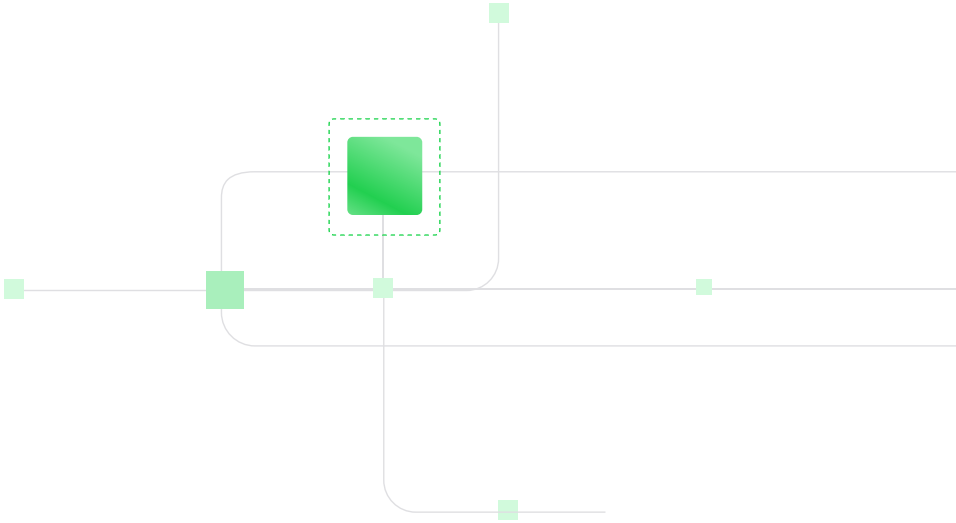
Finally, we strongly encourage customers to monitor their assessment programs for adverse impact. This involves comparing hiring rates (i.e., selection ratios) for members of different demographic groups (primarily gender and race/ethnicity) and determining if there are significant and practical differences between them. This is a critical activity to conduct because anti-discrimination and equal employment opportunity laws around the world place the burden of preventing discrimination on employers.

Of course, if you need assistance or guidance with monitoring adverse impact in your organization, our experienced team of I/O Psychologists and Engagement Managers are ready to help. Please contact your Customer Success Manager if you would like to discuss these or other services.

Conclusion

HackerRank's certified assessments provide you with an easy way to implement skill-based assessments you can trust in your technical hiring process. They are easy to use because we have taken on the heavy lifting to identify the job-related technical skills required for nearly any role, assemble valid and fair assessments that measure those skills, and maintain them overtime. You can trust them because of the rigor behind our approach to building certified assessments, which we have transparently described in detail here.

In closing, we should note that our Professional Services team (i.e., I/O Psychologists, Engagement Managers) is well-equipped and ready to help with any additional steps you would like to take as you implement certified assessments. In combination, they possess decades of experience conducting job analyses and local validation studies, determining appropriate cut-off scores, and evaluating and minimizing adverse impact. They are prepared and excited to be your trusted partners as you implement HackerRank's certified assessments.



References

Binning, J. F., & Barrett, G. V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. Journal of Applied Psychology, 74(3), 478–494. https://doi.org/10.1037/0021-9010.74.3.478

Dunnette, M.D. (1976). Aptitudes, abilities, and skills. In M.D. Dunnette (ed.), Handbook of industrial and organizational psychology (pp. 478-483). Rand McNally.

Lubinski, D. & Dawis, R.V. (1992). Aptitudes, skills, and proficiencies. In M.D. Dunnette & L.M. Hough (eds.), Handbook of industrial and organizational psychology, 2nd ed., Vol 3. (pp. 1-59). Consulting Psychologists Press.

Hattrup, K., & Schmitt, N. (1990). Prediction of trades apprentices' performance on job sample criteria. Personnel Psychology, 43(3), 453–466. https://doi.org/10.1111/j.1744-6570.1990.tb02392.x

Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. Personnel Psychology, 57(3), 639–683. https://doi.org/10.1111/j.1744-6570.2004.00003.x

Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2022). Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range. Journal of Applied Psychology, 107(11), 2040–2068. https://doi.org/10.1037/apl0000994