

*Standards and
Assessment Guide*
Software Testing

Introduction

Welcome to the Standards and Assessment Guide (SAG), a comprehensive resource designed to support excellence in assessment practices for all skills competitions hosted at the WorldSkills Competition. A key component of the SAG is the commitment to continuous improvement through ongoing review, ensuring that these guides consistently reflect the latest industry innovations and high standards of assessment in an open and transparent manner.

Purpose

The primary purpose of this guide is to provide a structured framework for the consistent and fair assessment of each skill competition. It aims to ensure that all assessments are conducted with a high degree of accuracy, reliability, validity, and reflect global industry standards. By adhering to these guidelines, assessors can maintain the integrity of the competition and uphold the high standards associated with WorldSkills.

Note: this guide is not a replacement for the Marking Scheme and should only be used as a reference when required during the assessment process. Competitors are not allowed to have a copy of this SAG in the workshop in either digital or paper form.

Scope

This document covers general assessment practices applicable to a diverse range of industry sectors including:

- Manufacturing and Engineering Technology
- Information and Communication Technology
- Construction and Building Technology
- Transportation and Logistics
- Social and Personal Services
- Creative Arts and Fashion

Each skill competition has its unique requirements and benchmarks, which are detailed within sections of this guide. The consistency of assessment across each skill is the overarching principle, ensuring a unified approach to evaluating Competitor performance to the highest standards.

Key Components

The guide is structured into several key components, each addressing critical aspects of the assessment process:

Assessment Principles and Objectives

- Emphasis on fairness, transparency, and objectivity in assessments.
- Alignment with global industry standards and best practices.

Assessment Methods and Tools

- Description of various assessment methods in Measurement and Judgement.
- Guidance on the application of appropriate assessment tools and equipment (if applicable).

Criteria and Benchmarks

- Examples of assessment criteria and aspects for each skill, outlining the expected competencies and performance standards.
- Benchmarks for different levels of proficiency i.e. Judgement award range 0, 1, 2, 3 or Yes/No criteria alongside clear measurable descriptors.

Expert assessment requirements

- Completion of the Access Programme.
- Approved Curriculum Vitae (CV).
- Reaching **100%** Expert preparedness.
- Full participation of online and in-person Mandatory Assessment Training (MAT).
- Completion of practical assessment testing at the Competition.

Quality Assurance and Improvement

- Monitoring and evaluation: Ensuring assessments are carried out correctly and produce reliable results.
- Continuous improvement: Using feedback from Experts and Skill Competition Managers to enhance future assessments.
- Periodic review: Regularly updating processes to reflect current industry and WorldSkills standards.

Implementation

For effective implementation, this guide should be used for guidance purposes only and in parallel with the Technical Description for each skill. Experts are encouraged to familiarize themselves thoroughly with both the general guidelines and the particular requirements of their respective skill areas. Collaborative efforts between the Skill Management Team, Experts, and Competitors are vital for achieving the ultimate goal of global skills development.

Conclusion

The Standards and Assessment Guide embodies the commitment of the WorldSkills movement to nurturing talent and promoting the highest standards of vocational education and training worldwide. By providing clear and comprehensive guidance, we aim to empower Experts to conduct marking and assessment practices that are not only rigorous and equitable but also inspiring and transformative for all participants.

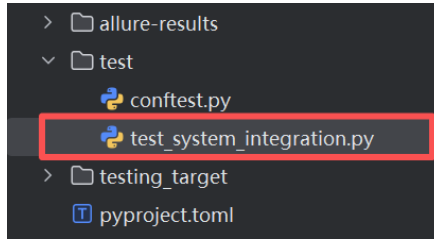
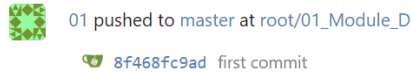
Thank you for your dedication to upholding these standards and contributing to the success of the WorldSkills Competition. Together, we can continue to champion skills excellence and celebrate the achievements of talented individuals from around the globe.

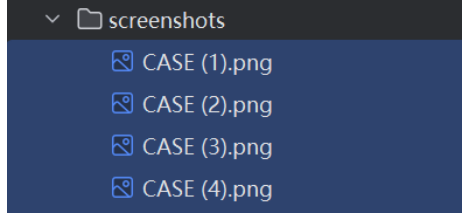
Skill 11

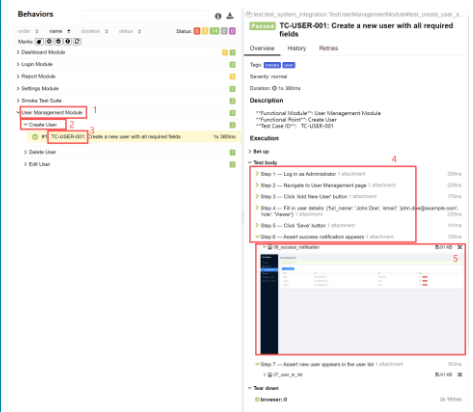
Measurement

Measurement is used to assess accuracy, precision, and other performance that can be measured objectively. It is used where ambiguity must be avoided.

The total marks allocated to measurement marking may vary from competition to competition depending on the Test Project.

Aspect	WSOS Section (as per TD)	Descriptor	Requirement	Marks	Resource
M	1	File naming and storage requirements are assessed based on objective compliance with the specified file structure and naming conventions.	The integration test module is stored using the exact required file name: test_system_integration.py. No alternative naming formats are accepted.	0.5	 Full marks are awarded if the file exists; zero marks will be given otherwise.
M	2	File upload is assessed by verifying the presence and location of required files on the Git Server.	All required test files are successfully uploaded to the correct repository on the Git Server and are accessible at the time of assessment.	0.2	 Full marks shall be awarded if the competitor submits the file to the Git Server; zero marks will be given otherwise.

Aspect	WSOS Section (as per TD)	Descriptor	Requirement	Marks	Resource
M	3	Evidence capture is assessed based on the completeness and availability of required screenshots.	Screenshots showing the execution results of all test cases are saved in full within the designated screenshots directory. Partial or missing screenshots are not acceptable.	1.0	 Full marks for all complete screenshot files; no marks if screenshots are missing or insufficient in quantity.
M	4	Test data design is assessed by verifying the use of valid equivalence class analysis.	Input and output conditions are analysed, and three valid test data inputs are designed and implemented for each functional method. All required test data must be present in the CSV file.	0.9	 Full marks for CSV file with 3 valid test datasets; no marks for missing file, incomplete or erroneous data.
M	5	Use of data-driven testing is assessed by verifying the implementation method and required variables.	0.25 marks for the integration testing script uses a data-driven testing approach implemented with pytest. 0.5 marks for test data is stored in a variable named paraDatas, containing all data sets required for execution.	0.75	 0.25 marks for integration testing script with pytest-implemented data-driven approach, 0.5 marks for all required execution datasets stored in variable paraDatas

Aspect	WSOS Section (as per TD)	Descriptor	Requirement	Marks	Resource
M	6	Test reporting completeness is assessed by verifying required report elements in the Allure test report.	For each test case, the Allure report includes: 1. functional module name, 2. functional point name, 3. test case identifier, 4. documented test steps (including operations, inputs, and assertions), 5. a corresponding screenshot. All elements must be present, 0.4 marks deducted per missing element or data error.	2.0	

Judgement

Judgement is used to assess the quality of performance about which there may be small differences of view when applying the external benchmarks.

The Resources column in the table below may include all kinds/formats of resources such as: a link to a YouTube video, a website link, an illustration, a photo, a reference to a book, and so on. It needs to be as detailed as possible to show the differentiation between 0, 1, 2, 3.

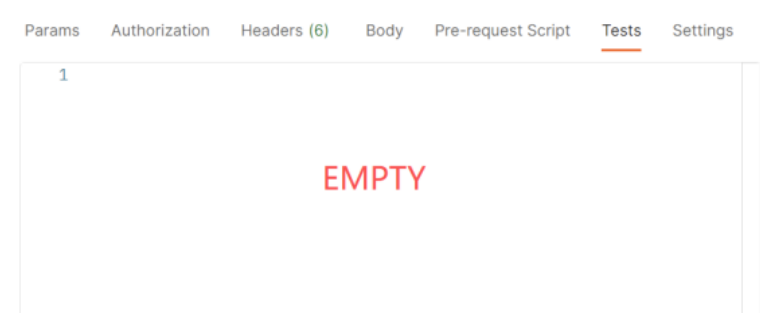
Generic rules:

The assessment group comprises three Experts plus one Expert who acts as the supervisor. The supervising Expert will replace an Expert in the marking group to prevent compatriot marking.

The difference of the three Experts' scores may not exceed 1. If this is the situation the Experts must re-score until there is a maximum difference of 1. As long as the three Experts judge within 1, the result can be entered into the CIS.

The total marks allocated to judgement marking may vary from competition to competition depending on the Test Project.

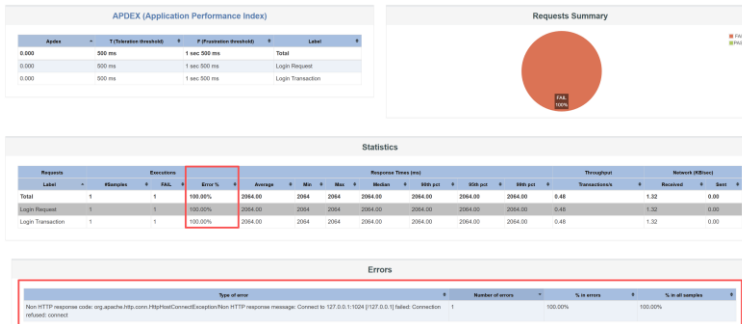
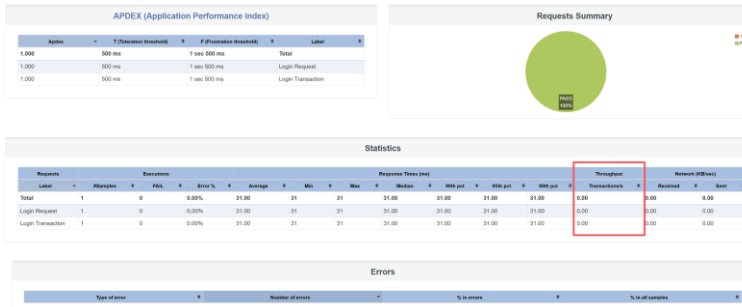
Aspect – Evaluate the assertion design for the login success response scenario

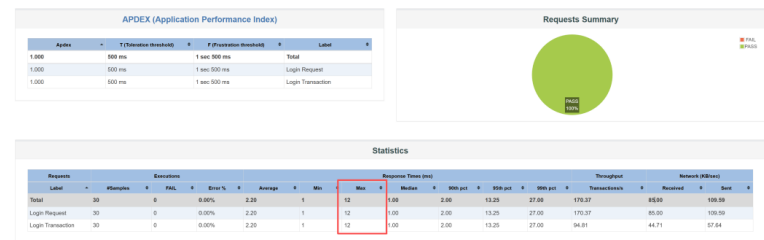
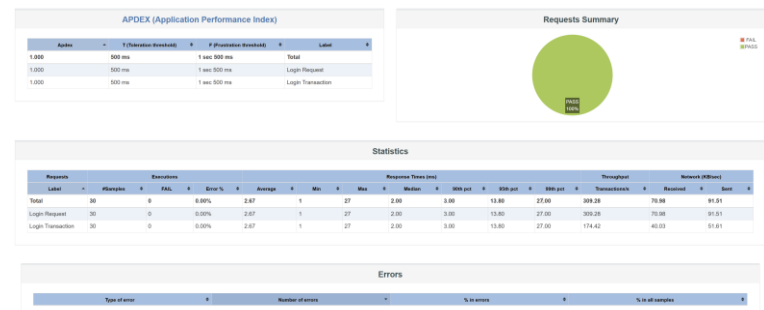
Aspect	Score	Descriptor	Resource
Judgement	0	No assertions; assertions irrelevant to success scenario; severe syntax/logic errors.	

Aspect	Score	Descriptor	Resource
	1	Valid assertions < 3 or errors ≥ 2; insufficient explanation of purposes.	Params Authorization Headers (6) Body Pre-request Script Tests Settings <pre> 1 // 1-Point Example: Only 1 valid assertion, 1 logical error, // missing purpose comments (Meets 1-point criteria) 2 // Valid Assertion 1: Verify HTTP status code (only correct // assertion) 3 pm.test("Status code 200", () => pm.response.to.have.status(200)); 4 // Missing purpose comment → insufficient explanation 5 6 // Invalid Assertion: Logical error (assert token is null for // success scenario) 7 pm.test("Token is null", () => { 8 pm.expect(pm.response.json().data.token).to.be.null; // Logical // error: Token must be non-null in success scenario 9 }); </pre>
	2	Valid assertions ≥ 3 but incomplete coverage; clear explanation of most purposes.	Params Authorization Headers (6) Body Pre-request Script Tests Settings <pre> 1 // 2-Point Example: 3 valid assertions (status code + core fields // only), no data type coverage (Meets 2-point criteria) 2 // Assertion 1: Verify HTTP status code (covers status code // dimension, clear purpose) 3 pm.test("HTTP response status code is 200 (API connectivity check) ", () => { 4 pm.response.to.have.status(200); 5 }); 6 7 // Assertion 2: Verify business code = 200 (covers core field // dimension, clear purpose) 8 pm.test("Business code 200 (login success flag)", () => { 9 const resData = pm.response.json(); 10 pm.expect(resData.code).to.eql(200); 11 }); 12 13 // Assertion 3: Verify token is not null (covers core field // dimension, clear purpose) 14 pm.test("Token is non-null (valid auth token returned)", () => { 15 const resData = pm.response.json(); 16 pm.expect(resData.data.token).not.to.be.null; 17 }); 18 </pre>

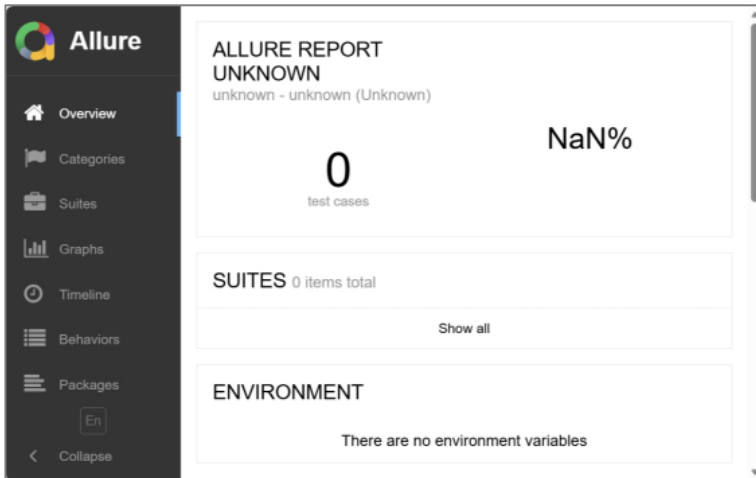
Aspect	Score	Descriptor	Resource
	3	≥3 non-redundant assertions covering status code, core fields, data types; clear & rigorous purpose explanations.	Params Authorization Headers (6) Body Pre-request Script Tests Settings <pre> 1 // 3-Point Example: 3 non-redundant assertions (full dimension // coverage), clear purpose comments (Meets 3-point criteria) 2 // Assertion 1: Verify HTTP status code (status code dimension - // basic connectivity check) 3 pm.test("Verify login API status code = 200 (ensure normal API // communication)", () => { 4 pm.response.to.have.status(200); 5 }); 6 7 // Assertion 2: Verify token is non-null (core field dimension - // auth validity check) 8 pm.test("Verify non-null token (ensure valid authentication for // follow-up operations)", () => { 9 const resData = pm.response.json(); 10 pm.expect(resData.data.token).not.to.be.null; 11 }); 12 13 // Assertion 3: Verify userId is number type (data type dimension - // spec compliance) 14 pm.test("Verify userId is number type (comply with API data // specification)", () => { 15 const resData = pm.response.json(); 16 pm.expect(resData.data.userId).to.be.a('number'); 17 }); </pre>

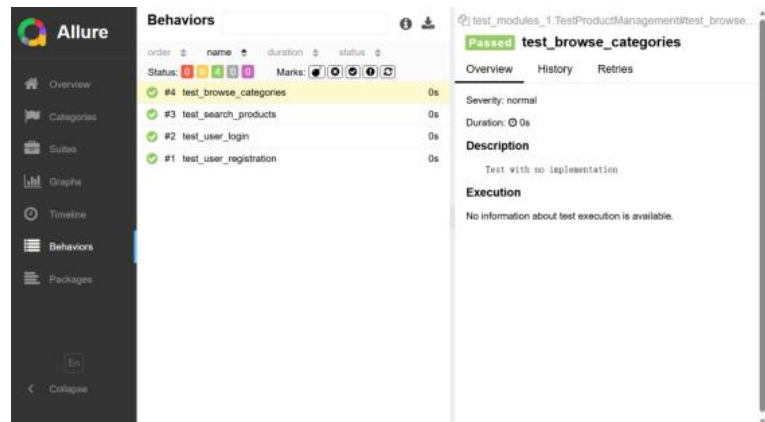
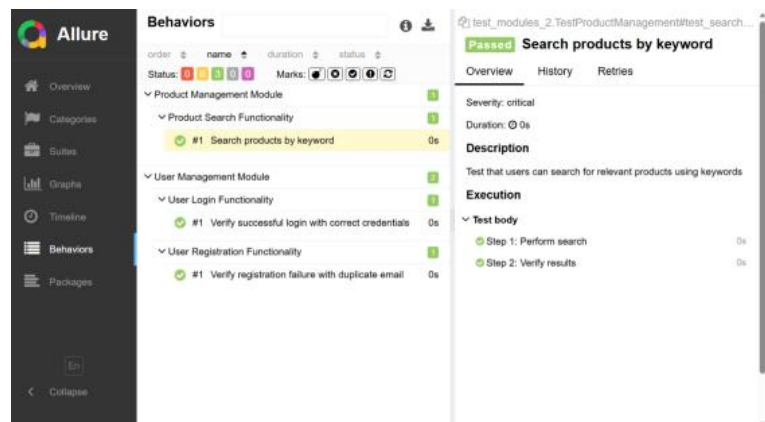
Aspect – Evaluate the compliance of JMeter login test execution results with expected system performance indicators

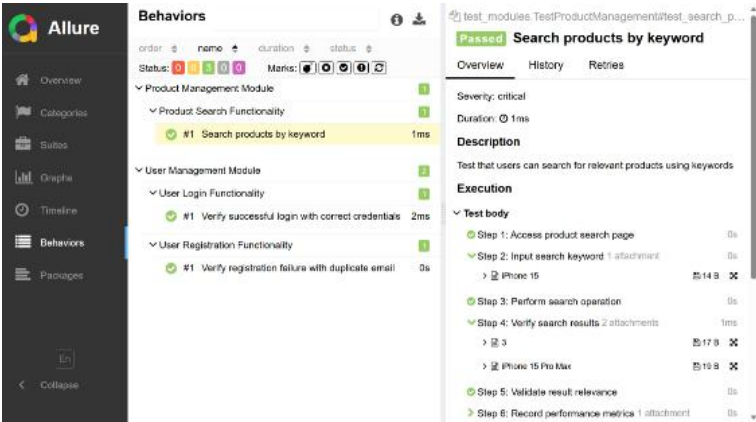
Aspect	Score	Descriptor	Resource
Judgement	0	Execution data irrelevant to login performance testing, no indicator records.	 The resource shows JMeter results for a failed login test. The 'Errors' section is highlighted with a red box, indicating a 100% failure rate for the 'Login Transaction'.
	1	Incomplete execution content, missing core login indicators (success rate, response time); indicators deviate severely from expectations; no anomaly explanations, unable to judge test outcomes.	 The resource shows JMeter results for a successful login test. The 'Statistics' section is highlighted with a red box, indicating a 0% failure rate for the 'Login Transaction'.

Aspect	Score	Descriptor	Resource
	2	Complete execution content covers core login indicators; most meet expectations with minor deviations; main anomalies labeled, brief analysis, content layout flaws.	 The screenshot shows the APDEX (Application Performance Index) dashboard. It includes a 'Requests Summary' section with a green circle indicating a 'PASS' status. Below this is a 'Statistics' table with columns for 'Requests', 'Executions', 'Response Time (ms)', 'Throughput', and 'Network (KB/sec)'. The 'Max' column in the 'Response Time (ms)' section is highlighted with a red box.
	3	All login performance indicators meet expected standards; precise anomaly location & specific analysis; clear logic, directly guides optimization and follow-up testing.	 The screenshot shows the APDEX (Application Performance Index) dashboard. It includes a 'Requests Summary' section with a green circle indicating a 'PASS' status. Below this is a 'Statistics' table with columns for 'Requests', 'Executions', 'Response Time (ms)', 'Throughput', and 'Network (KB/sec)'. The 'Max' column in the 'Response Time (ms)' section is highlighted with a red box. Additionally, there is an 'Errors' section at the bottom of the dashboard.

Aspect – Evaluate the content integrity of the Allure test report complies with industry standards

Aspect	Score	Descriptor	Resource
Judgement	0	Allure test report cannot be launched, or test report is empty.	 The screenshot shows the Allure test report interface. The status is 'UNKNOWN' with the subtext 'unknown - unknown (Unknown)'. It displays '0 test cases' and 'NaN%' pass rate. Below this, it shows 'SUITES 0 items total' with a 'Show all' link. The 'ENVIRONMENT' section states 'There are no environment variables'. A sidebar on the left contains navigation links: Overview, Categories, Suites, Graphs, Timeline, Behaviors, Packages, and a collapse button.

Aspect	Score	Descriptor	Resource
	1	The Allure test report is missing necessary content, or the test fails to achieve the inspection purpose.	
	2	The content of the Allure test report is complete, with the names of functional modules, functional test points, titles, etc., and there are test steps, but the test steps are not comprehensive and the content is not rich enough.	

Aspect	Score	Descriptor	Resource
	3	The content of the Allure test report is complete, with the names of functional modules, functional test points, titles, test steps, etc., and there are test steps, or other content descriptions. The test steps are comprehensive, and the content is detailed.	 The screenshot displays the Allure test report interface. On the left, a sidebar shows navigation options: Overview, Categories, Suites, Graphs, Timeline, Behaviors (selected), and Packages. The main area is titled 'Behaviors' and lists several test behaviors under different modules. The 'Product Search Functionality' module is expanded, showing a behavior '#1 Search products by keyword' with a duration of 1ms. To the right, a detailed view of this behavior is shown, including its status (Passed), severity (critical), duration (1ms), and a description: 'Test that users can search for relevant products using keywords'. The 'Execution' section shows a list of test steps, each with a status (green checkmark for passed, red X for failed) and a duration. The steps include: Step 1: Access product search page (0s), Step 2: Input search keyword '1' attachment (0s), Step 3: Perform search operation (0s), Step 4: Verify search results '2' attachments (1ms), Step 5: Validate result relevance (0s), and Step 6: Record performance metrics '1' attachment (0s).