

**ASSESSMENT PLATFORM DEMONSTRATION**

**Online Tests and Assessments Platform**

**REFERENCE NUMBER**

**RM6129**

**Attachment 11**

# INTRODUCTION

All Bidders successfully passing the Compliance Stage and the Selection

Stage (Attachment 2), will be required to provide an Assessment Platform Demonstration.

The Assessment Platform Demonstration will test in a live environment that the key output requirements of Scenario’s 1 to 7 can be delivered by Bidders. Bidders should familiarise themselves with the Scoring at paragraph 4 of this document.

# PRE-DEMONSTRATION ACTIVITY

Bidders will need to create a test creator and test taker user profile ahead of the demonstration; these profiles must be accessible on the day of the demonstration to support a successful outcome.

The demonstration session has a fixed duration - when the allocated time has concluded you will be required to stop whether you have finished or not. No-one may overrun.

Some of the demonstration scenarios include showing and using online tests that meet the acceptance criteria listed for each task. Depending on the time it takes to complete the scenarios you may wish to create the tests in advance. If they are created in advance you will need to walk through how the task acceptance criteria were met.

For example, if the scenario is to create a test, you may create this test in advance and then demonstrate each step of the test creation process in the session.

# Location and Availability

The dates of the Assessment Platform Demonstrations are set out in the ITT. The Authority will confirm the exact date, time and location of a Bidder’s Assessment Platform Demonstration via the eSourcing Suite.

All demonstrations will take place at the Civil Service building in London. This is likely to be Bush House (Bush House South West Wing, 44-46 Aldwych, WC2B 4RD) which will be confirmed when the dates have been confirmed. Wifi will be available. On your allocated day, Bidders will report to reception 45 minutes before the start of their demonstration.

Bidders will be escorted to a holding room, and 30 minutes before the start of the demonstration will be given a “Scenario Information Pack” based on the scenarios set out in paragraph 6 below.

# DEMONSTRATION

The requirements will be demonstrated to the evaluators following a series of scenarios as defined in this “Scenario Information Pack”. The scenarios will allow Bidders to manage the demonstration as they see fit, within the allotted timescale of 3 hours, rather than explicitly allocating timings to each requirement.



**It is the responsibility of the Bidder to ensure that they demonstrate all the required outputs within the allocated time.**

Evaluators may ask questions during the demonstration for clarification purposes only, but it is up to the Bidder to use the demonstration time effectively and to highlight to evaluators where the output requirements have been met.



**Bidders must tell the evaluators when they are moving from scenario to scenario.**

The evaluators will score the demonstration solely on the scenarios provided in this document. Bidders should not use the time to demonstrate functionality which is outside of the scope of the scenario, as this will not be taken into account by the Authority. All scenarios must be fully demonstrated within the allotted timescale of 3 hours.

The Bidder’s System Quality Demonstration must demonstrate a live test creation system, capable of delivering the Authority’s requirements. It must not demonstrate using a development, test or demonstration site that is not equivalent to the live system.

Should the Bidder fail to demonstrate the service, then you may be excluded from the Procurement.

Please note, that the time allocation will not commence until the Bidder has established the required IT connection and in the unlikely event that there are any technical challenges on the day, beyond the control of the Bidder, this will **not** result in exclusion from the Procurement.

# Scoring

Scoring for these scenarios is based on the number of tasks met within the scenario which is then weighted.

Number of scenario tasks met

X Weighting of the scenario

|  |
| --- |
| Total tasks in scenario |

The total mark is calculated by adding up the marks for each scenario.

For example – for AQC5 if 2 tasks are met this would then be divided by the total number of tasks (3 tasks) in the scenario multiplied by the weighting of the scenario (10).

AQC5 example:

2 tasks met

X 10 (weighting of the scenario) = 6.67

|  |
| --- |
| 3 tasks in scenario |

See table below for example scoring of all Scenario questions.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Scenario Questions** | **Number of Tasks Met** | **Number of Tasks** | **Weighting** | **Your weighted mark** |
| **SECTION B – SPECIFIC QUESTIONS** | | |  |  |
| **AQC1** | 3 Divided by | 5 x | 25 | 15.00 |
| **AQC2** | 3 Divided by | 4 x | 25 | 18.75 |
| **AQC3** | 1 Divided by | 1 x | 15 | 15.00 |
| **AQC4** | 3 Divided by | 3 x | 5 | 5.00 |
| **AQC5** | 2 Divided by | 3 x | 10 | 6.67 |
| **AQC6** | 3 Divided by | 4 x | 15 | 11.25 |
| **AQC7** | 1 Divided by | 1 x | 5 | 5.00 |
| **Quality Mark** | | | | **76.67** |
| **Weighted Quality Mark (35%)** | | | | **26.83** |

# SCENARIOS

These are the scenarios that will be used for demonstration of the Online Tests and Assessments Platform.

Please refer to paragraph 5 for how the marking scheme (number of tasks met) is scored and calculated.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Scenario Questions** | | **Marking Scheme**  **(Number of tasks met)** | | **Weighting** |
| **SECTION B – SPECIFIC QUESTIONS** | | |  | |
| **AQC1** | **Scenario 1** - Create a linear test containing multimedia content, multiple traits (or dimensions), and combined scoring. | 0/1/2/3/4/5 | 25 | |
| **AQC2** | **Scenario 2** - Create an item response theory (IRT) test using computerised adaptive testing (CAT) and convert raw scores to standardised scores using score norms. | 0/1/2/3/4 | 25 | |
| **AQC3** | **Scenario 3 -** Demonstrate that GOV.UK styling can be adopted for the tests. | 0/1 | 15 | |
| **AQC4** | **Scenario 4 -** Create a free-text (constructed) response test and show that the results can be reviewed and marked. | 0/1/2/3 | 5 | |
| **AQC5** | **Scenario 5 -** Schedule and deliver a test, using an anonymous identifier to generate a unique URL and then export the results. | 0/1/2/3 | 10 | |
| **AQC6** | **Scenario 6 -** Demonstrate that tests created on your platform are compatible with a range of browsers, devices and assistive technologies. | 0/1/2/3/4 | 15 | |
| **AQC7** | **Scenario 7 -** Demonstrate how management information can be accessed and reports generated. | 0/1 | 5 | |

# Online Systems Quality Evaluation

|  |  |  |
| --- | --- | --- |
| **AQC1 - Scenario 1 - Create a linear test containing multimedia content, multiple traits (or dimensions), and combined scoring.**  This scenario relates to creating a situational judgement test for recruitment. | | Met /  not-met |
| Notes | * Depending on how long this scenario will take to complete, you may choose to create the test in advance of the demonstration. In the session, you will need to walk through how each step was completed. * Question content and videos are just a vehicle to showing functionality. You are not expected to develop a true psychometric test. * You are not required to demonstrate a raw to standard score conversion for this scenario. | Met /  not-met |
| 1.1 | Create a linear test with 4 simple multiple choice questions. 2 of these questions must use a video as the scenario (stem) for the question. |  |
| 1.2 | The user can return to a previous question as well as progress to the next. This will be a “back” button or link. |  |
| 1.3 | The test must have 2 sections.  The video-content questions are one section and the other questions are a separate section. |  |
| 1.4 | The sections must generate scores separately for each trait and a total test score will be generated combining both sections.  Test section 1 is worth 40% of the overall score.  Test section 2 is worth 60% of the overall score. |  |
| 1.5 | At the end of the test provide a pass/fail outcome based on the outcome of the test.  If the overall score is >50% then it is a pass, if it is below then it is a fail.  This feedback page must also show the user’s scores for each section.  E.g.  You passed the test with a total score of 80%  In section 1 you got 1 out of 2 (50%) correct  In section 2 you got 2 out of 2 (100%) correct |  |

|  |  |  |
| --- | --- | --- |
| **AQC2 - Scenario 2 - Create an item response theory (IRT) test using computerised adaptive testing (CAT) and convert raw scores to standardised scores using score norms.**  This relates to creating adaptive verbal and numerical tests for recruitment. | | **Met / not-met** |
| Notes | * Depending on how long this scenario will take to complete, you may choose to create the test in advance of the demonstration. In the session, you will need to walk through how each step was completed. * Item content is just a vehicle to showing functionality. You are not expected to develop a true psychometric test. * IRT parameters do not need to be psychometrically/statistically accurate. They are just needed for demonstration purposes, just show how they can be applied and updated. |  |
| 2.1 | Create a test with 10 multiple choice questions with a range of difficulties that dynamically adapts based on which response a user selects. Create a starting and stopping parameter to illustrate with.  example test questions:  Q1: 10+10 = ?  If correct go to Q2 [harder]  else (incorrect) go to Q3 [easier]  Q2: 100+100=?  Q3: 1+1=?  Although there are 10 questions, when taking the test the user should not experience all of them. For example in the simple example above the user would see either Q2 or Q3, not both. |  |
| 2.2 | Change the IRT parameters (give some items a higher difficulty rating). |  |
| 2.3 | Change the stopping rule to increase the required accuracy level for score estimation. Demonstrate the change in terms of numbers of items to attempt before the stopping rule kicks in. |  |
| 2.4 | Convert the user’s score using norms and provide a) a pass/fail outcome, then b) a “processed” score (e.g. percentile). Then demonstrate a standard score conversion using STEN scaling. |  |

|  |  |  |
| --- | --- | --- |
| **AQC3 - Scenario 3 - Demonstrate that GOV.UK styling can be adopted for the tests.**  When an applicant takes a test they will be expecting a GOV.UK style site so that it is consistent with the wider recruitment journey (and their experiences of government digital services). | | **Met / not-met** |
| Notes | * We do not expect a perfect adoption of the GOV.UK style at this stage. The GOV.UK patterns are a default style that should be adopted based on user needs and user research, so there is flexibility. We need to see that your service is able to provide this flexibility to adopt the patterns and adapt them appropriately. |  |
| 3.1 | Show how these patterns/design elements can be edited by the CSHR users.  Where this would require effort/input by the Supplier (e.g. code changes by your developers) you must explain when this is needed and the process for managing it. |  |

|  |  |  |
| --- | --- | --- |
| **AQC4 - Scenario 4 - Create a free-text (constructed) response test and show that the results can be reviewed and marked.** | | Met / not-met |
| Notes: | * You may wish to create the free-text response in advance to paste in (if possible) rather than having to type it out. |  |
| 4.1 | Create a test with a free-text (constructed) response question that requires the user to fill in a detailed response (> 200 words). |  |
| 4.2 | As a test taker, complete the test and the written response. You can prepare the response in advance and paste it into the response during the session. |  |
| 4.3 | As a test administrator/marker, review the written response in full, add comments (e.g. notes and feedback) and apply a score to the response. |  |

|  |  |  |
| --- | --- | --- |
| **AQC5 - Scenario 5 - Schedule and deliver a test, using an anonymous identifier to generate a unique URL and then export the results.**  Most of our tests will be scheduled and issued via the integrated applicant tracking system(s) whereby the test link would be shown on their application hub page.  However, some tests need to be administered manually but without bringing personal data (e.g. email addresses) into the test platform. | | Met / not-met |
| Notes | * We expect this scenario to be fully demonstrated in the session. |  |
| 5.1 | For one of the tests created previously, generate two unique hyperlinks to provide to “applicants/test takers” through which the users to access directly. Their scores are collected against a unique identifier. |  |
| 5.2 | Using the unique hyperlink, access the test and complete it. |  |
| 5.3 | View and export the results of the test and detail the export formats available for test results. |  |

|  |  |  |
| --- | --- | --- |
| **AQC6 - Scenario 6 - Demonstrate that tests created on your platform are compatible with a range of browsers, devices and assistive technologies.**  It is crucial that the tests we create are accessible to all applicants (with only extreme exceptions where equivalency cannot be provided). You must explain what your platform supports and any additional effort required to make tests compatible. | | Met / not-met |
| Notes | * For 6.2 “resolutions”, you are able to use an emulator such as that provided in Google Chrome. <https://developers.google.com/web/tools/chrome-devtools/device-mode/> * Free trials for the assistive technologies listed above are available online. If you are unable to access these let us know at least 5 days ahead of your session. * Guidance on testing with assistive technology <https://www.gov.uk/service-manual/technology/testing-with-assistive-technologies> |  |
| 6.1 | Demonstrate your platform is capable of a range of browsers by showing a short test, which can be for a previous scenario, using:   * Google Chrome (latest version / n-1) * Mozilla Firefox * Internet Explorer 11 * Safari |  |
| 6.2 | Demonstrate the test using:   * Mobile resolution e.g. iPhone 8 * Tablet resolution e.g. IPad |  |
| 6.3 | Demonstrate the test using assistive technologies:   * screen reader (JAWS, VoiceOver or NVDA) * Screen magnifier (e.g. Zoomtext) |  |
| 6.4 | Demonstrate how your service notifies users when they access a test using an incompatible browser.  For example, if a user accesses a test using internet explorer 8 (assuming your tests are not compatible) your service may present a warning message. Alternatively, your service may provide a compatibility test feature as part of your tests. |  |

|  |  |  |
| --- | --- | --- |
| **AQC7 - Scenario 7 - Demonstrate how management information can be accessed and reports generated.**  Much of the complete management information/business intelligence analysis will be conducted using the applicant tracking system and other analysis systems. However, we expect to be able to do some basic reporting on the testing platform. | | Met / not-met |
| Notes | Demonstrate how you would access data on test delivery:   * How many tests have been completed this week? * How many times has test {*test example}* been completed in the last month? * How many unique users have taken tests today? * How many users are currently online taking tests? |  |
| 7.1 | Demonstrate how this MI data could be accessed and extracted for analysis (e.g. API access, data exports) |  |