Highfield Level 4 End-Point Assessment for ST0118 Data Analyst

End-Point Assessment Kit

Highfield Level 4 End-Point Assessment for ST0118 Data Analyst

EPA Kit

Contents

Please click on the headings below to navigate to the associated section of the EPA Kit.

Versions:

ST0118 / v1.1 / AP03

DA v1.4

How to use this EPA Kit

Welcome to the Highfield End-Point Assessment Kit for the Data Analyst apprenticeship standard.

Highfield is an end-point assessment organisation that has been approved to offer and carry out end-point assessments for the Level 4 Data Analyst apprenticeship standard.

The EPA Kit is designed to outline all you need to know about the end-point assessments for this standard and will also provide an overview of the on-programme delivery requirements. In addition, advice and guidance for trainers on how to prepare apprentices for the end-point assessment is included. The approaches suggested are not the only way in which an apprentice may be prepared for their assessments, but trainers may find them helpful.

In this kit, you will find:

- an overview of the standard and any on-programme requirements
- a section focused on amplification
- guidance on how to prepare the apprentice for gateway
- detailed information on which part of the standard is assessed by which assessment method
- suggestions on how to prepare the apprentice for each part of the end-point assessment
- a section focused on the end-point assessment method where the assessment criteria are presented in a format suitable for carrying out 'mock' assessments

Introduction

Standard overview

The broad purpose of the occupation is to ascertain how data can be used to answer questions and solve problems. For example: a data analyst may investigate social media trends and their impact on the organisation. In their daily work, an employee in this occupation interacts with internal or external clients.

A data analyst will be responsible for the creation and delivery of their own work. They will be responsible for working within the data architecture of the company and ensuring that data is handled in a compliant and secure manner. They will also understand and adhere to company data policy and legislation.

Data analysis is a fast-moving and changing environment and the analyst needs to stay up to date and engaged with changes and trends in the wider industry.

On-programme requirements

Although learning, development and on-programme assessment is flexible, and the process is not prescribed, the following is the recommended baseline expectation for an apprentice to achieve full competence in line with the Data Analyst apprenticeship standard.

The on-programme assessment approach will be agreed between the training provider and employer. The assessment will give an ongoing indication of an apprentice's performance against the final outcomes defined in the standard. The training provider will need to prepare the apprentice for the end-point assessment, including preparation for the professional discussion and collation of the portfolio of evidence (such as a provision of recordings of professional discussions or workplace evidence).

The training programme leading to end-point assessment should cover the breadth and depth of the standard using suggested on-programme assessment methods that integrate the knowledge, skills and behaviour components, and which ensure that the apprentice is sufficiently prepared to undertake the end-point assessment. Training, development and ongoing review activities should include:

- the achievement of level 2 English and maths. If the apprentice began their apprenticeship training before their 19th birthday, they will still be subject to the mandatory requirement to study towards and achieve English and maths. The requirements for English and maths are optional for apprentices aged 19+ at the start of their apprenticeship training.
- any qualifications specified by the employer
- the completion of a portfolio through which the apprentice gathers evidence of their progress

- structured one-to-one reviews of their progress with their employer and/or training provider

Portfolio of evidence

The apprentice must compile a portfolio of evidence during their time on-programme that is mapped against the knowledge, skills and behaviours assessed in the professional discussion underpinned by a portfolio of evidence.

It will typically contain **10 pieces of evidence**. Evidence may be used to demonstrate more than **1 knowledge, skill or behaviour** as a qualitative approach is suggested as opposed to a quantitative approach.

Evidence sources for the portfolio may include:

- progress review documentation, witness testimonies and feedback from colleagues and/or clients
- written accounts of activities
- photographic evidence and work products
- work instructions
- safety documentation
- technical reports
- drawings
- company policies and procedures

This is not a definitive list and other evidence sources are possible. Given the breadth of context and roles in which this occupation works, the apprentice will select the most appropriate evidence based on the context of their practice against the relevant knowledge, skills and behaviours.

The portfolio should not include reflective accounts or any methods of self-assessment. Any employer contributions should focus on direct observation of performance (for example, witness statements) rather than opinions.

The portfolio must be accompanied by a portfolio matrix. This can be downloaded from our website. The portfolio matrix must be fully completed, including a declaration by the employer and the apprentice to confirm that the portfolio is valid and attributable to the apprentice.

The portfolio of evidence **must** be submitted to Highfield at gateway. It is **not** directly assessed but underpins the professional discussion.

Use of artificial intelligence (AI) in the EPA

Assessments must be carried out in accordance with the published assessment plan and all work submitted must be the apprentice's own. AI tools must not be used to generate evidence in its entirety or to replace the apprentice's own judgement, performance or competence. Any use of AI must be transparent, limited and properly referenced.

Where AI has been used by the apprentice as part of normal work activity (for example, drafting a document, worksheet or PowerPoint) this may form part of the portfolio provided that:

The apprentice has materially authored, verified and taken responsibility for the content:

- AI use is clearly declared and referenced within the work (include tool name, purpose and how outputs were verified)
- Source prompts, system settings and the portions influenced by AI are retained and available for review
- AI outputs must not substitute for authentic demonstration of competence against the standard

If an AI tool is used at any stage of an assessment method (for example, to prepare a presentation outline or to organise notes), its use must be fully referenced in the submission or assessor records, and must not compromise authenticity, validity or security. Assessors must be satisfied that decisions remain rooted in the apprentice's knowledge, skills and behaviours, and in direct evidence gathered through observation, questioning and professional discussion.

AI tools must not be used to produce assessment evidence end-to-end, to fabricate logs/records or to simulate performance.

Readiness for end-point assessment

For an apprentice to be ready for the end-point assessments:

- the apprentice must have achieved level 2 English and maths. The requirements for English and maths are mandatory for all apprentices aged between 16-18 at the start of their apprenticeship training. The requirements for English and maths are optional for apprentices aged 19+ at the start of their apprenticeship training.
- the apprentice must have gathered a **portfolio of evidence** against the required elements to be used as the basis for the professional discussion.
- the line manager (employer) must be confident that the apprentice has developed all the knowledge, skills and behaviours defined in the apprenticeship standard and that the apprentice is competent in performing their role. To ensure this, the apprentice must attend a formal meeting with their employer to complete the Gateway Readiness Report.

- the apprentice and the employer should then engage with Highfield to agree a plan and schedule for each assessment activity to ensure all components can be completed within a **mandated** end-assessment window. Further information about the gateway process is covered later in this kit.

If you have any queries regarding the gateway requirements, please contact your EPA Customer Engagement Manager at Highfield Assessment.

Order of end-point assessments

There is no stipulated order of assessment methods. This will be discussed with the apprentice, training provider and/or employer with our scheduling team when scheduling the assessments, to ensure that the learner is provided with the best opportunity to attempt the assessment.

Click here to return to contents

The Highfield approach

This section describes the approach Highfield has adopted in the development of this end-point assessment in terms of its interpretation of the requirements of the end-point assessment plan and other relevant documents.

Specific considerations

Highfield's approach does not deviate from the assessment plan.

Dispensation

A temporary dispensation has been applied to the ST0118 version 1.1 end-point assessment plan for this apprenticeship. The dispensation will remain live until redundant apprentices identified in the dispensation request have completed their EPA, including any resits and retakes, when it will then be withdrawn. Highfield will implement the dispensation as required and agreed by IfATE (now Skills England), supported and monitored by Ofqual.

The key changes are:

- the training provider will step in as the employer for the duration of the EPA period in order to ensure that apprentices that have been made redundant can conduct their work-based project and complete their apprenticeship

The dispensation applies to the agreed apprentices only. Highfield must contact Skills England for each individual case.

Click here to return to contents

Gateway

How to prepare for gateway

After apprentices have completed their on-programme learning, they should be ready to pass through 'gateway' to their end-point assessment.

Gateway is a meeting that should be arranged between the apprentice, their employer and training provider to determine that the apprentice is ready to undertake their end-point assessment. The apprentice should prepare for this meeting by bringing along work-based evidence, including:

- customer feedback
- recordings
- manager statements
- witness statements

As well as evidence from others, such as:

- mid and end-of-year performance reviews
- feedback to show how they have met the apprenticeship standards while on-programme

In advance of gateway, apprentices will need to have completed the following. The requirements for English and maths listed below are mandatory for all apprentices aged between 16-18 at the start of their apprenticeship training. The requirements for English and maths listed below are optional for apprentices aged 19+ at the start of their apprenticeship training.

- Achieved level 2 English
- Achieved level 2 maths
- Compiled a portfolio

Therefore, apprentices should be advised by employers and providers to gather this evidence and undertake these qualifications during their on-programme training. It is recommended that employers and providers complete regular checks and reviews of this evidence to ensure the apprentice is progressing and achieving the standards before the formal gateway meeting is arranged.

The gateway meeting

The gateway meeting should be attended by the apprentice and a representative from the employer or training provider.

The **Gateway Readiness Report** should be used to log the outcomes of the meeting and agreed by **all** 3 parties. This report is available to download from the Highfield Assessment website.

The report should then be submitted to Highfield. If you require any support completing the Gateway Readiness Report, please contact your EPA Customer Engagement Manager at Highfield Assessment.

Reasonable adjustments
Highfield Assessment has measures in place for apprentices who require additional support. Please refer to the Highfield Assessment Reasonable Adjustments policy for further information/guidance.

ID requirements
Highfield Assessment will complete an identification check before starting any assessment and will accept the following as proof of an apprentice's identity:

- a valid passport (any nationality)
- a signed UK photocard driving licence
- a valid warrant card issued by HM forces or the police
- another photographic ID card, such as an employee ID card or travel card

Click here to return to contents

The Data Analyst apprenticeship standard

Below are the knowledge, skills and behaviours (KSBs) from the standard and related assessment criteria from the assessment plan. On-programme learning will be based upon the KSBs and the associated assessment criteria are used to assess and grade the apprentice within each assessment method.

Project with presentation and questioning		
Knowledge	**Skills**	**Behaviour**
K3 Principles of the data life cycle and the steps involved in **carrying out routine data analysis tasks.** **K4 Principles of data**, including open and public data, administrative data, and research data. **K8 Quality risks** inherent in data and how to mitigate or resolve these. **K9** Principal approaches to defining **customer requirements** for data analysis. **K11** Approaches to **organisational tools and methods** for data analysis. **K12** Organisational **data architecture**.	**S1** Use data systems securely to meet requirements and in line with **organisational procedures** and legislation including principles of Privacy by Design. **S2** Implement the stages of the **data analysis lifecycle**. **S3** Apply principles of **data classification** within data analysis activity, flexing approach as necessary. **S4** Analyse data sets taking account of different **data structures** and database designs. **S6 Identify and escalate quality risks** in data analysis with suggested mitigation or resolutions as appropriate. **S7** Undertake **customer requirements analysis** and implement findings in data analytics planning and **outputs**.	**B3** Work independently and collaboratively. **B4** Logical and analytical.

	S8 identify data sources and the **risks and challenges** to combination within data analysis activity. **S12 collaborate and communicate** with a range of internal and external stakeholders using appropriate styles and behaviours to suit the audience. **S15** select and apply the most **appropriate data tools** to achieve the optimum outcome.	

Pass criteria	Distinction criteria
PP1 Operates data systems in compliance with all organisational and legislative requirements including principles of Privacy by Design. (S1) **PP2** Outlines and applies the principles of **data analysis lifecycle** to the steps of data analysis. (K3, S2) **PP3** Describe the **principles of data** including open, public, administrative and research data and how they relate to the data used within the project. (K4) **PP4** Demonstrates a reasoned application of the principles of **data classification**. Explains where any flexibilities in application have been applied and their purpose. (S3) **PP5** Identifies **quality risks** in data analysis and outlines methods to mitigate, escalate and/or resolve them. (K8, S6)	***PP14*** *Evaluates the outcomes of data analysis and suggests alternative tools/methods which would be of benefit to all stakeholders. (K8, K11, B4)* ***PP15*** *Analyses the requirements of the customer to produce a data analysis plan which provides an optimum solution. (S7)* ***PP16*** *Demonstrates the use of data sets with different* ***data structures*** *and database designs to solve problems or improve the accuracy or efficiency of data analysis. (S4)*

PP6 Outlines and applies the principles for defining **customer requirements** and implements findings in data analytics planning and **outputs**. (K9, S7) **PP7** Demonstrates how data from different sources is combined and prepared for data analysis setting out how they identified the **risks and challenges** inherent in combining data within the project. (S8) **PP8** Describes the **tools and methods** used by their organisation for data analysis and identifies which were used within the project with reasoning for the choices made to achieve the best outcome. (K11, S15) **PP9** Analyses data sets taking account of different **data structures** and database designs. (S4) **PP10** Outlines the choice of organisational **data architecture**. (K12) **PP11** Communicates and collaborates with all relevant stakeholders and adapts communication style to meet audience and situational requirements. (S12) **PP12** Describes how they work independently and collaboratively detailing their impact on the work of others. (B3) **PP13** Acts independently to establish logical and analytical solutions such as exploring new data sets or resolving issues within the data. (B4)	
Amplification and guidance	
Carrying out routine data analysis tasks could include: • identifying the hypothesis and setting necessary objectives • identifying and collecting the data, whether as a one-time activity, or periodic reporting • transforming, merging, aggregating, quality checking and other processing activities	

- removing errors or duplicates, structuring the data and preparing it for analysis
- performing statistical and other analysis. Analysis may be:
    - descriptive
    - diagnostic
    - predictive
    - prescriptive
- presenting findings using appropriate methods such as reports, dashboards, visualisations and presentations

Principles of data could include:

- assets such as management of data assets throughout the lifecycle, such as:
    - managed data collection
    - version control, configuration management and backups
    - management of data quality
    - use of metadata and change audits
- data management such as management in an ethical, transparent and legally compliant way through all parts of the lifecycle, including how data is published
- reuse and linkage such as use and application of common standards, allowing for repeatability, data integration and data sharing
- security such as rules and governance around data accessibility, permissions and rules

Quality risks could include:

- data quality issues such as duplication, recording errors and data age
- unknown or unconfirmed origins of data, or unknown data definitions
- subjectivity and non-quantifiable bias as part of data collection or storage
- data conversion, migration or blending errors
- sufficiency of sample sizes

Customer requirements in relation to the outcomes of data analysis, such as what questions are being asked of the data that can then be clearly defined in the hypothesis.

Organisational tools and methods, software and hardware, and cloud services available to support data analysis processes could include:

- databases - relational in nature (RDBMs) or non-relational (noSQL) databases
- statistical packages - systems to support the application of statistical formula to the data
- data visualisation software – such as maps, graphs, tables and other visuals

Data architecture is a set of principles, rules, systems and processes relating to how data can be collected, stored and used, and is set out in a structured and standardised way. This could include:

- how data moves through a company, from how it is captured to how it is exported or reported
- its accessibility such as the systems or software in place to allow the right users to interact with it
- retaining confidentiality, security and integrity of the data
- documentation of the architecture

Organisational procedures, and data protection legislation, could include:

- secure access to data systems, with strong passwords, multi-factor authentication and role-based permission access
- storage of files and folders in approved and maintained locations, ensuring storage has been configured by the business
- awareness of cyber security threats and likely vectors for their transmission and technology such as anti-virus, encryption and malware protection
- application of elements of General Data Protection Regulation (GDPR) legislation

Data analysis lifecycle includes 6 stages and presentation of evidence for each of these stages, such as:

- the discovery stage - asking the questions around defining the problem or hypothesis to be solved, and understanding and engaging with stakeholders
- the data preparation stage - preparing what data needs to be collected to answer questions, what research is needed and the methods to capture, store, architect and secure the data

- the processing and model planning stage - using knowledge of data analysis tasks:
    - to create and transform data
    - to manage data integrity or quality risks
    - for data storage methods
    - for feasibility studies
    - for set up testing of datasets
    - to perform data cleansing
    - to merge, transform and prepare methods before analysis commences
- the analysis stage - uses tools, statistical analysis and other software to perform the analysis of the data to help identify patterns, interpret the data and assess the outcomes against the original hypothesis
- communicating the results - applies the appropriate technique to make the information and messages being delivered as clear as possible to the given audience
- acting on the outcomes - applies the outputs and insights provided by the analysis to solve a particular problem or to help inform debate or decision making to support a change or a business case

Data classification is a method of organising data into categories based on specific characteristics of the data. Classification of data often considers the sensitivity of the data or information. Examples could include:

- highly sensitive or confidential data, internal to specific people in the business, role-based control, and encryption and back up
- medium sensitive data, such as internal only or sharing only with specific customers
- low sensitive data, such as information in the public domain that can be published online or communicated out to all customers

Data structures and how these structures store, organise and manage data, in its simplest form, is a table comprising of rows representing the variables. This simple model represents a structured, tabular dataset, but some datasets consist of unstructured data that do not fit this traditional format.

Identify and escalate quality risks to identify them from the outset of any data project and ensure that these are recorded appropriately in some format of a risk register, along with suitable mitigations, controls, decisions on risk treatment and the impacts that would occur in the event that the risk occurred. Data risks could originate from many places but could include:

- poor or incomplete data sources, including sampling bias
- inaccurate data input or data preparation, such as inaccurate merging of datasets
- inappropriate storage, data relationship building or data corruption
- inappropriate or ineffective data analysis techniques
- incorrect interpretation of results
- data protection risks

Customer requirements analysis influences planning and requires an understanding of what kind of outputs are required. Consideration needs to be given to who the customers are and their drives and needs. Customer requirements analysis could take the form of qualitative data capture, such as collection through interviews and focus groups.

Outputs need to be presented in a suitable form so that they can be interpreted, manipulated and understood by stakeholders. Examples may be data dashboards and other visualisation tools.

Risks and challenges are important to understand to ensure that they do not create inaccuracies. These could include:

- inaccuracies, duplications or incomplete datasets requiring data validation and cleansing checks before any use for data analysis
- bias in data, or flaws in data collection methods requiring identification of potential bias and repeating or removing specific data collections
- data protection risks, such as consent for the use of data and a need to ensure all relevant consents and adherence to data protection legislation are in place
- incompatible or inconsistent data sets when trying to merge data sets and sources together

Understanding how risks and challenges can be overcome could include:

- standardised, documented approaches

- audits and validation checks
- using the right tools, including software and other technology to help identify issues and help resolve them

Collaborate and communicate at various points in the analysis lifecycle. There will be the need to engage with:

- customer requirements
- wider stakeholder management and interaction, potentially dealing with conflicting objectives, bias or desires from outcomes and different levels of understanding
- other analysts and the team working on the project who bring different skills sets and experience to the analysis

Appropriate data tools is critical to data analysis and the ability to identify accurate and informative outputs. Considerations for which tools to use could include:

- the context, size and sector of the business, and the type of data handled
- the data architecture, policies and governance already in place
- resources such as financial and willing to invest in analysis, and skills and capabilities within the business to perform the analysis
- the general 'appetite' to be data and business-intelligence driven as a business

Data tool selection may need to consider:

- the likely quality of the data and how much the system needs to handle 'missing' or incomplete data
- whether there are requirements to merge or transform data
- the need for presentation and visualisation software, and report systems
- cutting edge technologies or more 'traditional' systems and configurations
- security, data protection and ethical considerations
- cost and potential return on investment

Professional discussion with portfolio		
Knowledge	**Skills**	**Behaviours**
K1 Current **relevant legislation** and its application to the safe use of data. **K2** Organisational data and information security standards, policies and procedures relevant to **data management activities**. **K5** The **differences between structured and unstructured data.** **K6** The fundamentals of data structures, **database system design**, implementation and maintenance. **K7 Principles of user experience** and domain **context** for data analytics. **K10** Approaches to **combining data from different sources.** **K13** Principles of **statistics for analysing datasets**. **K14** The principles of **descriptive, predictive and prescriptive analytics**. **K15** The **ethical aspects** associated with the use of and **collation** of data.	**S5** Assess the impact on **user experience** and domain **context** on the data analysis activity. **S9** Apply **organizational architecture** requirements to data analysis activities. **S10** Apply **statistical methodologies** to data analysis tasks. **S11** Apply **predictive analytics** in the collation and use of data. **S13** Use a range of **analytical techniques** such as **data mining, time series forecasting** and **modelling techniques** to identify and predict trends and patterns in data. **S14** To collate and interpret **qualitative and quantitative** data and convert into infographics, reports, tables, dashboards, and graphs.	**B1** Maintain a productive, professional, and **secure working environment**. **B2** Shows initiative, being resourceful when faced with a problem and taking responsibility for solving problems within their own remit. **B5** Identifies issues quickly, enjoys investigating and solving complex problems and applies appropriate solutions. Has a strong desire to push to ensure the true root cause of any problem is found and a solution is identified which prevents recurrence. **B6** Demonstrates **resilience** by viewing obstacles as challenges and learning from failure. **B7** Demonstrates an ability to adapt to changing contexts within the scope of a project, direction of the organisation or Data Analyst role.
Pass criteria		**Distinction criteria**

PD1 Explains how current, **relevant legislation** impacts on the safe use of data and how their role contributes to a productive, safe, and **secure working environment**. (K1, B1) **PD2** Explains the relevant data policies and procedures for the organisation and identifies the data standards to be reached. (K2) **PD3** Describes the fundamentals of data structures and **database system design** and explains how they are implemented and maintained. (K6) **PD4** Explains approaches to **combining data from different sources** to improve accuracy and / or efficiency and / or maximise benefits to the organisation and / or customer. (K10) **PD5** Describes impact on **user experience** and domain **context** on data analysis. (S5) **PD6** Explains the **differences between Structured and Unstructured data**. (K5) **PD7** Explains the **ethical aspects** associated with the **collation** and use of data and justifies why this is important. (K15) **PD8** Describes the relevant tools or techniques used for working with the data systems **architecture** in their organisation. (S9) **PD9** Explains and applies the principles of **statistics for analysing datasets**. (K13, S10) **PD10** Identifies and explains challenges in their work and how they overcame them, providing an outline of lessons learned. (B6) **PD11** Explains how they have applied analytical techniques for **data mining** and **time series forecasting** and other **modelling techniques**. (S13) **PD12** Identifies areas of work where they adapted to changing contexts	***PD18*** *Critically evaluates the risks and benefits of* ***predictive analytics****. (K14, S11)* ***PD19*** *Compares and contrasts visual data representation approaches and how they aid understanding by stakeholders. (S14)* ***PD20*** *Evaluates the benefits and risks inherent in* ***combining data from different sources****. (K10)*

within the scope of a project, direction of the organisation or Data Analyst role. (B7) **PD13** Explains the principles of **descriptive**, **predictive** and prescriptive **analytics** and demonstrates how they have been applied within their own data analysis practice. (K14, S11) **PD14** Demonstrates data analysis activities involving the collation and interpretation of **qualitative and quantitative** data and displays results using visual representations. (S14) **PD15** Explains the **principles of user experience** and domain **context** for data analytics. (K7) **PD16** Describes how they have appropriately adapted their activities to meet minor, unexpected changes at work. (B2) **PD17** Describes how they have ensured the true root cause of any problem is found and a solution is identified which prevents recurrence. (B5)	

Amplification and guidance

Relevant legislation related to the use of data and information could include:

- the Data Protection Act
- the General Data Protection Regulation (GDPR)
- other legislation that can apply to a data analyst could include:
    - Computer Misuse Act
    - Privacy and Electronic Communications Regulation (PECR)
    - Copyright, Designs and Patents Act
    - Freedom of Information Act
    - The Waste Electrical and Electronic Regulation

- Payment Card Industry Data Security Standards
- Online Safety Bill
- Anti-terrorism, Crime and Security Act

Data management activities and processing could include:

- creation or engagement with organisational policies and procedures to ensure legislation requirements are met
- how these are formalised in documentation, training materials or other formats. For example:
  - privacy policies
  - user access controls (systems and physical)
  - data retention, classification and labelling
- how this information is communicated, whether to internal staff, customers, suppliers and other stakeholders
- registration with relevant authorities
- relevant roles and responsibilities within the business, such as a data protection officer (DPO) and/or an information management system team
- audits to identify and classify data, information or assets appropriately in relation to data protection

Differences between structured and unstructured data are:

- structured uses quantitative, categorised data that can be organised in a well-structured way, such as names, addresses, dates, heights and telephone numbers
- unstructured often uses qualitative data that does not have a predefined data model or structure, such as photographs, social media activity, internet of things (IoT) data and data in multiple file formats

Database system design means to consider what data is to be stored and whether this allows for a normalised, quantitative structured data model, or an unstructured one, such as:

- the purpose of the data and how this can inform the design of the database design
- what reporting requirements there are
- what systems there are

- who is accessing the data

A normalised, structured data model could include:

- how different data elements relate to one another to identify appropriate primary and foreign keys
- identification of data types to be used, such as Ints, Varchar and BLOB
- appropriate design to ensure data accuracy and integrity
- appropriate design to eliminate the risk of data redundancy

Principles of user experience could include:

- useful
- usable
- findable
- credible
- desirable
- accessible
- valuable

Context of the data also allows for:

- the ability to ask the right questions at data collection and data blending stages of the project
- identification and selection of appropriate datasets and areas of data that are appropriate to build effective reports and statistical models
- having the ability to validate (through previous experience) and interpret results effectively
- the ability to provide insights that are relevant to the audience

Combining data from different sources is a core part of data preparation and can involve compiling data from multiple sources, such as spreadsheets, written reports, sensors, databases and data held in a range of different file types.

Statistics for analysing datasets include:

- the study or experiment design identifying the hypothesis, the data that needs to be collected or used, the sample sizes, and the dependent and independent variables
- data types, the number of variables, sample types and the number of samples to compare
- how to interpret the data such as understanding of confidence levels applied, the concepts of statistical significance, validity, and the different outputs depending on the statistical method applied

Descriptive analytics such as analysing past performance or data to model how this impacts the future. Most statistical analysis falls into this category. Tools to support this type of analysis include the more 'traditional' data analysis techniques of data collection, aggregation and data mining.

Predictive analytics such as analysing potential future outcomes and which outcomes are most likely. Trying to predict what might happen based on probabilities and interpolation of existing data. Statistical analysis is commonly used to develop predictive models.

Ethical aspects of data **collation**, analysis and use of data could include:

- objectivity which includes impartiality, non-discriminatory and independence in the way that the analysis has been performed
- reassurances around honesty and trustworthiness, to ensure that the analysis and the outputs are treated with confidence by the stakeholders

User experience and the **context** can greatly influence the effectiveness and the outcomes of analysis, such as:

- the usability of data analysis tools and outputs making it easier to understand the data, conduct the analysis and interpret results
- providing interactive features can engage stakeholders more effectively, such as allowing filtering, sorting, and adjusting parameters to help identify patterns and relationships
- the relevance of the data to the domain context so stakeholders can interpret the data correctly and draw valid conclusions

Architecture is a set of principles, rules, systems and processes relating to how data can be collected, stored and used. The data is set out in a structured and standardised way.

Organisational architecture analyses the use of shared data, definitions, vocabulary governance and controls, which leads to more confidence on analysis outcomes, with greater understanding.

Statistical methodologies are applied to help draw the necessary conclusions from analysis activities.

Predictive analytics analyses data to identify potential future outcomes and which outcomes are most likely. This involves trying to predict what might happen based on probabilities and interpolation of existing data. Statistical analysis is commonly used to develop predictive models.

Predictive analysis methods, could include:

- regression describing the relationship between 2 variables where one of the variables can be predicted by the other
- a decision tree model, detailing the possible decisions and consequences
- a time series model, predicting future behaviour based on historical data
- machine learning and the training of models with previous data

Data mining techniques are commonly aligned to predictive analysis methods and identifying patterns in data, often in the context of large, complex datasets, such as those contained within data warehouses rather than simpler datasets. Techniques could include:

- grouping, or clustering groups of data points together to help identify patterns in the data
- assigning data points to predefined categories based on characteristics

Time series forecasting is based on the analysis of historical time-ordered data to predict future values. The tools and techniques for this aim to remove outliers and short-term fluctuations to 'smooth' out the data, which highlights trends over time. Time series forecasting tools can also be used to complete 'missing' data from time-ordered data.

Modelling techniques use a variety of techniques based on the complexity of the model and the number of data points, connections and rules. For simple models, regression and decision trees may adequately model the scenario. For more complex models, the range of data mining techniques may need to be utilised.

Qualitative and quantitative data sets can complement or support the outcomes of an activity more than if just 1 type is used. As such, they may need to be presented and communicated together.

The types of visualisation may differ depending on the data type, for example:

- qualitative data can be presented as thematic illustrations, word clouds and icons
- quantitative data can be presented through charts, graphs and tables

Secure working environment could include using encrypted connections, ensuring data is anonymised where necessary and maintaining a clean and organised digital workspace.

Resilience means to solve problems in the face of challenges and recovering from mistakes.

Click here to return to contents

Assessment summary

The end-point assessment for the Data Analyst apprenticeship standard is made up of 2 assessment methods:

1. A 3500-word (+/-10%) project, followed by a 40-minute presentation with questioning. Typically, 20-minutes for the presentation and 20 minutes for the questions and answers
2. A 60-minute professional discussion with portfolio

As an employer/training provider, you should agree a plan and schedule with the apprentice to ensure all assessment components can be completed effectively.

Each component of the end-point assessment will be assessed against the appropriate criteria laid out in this kit, which will be used to determine a grade for each individual. The grade will be determined using the combined grades.

Project with presentation and questioning

The project report and presentation is weighted equally with all other assessment methods. Apprentices will be marked against the pass and distinction criteria outlined in this kit.

- To achieve a **pass**, apprentices must achieve all of the pass criteria
- To achieve a **distinction**, apprentices must achieve all of the pass criteria **and** all of the distinction criteria
- **Unsuccessful** apprentices will not have achieved all of the pass criteria

The project report and presentation can be conducted face to face or via remote technology.

Professional discussion with portfolio

The professional discussion is weighted equally with all other assessment methods. Apprentices will be marked against the pass and distinction criteria outlined in this kit.

- To achieve a **pass**, apprentices must achieve all of the pass criteria
- To achieve a **distinction**, apprentices must achieve all of the pass criteria **and** all of the distinction criteria
- **Unsuccessful** apprentices will not have achieved all of the pass criteria

The professional discussion may be conducted using technology such as video link, as long as fair assessment conditions can be maintained.

Grading

The apprenticeship includes pass and distinction grades, with the final grade based on the apprentice's combined performance in each assessment method.

To achieve a pass, the apprentice is required to pass each of the 2 assessment methods.

To achieve a distinction, the apprentice must achieve a distinction in the project with presentation and questioning and a distinction in the professional discussion with portfolio.

Where the apprentice achieves a pass in one assessment method and a distinction in the other, they will achieve a merit overall.

The overall grade for the apprentice is determined using the matrix below:

Project with presentation and questioning	Professional discussion with portfolio	Overall grade awarded
Fail any of the 2 assessment methods		**Fail**
Pass	Pass	**Pass**
Pass	Distinction	**Merit**
Distinction	Pass	**Merit**
Distinction	Distinction	**Distinction**

Retake and resit information

If an apprentice fails an end-point assessment method, it is the employer, provider and apprentice's decision whether to attempt a resit or retake. If a resit is chosen, please call the Highfield scheduling team to arrange the resit. If a retake is chosen, the apprentice will require a period of further learning and will need to complete a retake checklist. Once this is completed, please call the Highfield scheduling team to arrange the retake.

An apprentice who fails one or more assessment methods, and therefore the EPA in the first instance, will be required to resit or retake the failed assessment method(s) only.

A resit is typically taken within 2 months of the EPA outcome notification. The timescale for a retake will be dependent on how much retraining is required but is typically taken within 4 months of the EPA outcome notification.

When undertaking a resit or retake, the assessment method(s) will need to be reattempted in full, regardless of any individual assessment criteria that were passed on any prior attempt. The EPA Report will contain feedback on areas for development and resit or retake guidance.

Apprentices will **not** need to complete a different project where a resit or retake is required but will need to revise their existing project report and/or presentation. Apprentices will be asked different questions in the case of a resit or retake.

Any EPA component resit/retake must be taken within a 6-month period, otherwise, the entire EPA must be retaken in full. Apprentices should have a supportive action plan to prepare for the resit/retake.

Apprentices who achieve a pass grade cannot resit or retake the EPA to achieve a higher grade.

The overall grade awarded for this apprenticeship standard is **not** capped, meaning that an apprentice can still achieve a maximum EPA grade of distinction if any assessment method needs to be resat or retaken.

Click here to return to contents

Assessing the project with presentation and questioning

This end-point assessment method consists of 2 components:

- data analysis project
- presentation with questioning

Component 1: Data Analysis Project

The project starts after the apprentice has gone through gateway. Together with the employer, the apprentice must select the most appropriate project title by reference to the work available to the apprentice. The employer will ensure it has a real business application and Highfield will ensure it meets the requirements of the EPA (including suitable coverage of the KSBs assignment to this assessment method). The duration of the project is a maximum of **8 weeks**, with the project typically taking **6 weeks** and then a further **2 weeks** to write the report. The employer should ensure that the apprentice has sufficient time and the necessary resources within this period to plan and undertake the project.

The project will be designed to give the apprentice the opportunity to demonstrate the KSBs mapped to the project with presentation and questioning within their day-to-day work and may cover the following project ideas to enable them to demonstrate competence (these are examples and other relevant project ideas are permitted):

- Patterns/trends and predictions
- Presenting statistical analysis results to inform decisions
- Optimising data models using statistical measures
- Reporting, dashboarding and data visualisation
- Data needs analysis, requirements and data requirements analysis
- Data architecture and data policy, procedures and legislation
- Real time and predictive
- Big data supports and optimises data science
- The role of AI in data analytics
- Data mining techniques

Given the large number of projects that will be completed per year, Highfield is not expected to sign-off each project title before the project commences. However, Highfield has provided detailed specifications and suggested project titles below to enable the employer to select a suitable project that will meet the requirements of the EPA:

- analysing trend in business performance
    - automating data flow from source by carrying suitable tasks such as cleansing and transformation

    - carrying out data manipulation using relevant tools leading to loading and visualisation
    - Producing reports and output to aid decision making
- carrying out HR performance tracking
    - extracting data from human resources databases, filtering and cleansing these to meet set requirements.
    - manipulating data to detect trends and patterns in abnormal staff absences
    - reporting the information and take remedial actions
- using data mining to support real-time risk detection and mitigation
    - identify patterns and anomalies within high-volume or real-time data streams
    - automatically uncover relationships and associations between variables or behaviours
    - use machine learning or statistical models to predict the likelihood of a particular risk or event occurring
- application of predictive analytics to support strategic planning and outcomes
    - analyse large datasets to identify trends or early indicators
    - use statistical or machine learning models to forecast future outcomes
    - apply findings to develop or recommend new strategies or improvements

Although it is not a requirement for Highfield to sign off project titles, apprentices and their employers can seek advice and guidance from Highfield on the suitability of a project the apprentice is planning to work on.

The project outcome should be in the form of an electronic report and should include an explanatory narrative, which sets out the tasks undertaken as well as accompanying infographics.

As a minimum, the project report **must** contain the following:

- an introduction
- the scope of the project (including key performance indicators)
- project outcomes and how the outcomes were achieved
- a project plan
- consideration of legislation, regulation, industry and organisational policies, procedures and requirements
- analysis
- research and findings
- recommendations and conclusions

The project report should also acknowledge sources. The project must also map in an appendix how it evidences the KSBs relevant to this assessment method.

The report has a maximum word limit of **3,500 words** (+/- 10%) and appendices, references and diagrams will **not** be included in this total. End-point assessors will only mark reports up to **3,850 words**, at which point, assessors will **stop** marking and only credit the criteria covered to that point. Reports that fall short of the word count will be marked in full, against all criteria. The assessor will review and assess the project report in advance of the presentation with questioning.

The apprentice may work as part of a team which could include technical internal or external support. However, the report will be the apprentice's own work and will be reflective of their own role and contribution. The report **must** be uploaded in PDF format and **must** be accompanied by the **written submission sheet** which is available to download from the Highfield Assessment website. On the written submission sheet, the apprentice and their employer must verify that the submitted report is the apprentice's own work and must map how it evidences the relevant KSBs for this assessment method, as outlined in this kit.

Component 2: presentation with questioning

The presentation will be based on the project report and will cover:

- a summary of the main aspects of the project
- context/implications/recommendations from the project
- practical application of knowledge, skills and behaviours
- business recommendations/outcomes of the project, including visualisations
- any follow-on outcomes
- actions and next steps

The presentation will be completed and submitted after gateway, within 8 weeks of the EPA start date, typically at the same time as the project report is submitted to Highfield. The apprentice must outline details of visual aids to be used and specify any equipment required for the presentation. The presentation will be presented to an end-point assessor, either face-to-face in a suitable controlled environment or via online video conferencing. The apprentice will have 10 days' notice of the presentation date to prepare for the presentation. The way in which the content of the presentation is delivered is **not** prescriptive.

The presentation and questioning will last **40 minutes**, typically including a **20-minute** presentation followed by **20 minutes** of questioning, where the end-point assessor will ask a minimum of **8 questions**. The end-point assessor has the discretion to increase the time of the presentation by up to 10% to allow the apprentice to complete their last point or respond to a question. Only the end-point assessor will observe the presentation. **The employer will not be present at the presentation. It will be presented by the apprentice to the end-point assessor.**

Before the assessment

Employers/training providers should:

- give the apprentice time to work on their project and report during the end-point assessment window
- ensure the apprentice knows the date, time and location of the assessment
- ensure the apprentice knows which data analyst criteria will be assessed (outlined on the following pages)
- encourage the apprentice to reflect on their experience and learning on-programme to understand what is required to meet the standard and identify real-life examples
- be prepared to provide clarification to the apprentice, and signpost them to relevant parts of their on-programme experience as preparation for this assessment

Grading the project with presentation and questioning

Apprentices will be marked against the pass and distinction criteria included in the tables on the following pages (under project report and presentation with questioning criteria).

- To achieve a **pass**, apprentices must meet all of the pass criteria
- To achieve a **distinction**, apprentices must meet all of the pass **and** distinction criteria

Project with presentation and questioning mock assessment

It is suggested that a mock assessment is carried out by the apprentice in advance of the end-point assessment with the training provider/employer giving feedback on any areas for improvement. It is the employer/training provider's responsibility to prepare apprentices for their end-point assessment. Highfield recommends that the apprentice experiences a mock presentation in preparation for the real thing. The most appropriate form of mock assessment will depend on the apprentice's setting and the resources available at the time.

When planning a mock assessment, the employer/training provider should include the following elements:

- mock presentations should be **40 minutes**, typically with the presentation lasting **20 minutes** followed by **20 minutes** of questioning
- consider a recording of the mock assessment and allow it to be played back to other apprentices, especially if it is not practicable for the employer/training provider to carry out a separate mock assessment with each apprentice
- ensure that the apprentice's performance is assessed by a competent trainer/assessor, and that feedback is shared with the apprentice to complete the learning experience
- mock assessment sheets are available to download from the Highfield Assessment website and may be used for this purpose
- structured, 'open' questions should be used as part of the questioning that do not lead the candidate but allow them to express their knowledge in a calm and comfortable manner. Some examples of this may include the following:
    - how do you ensure you operate data systems in compliance with organisational requirements?
    - describe the principles of open data.
    - what quality risks did you identify in your data analysis?
    - what data tools did you use in your project and why?
    - how did you ensure you collaborated with all relevant stakeholders during your project?
    - how did you work independently when completing your project, and how did the work of others impact you and your project?
    - how did analysing the requirements of the customer help you to produce a data analysis plan for your project?
    - provide an example of a data set you used within your project and explain how this helped you to solve problems.

Project with presentation and questioning criteria

Throughout the project with presentation and questioning, the assessor will review the apprentice's competence in the criteria outlined below.

Apprentices should prepare for the project with presentation and questioning by considering how the criteria can be met and reflecting on their past experiences.

Project with presentation and questioning
To pass, the following must be evidenced.
PP1 Operates data systems in compliance with all organisational and legislative requirements including principles of Privacy by Design. (S1)
PP2 Outlines and applies the principles of data analysis lifecycle to the steps of data analysis. (K3, S2)
PP3 Describe the principles of data including open, public, administrative and research data and how they relate to the data used within the project. (K4)
PP4 Demonstrates a reasoned application of the principles of data classification. Explains where any flexibilities in application have been applied and their purpose. (S3)
PP5 Identifies quality risks in data analysis and outlines methods to mitigate, escalate and/or resolve them. (K8, S6)
PP6 Outlines and applies the principles for defining customer requirements and implements findings in data analytics planning and outputs. (K9, S7)
PP7 Demonstrates how data from different sources is combined and prepared for data analysis setting out how they identified the risks and challenges inherent in combining data within the project. (S8)
PP8 Describes the tools and methods used by their organisation for data analysis and identifies which were used within the project with reasoning for the choices made to achieve the best outcome. (K11, S15)
PP9 Analyses data sets taking account of different data structures and database designs. (S4)
PP10 Outlines the choice of organisational data architecture. (K12)
PP11 Communicates and collaborates with all relevant stakeholders and adapts communication style to meet audience and situational requirements. (S12)
PP12 Describes how they work independently and collaboratively detailing their impact on the work of others. (B3)
PP13 Acts independently to establish logical and analytical solutions such as exploring new data sets or resolving issues within the data. (B4)

To gain a distinction, the following must be evidenced
PP14 *Evaluates the outcomes of data analysis and suggests alternative tools/methods which would be of benefit to all stakeholders. (K8, K11, B4)* ***PP15*** *Analyses the requirements of the customer to produce a data analysis plan which provides an optimum solution. (S7)* ***PP16*** *Demonstrates the use of data sets with different data structures and database designs to solve problems or improve the accuracy or efficiency of data analysis. (S4)*

Click here to return to contents

Assessing the professional discussion with portfolio

In the professional discussion with portfolio, the assessor and the apprentice will have a formal two-way conversation. It will consist of the independent assessor asking the apprentice questions to assess their competence against the relevant criteria outlined in this kit.

The apprentice can bring a copy of the portfolio with them to the professional discussion. The portfolio will be used by the apprentice to refer to and exemplify a point. However, the portfolio of evidence is **not** directly assessed.

The professional discussion will be scheduled at least **2 weeks** after gateway. It will take place in a suitable environment and can be conducted by video conferencing. It will last for **60 minutes**. The independent assessor can increase the time of the professional discussion by up to 10% to allow the apprentice to complete their last answer.

The assessor will ask **at least 10 open questions**.

Before the assessment

Employers/training providers should:

- ensure the apprentice knows the date, time and location of the assessment
- ensure the apprentice knows which criteria will be assessed (outlined on the following pages)
- encourage the apprentice to reflect on their experience and learning on-programme to understand what is required to meet the standard
- be prepared to provide clarification to the apprentice, and signpost them to relevant parts of their on-programme experience as preparation for this assessment

Grading the professional discussion with portfolio

Apprentices will be marked against the pass and distinction criteria included in the tables on the following pages (under 'Professional discussion with portfolio criteria').

- To achieve a **pass**, apprentices must achieve all of the pass criteria
- To achieve a **distinction**, apprentices must achieve all of the pass criteria **and** all of the distinction criteria
- **Unsuccessful** apprentices will have not achieved all of the pass criteria

Professional discussion with portfolio mock assessment

It is the employer/training provider's responsibility to prepare apprentices for their end-point assessment. Highfield recommends that the apprentice experiences a mock professional discussion with portfolio in preparation for the real thing. The most appropriate form of mock professional discussion underpinned by a portfolio of evidence will depend on the apprentice's setting and the resources available at the time.

In designing a mock assessment, the employer/training provider should include the following elements in its planning:

- the mock professional discussion with portfolio should take place in a suitable location.
- a 60-minute time slot should be available if it is intended to be a complete professional discussion covering all relevant standards. However, this time may be split up to allow for progressive learning.
- consider a video or audio recording of the mock professional discussion with portfolio and allow it to be available to other apprentices, especially if it is not practicable for the employer/training provider to carry out a separate mock assessment with each apprentice.
- ensure that the apprentice's performance is assessed by a competent trainer/assessor, and that feedback is shared with the apprentice to complete the learning experience. Mock assessment sheets are available to download from the Highfield Assessment website and may be used for this purpose.
- use structured, 'open' questions that do not lead the apprentice but allows them to express their knowledge and experience in a calm and comfortable manner. For example:
    - explain how current legislation impacts how you use data.
    - what data policies and procedures do you adhere to in your organisation?
    - how do you implement data structures in your daily role?
    - what is your approach to combining data from different sources to maximise benefits to the organisation?
    - describe the impact on domain context in data analysis.
    - why are the ethical aspects associated with the collation of data important?
    - how do you apply the principles of statistics when analysing datasets?
    - how do you evaluate the risks and benefits of predictive analytics?
    - how do you compare visual data representation approaches within your daily role?

- what are the ingrained risks when combining data from different sources?

Professional discussion with portfolio criteria

Throughout the **60-minute** professional discussion with portfolio, the assessor will review the apprentice's competence in the criteria outlined below. Apprentices should prepare for the professional discussion with portfolio by considering how the criteria can be met.

Professional discussion with portfolio criteria
To pass, the following must be evidenced.
PD1 Explains how current, relevant legislation impacts on the safe use of data and how their role contributes to a productive, safe, and secure working environment. (K1, B1)
PD2 Explains the relevant data policies and procedures for the organisation and identifies the data standards to be reached. (K2)
PD3 Describes the fundamentals of data structures and database system design and explains how they are implemented and maintained. (K6)
PD4 Explains approaches to combining data from different sources to improve accuracy and / or efficiency and / or maximise benefits to the organisation and / or customer. (K10)
PD5 Describes impact on user experience and domain context on data analysis. (S5)
PD6 Explains the differences between Structured and Unstructured data. (K5)
PD7 Explains the ethical aspects associated with the collation and use of data and justifies why this is important. (K15)
PD8 Describes the relevant tools or techniques used for working with the data systems architecture in their organisation. (S9)
PD9 Explains and applies the principles of statistics for analysing datasets. (K13, S10)
PD10 Identifies and explains challenges in their work and how they overcame them, providing an outline of lessons learned. (B6)
PD11 Explains how they have applied analytical techniques for data mining and time series forecasting and other modelling techniques. (S13)
PD12 Identifies areas of work where they adapted to changing contexts within the scope of a project, direction of the organisation or Data Analyst role. (B7)
PD13 Explains the principles of descriptive, predictive and prescriptive analytics and demonstrates how they have been applied within their own data analysis practice. (K14, S11)
PD14 Demonstrates data analysis activities involving the collation and interpretation of qualitative and quantitative data and displays results using visual representations. (S14)
PD15 Explains the principles of user experience and domain context for data analytics. (K7)
PD16 Describes how they have appropriately adapted their activities to meet minor, unexpected changes at work. (B2)
PD17 Describes how they have ensured the true root cause of any problem is found and

Professional discussion with portfolio criteria
To pass, the following must be evidenced.
a solution is identified which prevents recurrence. (B5)
To gain a distinction, the following must be evidenced.
PD18 *Critically evaluates the risks and benefits of predictive analytics. (K14, S11)* ***PD19*** *Compares and contrasts visual data representation approaches and how they aid understanding by stakeholders. (S14)* ***PD20*** *Evaluates the benefits and risks inherent in combining data from different sources. (K10)*

Click here to return to contents