

Value Added measure for 16-19 year olds specified at subject and qualification level

Value Added technical specification (Summary)

Purpose of Specification

This specification details:

- What data are used in the calculation of 16-19 Value Added
- Which learners/ qualifications/ exams are included in the calculation
- How to calculate prior attainment
- How to calculate national coefficients of expected attainment
- How to calculate Value Added scores per subject/ qualification
- How national samples of less than 80 aims per qualification/ subject are treated
- How to aggregate Value Added scores

Version 1.2

21 February 2006

Contents

Introduction

Stage I - Data preparation

Section 1 – Calculation of prior and output attainment

Section 2 – Learners/ Exams to include in calculations

Section 3 – Production of datasets for calculations of coefficients

Stage II - Specification of data subsets

Section 4 – Selection of data subsets

Section 5 – Specification of data for Multi-Level Modelling software

Section 6 – Storing and saving catalyst file

Section 7 – Grouping of subjects

Stage III – Multi-Level Modelling estimation

Section 8 – Creation of master files and subsets

Section 9 – Defining the specification/ equation

Section 10 – Output from the Multi-Level Modelling (national coefficients)

Stage IV - Calculating Value Added components

Section 11 – Calculation of key coefficients and variance covariance matrix

Stage V - Calculation of Value Added scores

Section 12 – Calculation of individual qualification aim Value Added scores

Section 13 – Data output file after calculating Value Added scores

Section 14 – Calculation of institution Value Added score per subject/ qualification

Stage VI - Calculation of institution Value Added curve

Section 15 – Calculation of the Value Added curve for an institution per subject/ qualification

Section 16 – Calculation of confidence intervals for the institution Value Added curve

Stage VII - Aggregation of Value Added scores

Section 17 – Calculation of institution Value Added score at qualification and institution level

Section 18 – Calculation of Value Added score at Sector Subject Area within qualification and Sector Subject Area across qualifications

Section 19 – Calculation of significance of institution Value Added scores

Section 20 – Specification of aggregated Value Added output file

Annex A: Modelling rules for Multi Level Modelling

Annex B: List of data tables used in Value Added calculation

Annex C: Example of Learner Achievement Tracker output charts

Stage I - Data preparation

Input: Data processed by matching contractor

Processes:

- calculation of prior attainment (section 1)
- calculation of output attainment (section 1)
- selection of learners/ exams to include in calculation (section 2)
- production of data set for calculations of coefficients (section3)

Output: Annex B Table 1.

Stage I - Data preparation

This section details the work undertaken by the checking contractor and also the data required to calculate VA.

Section 1 – Calculation of prior and output attainment

Data to enable the LSC to measure value added (VA) for 16-19 year olds in 2004/05 were provided by Forvus. These data included a number of fields that were calculated by Forvus, a data matching contractor, on our behalf, including prior and output attainment.

Calculation of prior attainment

1. For the VA measure, prior attainment is defined as: 'Average attainment in qualifications at Level 2 and below approved for use pre-16 under Section 96'
2. For 16 and 17¹ year olds, prior attainment is cumulative attainment in qualifications at Level 2 and below up to and including age 15. For 18¹ year olds, prior attainment is cumulative attainment in qualifications at Level 2 and below, up to and including age 16.

Points scores

3. Prior attainment is calculated using the Qualification and Curriculum Authority (QCA) point score system. Full details of the qualifications that are included in prior attainment calculation and the points assigned to each qualification are included in the Department for Education and Skills (*DfES*) *Production Tables*. The assignment of points is described in the Value Added and Distance Travelled Quality and Improvement Pack (*VA/DT QIP*). Briefly, points are assigned to a qualification depending upon the size of the qualification and the challenge level of the levels of attainment (for example grades) that can be achieved in that qualification.

Calculation of scores

4. For learners currently **aged 16 and 17** average prior attainment is calculated as:

$$\frac{\text{Total point score in qualifications at Level 2 and below achieved up to and including age 15}}{\text{Total size of qualifications at Level 2 and below taken}^2 \text{ up to and incl. age 15}}$$

5. For learners currently **aged 18**, average prior attainment is calculated as:

$$\frac{\text{Total point score in qualifications at Level 2 and below achieved up to and including age 16}}{\text{Total size of qualifications at Level 2 and below taken}^2 \text{ up to and incl. age 16}}$$

Institution checks and adjustments to data

6. All institutions included in the calculation of VA are required to check the data used in the calculation of prior attainment and the outcome attainment as part of the production of the Achievement and Attainment Table. Where exam results are deleted, withdrawn or claimed internally they are excluded from the calculation of prior attainment. The amendments at exam level also include certain learner level amendments as these amendments impact on the inclusion or exclusion of exam results. Where the learner is noted as transferring out, not on roll or double counted the results are excluded. For all other amendments the results are included.

Calculation of output attainment

7. Outcome attainment is calculated using the QCA point scores system. This follows the policy for reporting results in the School and College Achievement and Attainment Tables (SCAAT).

¹ Age of learner as at 31 August 2004.

² Qualification taken means "exam entered" for all qualification through examination; and "qualification completed" for all qualification through assessment.

8. Provision for 16-19 year olds is extremely diverse with different subject and qualification mixes being offered by a range of schools, further education, tertiary, specialist and sixth form colleges. To allow fair comparisons to be made while recognising this diversity, the VA for 16-19 learners is calculated by qualification and by subject. Therefore, for example, learners undertaking A level English can be compared with other learners taking English and not with learners taking AS level Physics or BTEC National Diploma in Performing Arts, for example.

9. Output attainment is measured separately for each subject (or subject group) for each qualification an individual has outcome attainment. Output attainment is calculated for all subjects in a qualification which can be identified. Subjects are identified using the QCA syllabus classification codes.

10. Qualifications to be included in the calculation of national lines and individual and institutional VA scores will include all approved qualifications at Level 3 with graded attainment that can be reliably identified.

The full list of Qualifications identified for inclusion in the VA measure is shown below:

- A level (legacy)³
- GCE A level
- AS level (Advanced Supplementary)³
- GCE AS level
- GCE AS level Double Award
- International Baccalaureate
- Free standing Maths Qualification Level 3
- Vocational GCE Single Award (VCE A)
- Vocational GCE AS Single Award (VCE AS)
- Vocational GCE Double Award (VCE DA)
- BTEC National Award
- BTEC National Certificate
- BTEC National Diploma
- OCR National Certificate at Level 3
- OCR National Diploma at Level 3
- OCR National Extended Diploma at Level 3
- ABC Diploma in Foundation Studies (Art & Design)
- CACHE Diploma in Child Care
- BTEC Diploma in Foundation Studies (Art & Design)

11. The output attainment is the QCA points assigned to the grade attained by the learner in the specific subject in that qualification, after the application of discounting rules.

Grouping with small cohort

12. Any subject/ qualification with a cohort greater or equal to 80 aims is analysed independently, otherwise such subjects must be analyzed at Subject Sector Area (SSA) level. For further information on this see section 7.

Treatment of AS levels

13. For AS levels and VCE AS levels, learner attainment is included for the 2003/04 and 2004/05 academic years for those AS levels that the learner has not taken a GCE/ VCE A level in the same subject. The rationale is explained in the VA/ DT QIP. Briefly, whether the AS level is claimed in the year the learner took the AS level, or the year they completed the Level 3 programme is as much due to the policy of the institution as any other factor. Therefore, it is not possible to reliably identify the AS levels taken in any one year by those learners included in the VA measure. Further, the intention is to include only those AS levels that are not taken as part of an A level, so as to avoid 'double counting'.

³ Legacy qualification codes are included for completeness, although there are no records in the data used.

Age of learner

14. Output attainments of learners who are 16, 17 and 18 years old at the commencement of the academic year they take the exam (that is 31 August 2004) age included in the VA calculation. Resits will not be included; that is, discounted against first award.

Treatment of fails and partial achievements

15. Fails are excluded from the calculation of VA scores whether for individual, institution or LSC/ Local Education Authority for the pilot year. This decision is to be reviewed for future years after the publication of this technical specification

16. If partial achievement is possible, those qualifications are given 0 points and are excluded from the calculation of national lines and VA scores.

Treatment of learner drop out

17. Learners who fail to complete their programme of study are excluded from all calculations, as reliable data are not available across the sector. Where there is reliable data, these learners are included in the Qualification Success Rate measure. Therefore, only learners who have entered and received an award are included in the VA measure.

Section 2 – Learners/ Exams to include in calculations

Learners/ Exams included in national lines

1. Exams are included in the calculation of the national lines (that is, the regression line that represents the national estimate of expected performance) if they meet all of the following criteria:

- Qualification is an approved Level 3 qualification within the scope of VA taken in the 2004/05 academic year
- Qualification is an AS level taken in the 2003/04 academic year
- Exam record is not discounted and has not been deleted, claimed internally or withdrawn by the institution
- QCA points for exam record are greater than zero (that is, not a fail)
- Learner passes “pilot trigger” (that is, has taken at least one qualification equal or greater in size than an A level in 2004/05 and has been deemed to be at the ‘end of sixth form by the institution they attended)
- Average prior attainment of the learner is greater than zero
- Records passing these criteria are given a value of ‘1’ on a flag NATINC. Learners passing these criteria are given a value of ‘1’ on flag INVA.

Learners/ Institutions included in VA calculations

2. Exam level VA scores are calculated for all results where INVA = 1

3. Exams are included in institutional averages if INVA = 1 & the institution is included in the DfES Achievement and Attainment Table (flag PSCHRES = 1).

Treatment of data from Consortia.

4. Where possible, disaggregated results should be included in the exam-level file, that is, results that can be allocated to a specific institution, with whole consortia results being excluded from the file to avoid double-counting.

5. Where consortia do not disaggregate their results, whole consortium post-16 results should be allocated to each institution in the exam-level file, as there is no way of knowing which results belong to which institution. In addition, whole consortia results should be excluded from the exam-level file to avoid double-counting of results.

Section 3 – Production of datasets for calculations of coefficients

1. Once all the calculations and data checks necessary to prepare the data for analysis have been performed an exam level file is sent to the LSC for processing by the Multi-Level Modelling (MLM) unit to produce national lines using the MLM software package MLWIN and then in the Learner Achievement Tracker (LAT) to produce VA scores.
2. No records are removed from this data file prior to it being sent to the LSC/ MLM unit. While three versions of this file are received, according to the extent the data has been checked by institutions. It is the final amended data set that is used in the calculation of VA scores.
3. A description of this file is given in Table 1 in annex B.

Stage II - Specification of data subsets

Input: Annex B Table 1.

Processes:

- selection of data subsets (section 4)
- specification of data for MLWIN (section 5)
- calculation of variables for MLWIN (section 5)
- storing and saving catalyst file (section 6)
- grouping of subjects (section 7)

Output: Annex B Table 2, Table 3 and Table 4.

Stage II - Specification of data subsets

This section provides the specification of the data sets that are used in the calculation of the national coefficients of the national lines of expected attainment and the calculation of VA scores. The input needed to generate these outcomes is given in Table 1 in annex B.

Section 4 – Selection of data subsets

1. To facilitate processing of the data, a selection of the fields is made. This selection is specified in Table 2 in annex B. This file has two uses as:
 - a) the base data file as input to the calculation of the National coefficients
 - b) the base data file as input to the calculation of the VA scores.

Section 5 – Specification of data for MLWIN

1. Once the appropriate learners and qualifications have been selected, the files need to be prepared for MLM.
2. The measure can include information regarding Sector Subject Area (SSA) in the calculations, which is derived from the Local Education Authorities Project (LEAP) subject mapping or Learn Direct Classification System (LDCS) code, whichever is contained in the data from Forvus.
3. Data are checked to ensure that there are no missing values, there are no repeating rows and that prior attainment is greater than 0. Records not meeting these criteria are deleted.
4. Prior and output attainment is tested for normality, outliers and extreme values. Values causing departure from normality are examined, checked and adjusted or removed
5. The following fields are deleted: GENDER, SURNAME, and FORNAME. The gender variable is not needed for the MLWIN calculation since there is no dummy variable for this. Surname and Forename have to be eliminated since such long string variables cause problems when being uploaded into MLWIN.

Calculation of variables for MLWIN

6. Once the basic data set has been created using the structure described above, it is necessary to transform the prior attainment variables before uploading the data into MLWIN. This process saves time in subsequent calculations and allows for quality assurance (QA) checks
7. The following calculations take place only once for the entire exercise:
 - i. Calculate the average of *PRIOR*, (*c*) of the **whole of the data set**
 - ii. Calculate *PRIOR* centred (*PRIORC*); ($PRIORC = PRIOR - c$).
 - iii. Calculate the square, cubic and quartic version of *PRIOR*; that is, *PRIOR2*, *PRIOR3*, *PRIOR4*.
 - iv. For these variables (*PRIOR2*, *PRIOR3*, *PRIOR4*), calculate their respective averages (c_2 , c_3 , c_4) and standard deviations (s_2 , s_3 , s_4) across the entire data set. These values, including *c*, are referred to as the catalysts, and are used later on in the process.
 - v. Centre and standardize *PRIORC* square, cubic and quartic, and refer to them as *PRIOR2C*, *PRIOR3C*, and *PRIOR4C*.
 - vi. Generate a vector of 1s and name it as “cons” (that is, add a constant)
8. It is important to make sure that the variables are created correctly, in order to avoid statistical complications in the estimation of the regressions. For this, it is recommended to check that the Pearson correlation amongst the variables is low. The comparison must be done between *PRIORC* and *PRIOR3C*, and then between *PRIOR2C* and *PRIOR4C*. If there is suspicion that the

correlation is high, it is recommended to repeat the calculation of, averages (c, c_2, c_3, c_4), standard deviations (s_2, s_3, s_4); and or in the calculation of *PRIORC PRIOR2C, PRIOR3C, PRIOR4C*.

Section 6 – Storing and saving catalyst file

1. The averages (c, c_2, c_3, c_4) and standard deviations (s_2, s_3, s_4) of the whole data set, used for the transformations of the variables are kept in a text file, and saved in a secure folder, since these are used again later on in the process. Table 4 in Annex B provides an example of the format for this file.

Section 7 - Grouping of subjects

1. It is not possible to obtain reliable results for a subjects in a qualification when less than 80 exams have been achieved nationally.

2. For these 'small subjects' the attainment is compared against different national lines according to the following rules:

Level 1: Assess against all exams in the SSA if the number of exams in the subject is less than 80, and the number of exams in the SSA is greater or equal to 80

Level 2: Assess against all exams in the qualification, if the number of exams in the subject is less than 80, and the number of exams in the SSA is less than 80.

3. VA scores can then be calculated for each small subject, but by comparing performance against the SSA/ qualification or qualification national line.

4. A file is then saved, which is detailed in Table 3 in annex B.

Stage III – MLM Estimation

Input: Annex B Table 3.

Processes:

- creation of master file (section 8)
- creation of subsets for subject/ qualification (section 8)
- defining equations (section 9)
- running MLM regressions (section 9)
- output from MLM (section 10)

Output: Annex B Table 5.

Stage III - MLM estimation

Section 8 – Creation of master files and subsets

1. The initial dataset is uploaded into MLWIN and sorted by exam record and institution. The former represents level 1 of the model and the latter level 2.

Creation of the master file

2. For all qualifications a single master file is created, containing the data for all subjects within that specific qualification. This master file must have all variables already created and a model specification in place before continuing in the next stage. The basic model that must be set up is model type A. For a complete of types of model see section 9, Table A and annex A.

Creation of subsets

3. Subsets of data are generated for each subject/ qualification to allow calculation of the national lines for each subject/ qualification.
- i. Define a subset of data based on a specific subject and save as an independent file, for example, (subject = 11010) & (qual = 111) would be the criteria to select all exam records for A-level Biology.
 - ii. Create a file for the subset of data for the calculation of the specific national line

Section 9 - Defining the specification/ equation

1. The core of the calculation of the VA scores is the statistical model that establishes the relationship between the output attainment and the prior attainment of an individual in a subject qualification. Such a relationship is determined by means of the following model and by applying a statistical technique called MLM. As mentioned in section 8, there are two levels in this model: level 1 is the exam record, and level 2 is institution.

2. The model specified allows for random parameters from the intercept up to the quadratic term of the equation. The cubic and quartic terms are fixed, and thus homogeneous for all cases. The default model is presented below.

Inputs/ outputs/ transients

y_{ij}	is the output attainment
x_{ij}	is the average prior attainment
β_{ij}	is the coefficient corresponding to each variable
μ_{ij}	is the term for the institutional effect
ε_{ij}	is the error term of the model
σ_{μ}	is the variance or covariance of the coefficients
σ_{ε}	is the variance of the error term

Default model

$$y_{ij} = \beta_{0ij} + \beta_{1j}x_{ij} + \beta_{2j}x_{ij}^2 + \beta_3x_{ij}^3 + \beta_4x_{ij}^4$$

$$\beta_{0ij} = \beta_o + \mu_{0j} + \varepsilon_{ij}$$

$$\beta_{1j} = \beta_1 + \mu_{1j}$$

$$\beta_{2j} = \beta_2 + \mu_{2j},$$

$$\begin{bmatrix} \mu_{0j} \\ \mu_{1j} \\ \mu_{2j} \end{bmatrix} \sim N(0, \Omega_{\mu}): \Omega_{\mu} = \begin{bmatrix} \sigma_{\mu 0}^2 & \sigma_{\mu 01} & \sigma_{\mu 02} \\ \sigma_{\mu 01} & \sigma_{\mu 1}^2 & \sigma_{\mu 12} \\ \sigma_{\mu 02} & \sigma_{\mu 12} & \sigma_{\mu 2}^2 \end{bmatrix}$$

$$[\varepsilon_{ij}] \sim N(0, \Omega_{\varepsilon}): \Omega_{\varepsilon} = [\sigma_{\varepsilon}^2]$$

3. Depending on the number of observations per file, a specific model has to be run. Table A below shows the different model options given the number of examination records available in each file.

Table A: Model specification according to the number of examination records

Type of Model	Number of exam records	Conditions	Model
A	More than 5000	None	Fixed quartic with quadratic institution effects.
B	1001-5000	None	Fixed cubic with linear institution effects
C	501-1000	Number of exam records must be greater or equal to eight times the number of institutions. (exam records $\geq 8 \times$ institutions)	Fixed cubic with linear institution effects
D	501-1000	Number of exam records must be less than eight times the number of institutions. (exam records $< 8 \times$ institutions)	Fixed cubic with constant institution effects
E	80-500	None	Fixed quadratic with constant institution effects

4. If the number of examination records in the model is less than 80 then it is not recommended to be used to estimate any model, but rather is combined with another subject following the criteria defined in section 7.

5. Depending on the criteria shown in Table A above, the model is adjusted in each file before running the regressions. All model specifications are in annex A.

Running the MLM regressions

6. The procedures describe how the national lines are generated, that is, the vector of coefficients betas (B) and the variance covariance matrix that describe the national estimates of expected attainment per subject per qualification.

7. Iterative Generalized Linear Squares (IGLS) is used to determine starting point for the regressions based on Markov Chain Monte Carlos (MCMC) process . When conducting the IGLS estimation, it is important to ensure that the matrix of variance-covariance of the terms is positive definitive.

8. Markov Chain Monte Carlos (MCMC)

- i. MCMC is a simulation technique, and it does not converge to a single best solution. Expertise in interpreting the run diagnostics is required to decide when the simulation has reached a local optimum.
- ii. When analysing the MCMC diagnostics the Brooks-Draper mean (Nhat) should ideally be under the length of the monitoring chain. Also, the Monte Carlo standard error (MCSE) should be at a level that indicates that extending the run chain would only result in a marginal reduction of the MCSE.

Section 10 - Output from the MLM (national coefficients)

1. The output of the estimation is then going to be used in the calculation of the VA scores. Tables B and C show the structure of the MLWIN outputs. These outputs will be used in the VA score calculation process. The output of the variance-covariance matrix is stored in a specific column of the MLWIN data set, and its format is shown in Table B. Similarly, for the coefficients, the output is stored using a default format by MLWIN. Table C details this format.

Table B: Structure of MLWIN output for variance-covariance matrix depending on the model estimated.

Row	Parameter	Meaning	Model A	Model B	Model C	Model D	Model E
1	$\sigma_{\mu 0}^2$	Variance of B_0	Value	Value	Value	Value	Value
2	$\sigma_{\mu 0, \mu 1}$	Covariance of B_0, B_1	Value	Value	Value	Value*	Value*
3	$\sigma_{\mu 1}^2$	Variance of B_1	Value	Value	Value	0	0
4	$\sigma_{\mu 0, \mu 2}$	Covariance of B_0, B_2	Value	Value*	Value*	0	0
5	$\sigma_{\mu 1, \mu 2}$	Covariance of B_1, B_2	Value	0	0	0	0
6	$\sigma_{\mu 2}^2$	Variance of B_2	Value	0	0	0	0
7	ε	Variance of error term	Value	0	0	0	0

* Value corresponds to the variance of the error term in these cases.

2. From Table C below, it is important to note that there is not a specific coefficients for the institutional (μ_{ij}) and individual (ε_{ij}) terms of the regression. These are specific per institution and exam record in subject.

Table C: Structure of MLWIN output for coefficients depending on model estimated.

Row	Parameter	Meaning	Model A	Model B	Model C	Model D	Model E
1	B_0	Intercept	Value	Value	Value	Value	Value
2	B_1	PRIORC	Value	Value	Value	Value	Value
3	B_2	PRIORC square	Value	Value	Value	Value	Value
4	B_3	PRIORC cubic	Value	Value	Value	Value	0
5	B_4	PRIORC quartic	Value	0	0	0	0

Stage IV - Transposition of coefficients

Input: Annex B Table 4 and Table 5.

Processes:

- calculating key coefficients and variance-covariance matrix (section 11)

Output: Annex B Table 6.

Stage IV - Calculating VA components

This stage specifies the calculation of two components of VA using the national coefficients and the MLM catalysts.

Section 11 – Calculation of key coefficients and variance-covariance matrix

- Once the base dataset is ready, it is necessary to set up the elements that are used to calculate the VA scores.

Inputs/ outputs/ transients

$C, C_2, C_3, C_4, S_2, S_3, S_4$	Catalyst values taken from the supplied MLM catalyst file
$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$	MLWIN Coefficients, as generated by the MLM unit and provided in MLM compiled file
S^*	Variance – covariance matrix from MLM compiled file
$\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4$	Key Coefficients, to be derived from MLWIN coefficients.
S	National Covariance Matrix to be derived from MLM catalyst values and variance-covariance matrix

Calculation of national covariance matrix

- Construct matrix A, using catalyst values, denoted by the following matrix structure

$$A = \begin{pmatrix} 1 & 0 & 0 \\ -c & 1 & 0 \\ \frac{(c^2 - c_2)}{s_2} & \frac{-2c}{s_2} & \frac{1}{s_2} \end{pmatrix}$$

- Calculate structure of S (national covariance matrix) using the following equation

$$S = A^T S^* A \quad ^4$$

where S^* is the variance-covariance matrix from the MLWIN for a given qualification / subject, the structure of which is denoted by the following matrix structure⁵

⁴ Throughout this document super-script T denotes Transpose of a matrix.

⁵ Note this is a symmetric matrix

$$S^* = \begin{pmatrix} VARB0 & COVARB01 & COVARB02 \\ COVARB01 & VARB1 & COVARB12 \\ COVARB02 & COVARB12 & VARB2 \end{pmatrix}$$

4. Values for variables in the matrix are taken from the MLM compiled file as supplied by the MLM unit.

Calculation of key coefficients

5. Calculation of each of the key coefficients $\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4$ is given by the following set of equations

$$\gamma_0 = \beta_0 - \beta_1 c + \frac{\beta_2}{s_2} (c^2 - c_2) + \frac{\beta_3}{s_3} (-c^3 - c_3) + \frac{\beta_4}{s_4} (c^4 - c_4)$$

$$\gamma_1 = \beta_1 - 2c \frac{\beta_2}{s_2} + 3c^2 \frac{\beta_3}{s_3} - 4c^3 \frac{\beta_4}{s_4}$$

$$\gamma_2 = \frac{\beta_2}{s_2} - 3c \frac{\beta_3}{s_3} + 6c^2 \frac{\beta_4}{s_4}$$

$$\gamma_3 = \frac{\beta_3}{s_3} - 4c \frac{\beta_4}{s_4}$$

$$\gamma_4 = \frac{\beta_4}{s_4}$$

6. The calculation of the key coefficients and corresponding national covariance matrix takes place for each record set within the MLM compiled file. This corresponds to each qualification/subject pairing.

Stage V - Calculation of VA scores

Input: Annex B Table 2.

Processes:

- calculate exam VA scores per exam record (section 12)
- set up data file after calculating VA scores (section 13)
- calculate institution VA scores per subject qualification (section 14)
(including calculation of 'shrinkage factor')

Output: Annex B Table 7 and Table 8.

Stage V - Calculating VA scores

Section 12 - Calculation of individual qualification aim VA scores

1. The calculation of an individual exam-level VA score is relatively straightforward, involving calculating the expected attainment for that individual in that subject/ qualification, given their institution membership and prior attainment and then subtracting this expected attainment from the individual's actual attainment for that subject in that qualification.

Inputs/ outputs/ transients

$\Gamma = \gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4$	Key coefficients for given qualification/ subject at national level
S	National covariance matrix
σ^2	National variance of error, taken from MLM compiled file
n	Number of exam records in subject/ qualification per provider
ϕ	Variance ratio per subject/ qualification per provider
λ	Shrinkage factor per subject/ qualification per provider
x_n record	Prior attainment for n_{th} exam
\underline{u}	Array of exam record VA scores
ψ	Standard deviation per subject/ qualification per provider
τ^2	Scalar variance of subject/ qualification per provider (intermediate value)

Calculation of Individual exam record VA scores

2. The calculation of the individual exam record VA scores is realised through the following equation

$$\underline{u} = Y - (\Gamma X^T)^T$$

3. \underline{u} is a $n \times 1$ matrix of the individual exam record VA scores and Γ is a 1×5 matrix containing the five key coefficients for this particular qualification / subject. Y is an $n \times 1$ matrix of the attainment of each exam record for this particular qualification / subject and X is an $n \times 5$ matrix, where each row of X is given

$$(1, x_n, x_n^2, x_n^3, x_n^4)$$

For example,

$$Y = \begin{pmatrix} 210 \\ 240 \\ 270 \end{pmatrix}, \quad X = \begin{pmatrix} 1 & 45 & 2025 & 91125 & 4100625 \\ 1 & 40 & 1600 & x_n^3 & x_n^4 \\ 1 & x_n & x_n^2 & x_n^3 & x_n^4 \end{pmatrix}$$

4. Therefore an example of \underline{u} would be

$$\underline{u} = \begin{pmatrix} 41.125 \\ -10.8 \\ 46.7 \end{pmatrix}$$

Section 13 - Data output file after calculating VA scores

1. Table D contains an example of the output after the calculation of VA scores has taken place. Table 7 in annex B describes the structure.

Table D: Example of Base VA scores file

ID	CAND	QUAL	SUBJ	POINTS	PRIOR	GENDER	INST	SURNAME	FORNAME	VA	UPIN	VA
1	1001	111	1010	270	45.00	0	9234567	Smith	Joe	41.12	100101	41.12
2	1001	111	1030	240	45.00	0	9234567	Smith	Joe	3.4	100101	3.4
3	1001	121	2210	120	45.00	0	9234567	Smith	Joe	-30.3	100101	-30.3
4	1002	121	2210	105	40.00	1	9234567	Smith	Jane	0.4	100101	0.4
5	1002	111	1010	210	40.00	1	9234567	Smith	Jane	-10.80	100101	-10.80
6	1002	111	1850	270	40.00	1	9234567	Smith	Jane	10.4	100101	10.4
7	1003	111	1010	270	41.50	0	9234567	Smith	John	46.7	100101	46.7

Note: rows 1, 5 and 7 in Table D contain the same data as shown in the example of section 12.

Section 14 - Calculation of institution VA score per subject/ qualification

1. The VA score at the level of subject per qualification for each institution is the average of the exam records' VA scores weighted by the number of exam records in that subject/ qualification and the corresponding shrinkage factor. All these calculations must be done at institutional level, that is, separately per institution.

Calculation of institution VA score for a qualification / subject

2. Utilising the matrix of individual exam record VA scores the overall institution level VA score for the given qualification / subject is calculated as follows.

3. Firstly, calculate the average VA score

$$VA_{avg} = \frac{\sum_1^n \underline{u}_n}{n}$$

\underline{u}_n denotes the VA score of the n_{th} exam record.

4. Next, calculate the average prior attainment x_{avg}

$$x_{avg} = \frac{\sum_1^n x_n}{n}$$

5. The shrinkage factor λ is then given by the following.

Calculate an intermediate value τ^2 given by

$$\tau^2 = \underline{x} \cdot S \cdot \underline{x}^T$$

where \underline{x} is defined as $(1, x_{avg}, x_{avg}^2)$ and S is the National Covariance matrix for the particular qualification / subject for which the calculations are being performed.

Then calculate the variance ratio ϕ given by

$$\phi = \sigma^2 / \tau^2$$

Finally, calculate the shrinkage factor λ

$$\lambda = n / (n + \phi)$$

6. The shrinkage factor is an element that controls the volatility of the VA scores at exam level when aggregated in small cohorts. By doing this, the VA score at subject level is unbiased and more robust for small groups. Large and normal size cohorts do not have their VA scores affected by the shrinkage factor.

7. The overall institution VA score U for the given qualification / subject is then given by

$$U = VA_{avg} \times \lambda$$

Calculation of confidence intervals about institution VA score for a qualification / subject

8. As the institution VA score for each subject in each qualification is a mean, then the standard error is used to calculate the confidence intervals.

The standard error of U is given by

$$\psi = \sqrt{\sigma^2 / (n + \phi)}$$

9. The 95% confidence intervals are given by

$$U \pm 1.96\psi$$

Stage VI - Calculation of Institution VA curve

Input: Annex B Table 7 and Table 8.

Processes:

- calculation of the institution VA curve per subject/ qualification (section 15)
- calculation of the confidence intervals for the institution VA curve (section 16)

Output: LAT output charts (example attached in annex C).

Stage VI - Calculation of institution VA curve

Section 15 - Calculation of the VA curve for an institution per subject/ qualification

1. The institution VA curve is a quadratic curve which, when added to the national curve, represents the expected levels of attainment at the institution.

Inputs/ outputs/ transients

n Number of exam records per subject/ qualification per provider

M $n \times 3$ matrix

S National covariance matrix

I Identity matrix

σ^2 National variance of error

\underline{u} Array of exam record VA scores

x_n Prior attainment for n_{th} exam record

exp Attain Matrix of expected attainment

$\Gamma = \gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4$ Key coefficients for given qualification/ subject at national level

2. In order to calculate the institution VA curve for a given qualification/ subject the following process is adhered to.

Firstly, construct a $n \times 3$ matrix denoted by M where each row is given by

$$\left(1, x_n, x_n^2\right)$$

Then perform the intermediate calculation of a variable P given by

$$P = MSM^T + \sigma^2 I$$

The identity matrix takes the following form

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

3. Next, the matrix of expected attainment $\underline{\text{exp Attain}}$ is calculated as follows

$$\underline{\text{exp Attain}} = (MS)^T P^{-1} \underline{u}$$

4. It is now possible to calculate the national and provider curves.

Define two values, $\underline{\text{priorAttain}}(x)$ and $\underline{\text{priorAttain}}_4(x)$ as follows

$$\underline{\text{priorAttain}}(x) = (1, x, x^2)$$

$$\underline{\text{priorAttain}}_4(x) = (1, x, x^2, x^3, x^4)$$

where x denotes a point on the x-axis of a chart. For the purposes of the calculation the values of x range from 0 to 58 in increments of 0.25.

5. Therefore, in order to calculate the full curves for the given qualification / subject one must calculate the following values for each x point.

$$Nat(x) = \underline{\text{priorAttain}}_4(x) \cdot \Gamma$$

where $Nat(x)$ represents the National curve for the given qualification / subject.

6. The Provider curve $Pro(x)$ is given by the following

$$Pro(x) = (\underline{\text{priorAttain}}(x) \cdot \underline{\text{exp Attain}}) + Nat(x)$$

Section 16: Calculation of confidence intervals for the institution VA curve

1. The calculation of the confidence intervals is directly derived from the vectors used to generate the institution curve and use the variation associated with the national and institution curves.

Inputs/ outputs/ transients

P	Variable as calculated for VA curve
S	National covariance matrix
M	Matrix as calculated for VA curve
Q	Covariance matrix for coefficients of <u>$\exp Attain$</u>
<u>$priorAttain(x)$</u>	As calculated for VA curve
ψ	Variance per subject/ qualification per provider

1. Firstly, the values for the matrix Q are calculated as follows

$$Q = S - (MS)^T P^{-1} (MS)$$

2. The standard deviation of $Pr o(x)$ is then given by

$$\psi = \sqrt{\left(\underline{priorAttain(x)} \cdot Q \cdot \underline{priorAttain(x)}^T \right)}$$

3. Such that the 95% confidence intervals for $Pr o(x)$ are

$$Pr o(x) \pm 1.96\psi$$

Stage VII - Aggregation of VA scores

Input: Annex B Tables 7 and 8.

Processes:

- calculation of institution VA scores at qualification and institution level (section 17)
- calculation of VA scores at SSA within qualification and SSA across qualifications (section 18)
- calculation of the significance of institution VA scores (section 19)
- specification of aggregated VA output file (section 20)

Output: Annex B Table 9 and Table 10 and Table 11.

Stage VII - Aggregation of VA scores

Section 17 - Calculation of institution VA score at qualification and institution level

1. To calculate an institutional VA score for across subjects in a qualification, V, it is necessary to combine the results of two or more subsets. The standard formula to combine variances is $\text{Var}(a) + \text{Var}(b) + 2\text{Cov}(a,b)$. However, the fact that the number of aims per subject, per provider is not uniform across different subjects, even within the same qualification, implies that it is not very feasible to estimate the covariance of the scores for any subject combination. Given that there is a correlation between the subject variance, it is recommended that this approach is not used if the covariance cannot be estimated. Otherwise, the risk of incurring a type one error increases substantially. Therefore, the more conservative approach was employed using the pooled estimate of variance.

Note: The following calculations must be done independently per provider.

Inputs/ outputs/ transients

$VA_{QualSubj}$	VA score for a particular subject within a given qualification at institutional level
VA_{Qual}	Aggregate VA score across subjects within a given qualification at institutional level
$n_{ExamQualSubj}$	Number of exams for a particular qualification/ subject at institutional level
$n_{ExamQuals}$	Number of exams across all subjects within a given qualification at institutional level
$\psi_{QualSubj}$	Standard error for given qualification/ subject at institutional level
$TotalSubjs$	Total number of subjects within the institution

2. The VA aggregate score for a given qualification is given by

$$VA_{Qual} = \frac{\sum_1^{TotalSubjs} VA_{QualSubj} \cdot n_{ExamQualSubj}}{n_{ExamQuals}}$$

3. The corresponding standard error for VA_{qual} is given by

$$\psi_{VA_{Qual}} = \sqrt{\sum_1^{n_{ExamQualSubj}} \left(\frac{n_{ExamQualSubj}}{n_{ExamQuals}} \right) \cdot \psi_{QualSubj}^2}$$

4. The 95% confidence intervals are then given by

$$VA_{Qual} \pm 1.96\psi_{VA_{qual}}$$

5. The production of an overall Institution VA score required the aggregation of VA scores across qualifications.

Inputs/ outputs/ transients

$VA_{AcrossQual}$ Aggregate VA score across all qualifications per institution.

Vol_{Qual} Volume for given qualification

$n_{ExamQuals}$ Number of exams within a qualification in that institution

$TotalQuals$ Total number of qualifications within the institution

6. The VA aggregate score across all qualifications is given by

$$VA_{AcrossQuals} = \frac{\sum_1^{TotalQuals} VA_{qual} \cdot n_{ExamQuals}}{\sum_1^{TotalQuals} n_{ExamQuals} \cdot Vol_{Qual}}$$

7. The corresponding standard error for $VA_{AcrossQual}$ is given by

$$\psi_{VA_{AcrossQuals}} = \sqrt{\sum_1^{TotalQuals} \psi_{VA_{qual}}^2 \cdot \left(\frac{n_{ExamQuals} \cdot Vol_{Qual}}{\sum_1^{TotalQuals} n_{ExamQuals} \cdot Vol_{Qual}} \right)}$$

8. The 95% confidence intervals are then given by

$$VA_{AcrossQuals} \pm 1.96\psi_{VA_{AcrossQual}}$$

Section 18 - Calculation of VA score at SSA within qualification and SSA across qualifications

1. The aggregation of scores within SSA within a qualification is built up from the estimation of the VA scores per subject per qualification. The main difference is that subject scores are grouped by SSA within qualification, and thus they are added and weighted differently than from previous calculations. The SSA subscript indicates the corresponding SSA that the subject belongs to.

Calculation of VA scores at SSA level within qualification

Inputs/ outputs/ transients

$VA_{QualSubjSSA}$	VA score for a particular subject within a given subject qualification in an SSA per provider
$VA_{QualSSA}$	Aggregate VA score across subjects within a given qualification in a specific SSA per provider
$n_{ExamQualSubjSSA}$	Number of exams for a particular qualification/ subject in a given SSA per provider
$n_{ExamQualsSSA}$	Number of exams across all subjects within a given qualification within a SSA per provider
$\psi_{QualSubjSSA}$	standard error for given qualification / subject in a specific SSA per provider
$TotalSubjsSSA$	Total number of subjects at SSA level within a qualification
$TotalSSA$	Total number of qualifications within an SSA

2. The VA aggregate score for a given qualification in a specific SSA is given by

$$VA_{QualSSA} = \frac{\sum_1^{TotalSubjsSSA} VA_{QualSubjSSA} \cdot n_{ExamQualSubjSSA}}{n_{ExamQualsSSA}}$$

3. The corresponding standard error for $VA_{qualSSA}$ is given by

$$\psi_{VA_{QualSSA}} = \sqrt{\sum_1^{TotalSubjsSSA} \left(\frac{n_{ExamQualSubjSSA}}{n_{ExamQualsSSA}} \right) \cdot \psi_{QualSubjSSA}^2}$$

4. The 95% confidence intervals are then given by

$$VA_{QualSSA} \pm 1.96\psi_{VA_{qualSSA}}$$

5. The aggregation of scores across qualifications within a SSA is very similar to those detailed previously. The main difference lays in the fact that subjects are grouped following the classification of the SSA.

Calculation of VA scores at SSA level across qualifications

Inputs/ outputs/ transients

$VA_{SSA_{acrossQual}}$	Aggregate VA score across all qualifications within a SSA
$VA_{QualSSA}$	Aggregate VA score across subjects within a given qualification in a specific SSA
Vol_{Qual}	Volume for given qualification
$n_{ExamQualsSSA}$	Number of exams within a qualification in a SSA
$TotalSSA$	Total number of qualifications within the SSA

6. The VA aggregate score across all qualifications is given by

$$VA_{SSA_{acrossQuals}} = \frac{\sum_1^{TotalSSA} VA_{qualSSA} \cdot n_{ExamQualsSSA}}{\sum_1^{TotalSSA} n_{ExamQualsSSA} \cdot Vol_{Qual}}$$

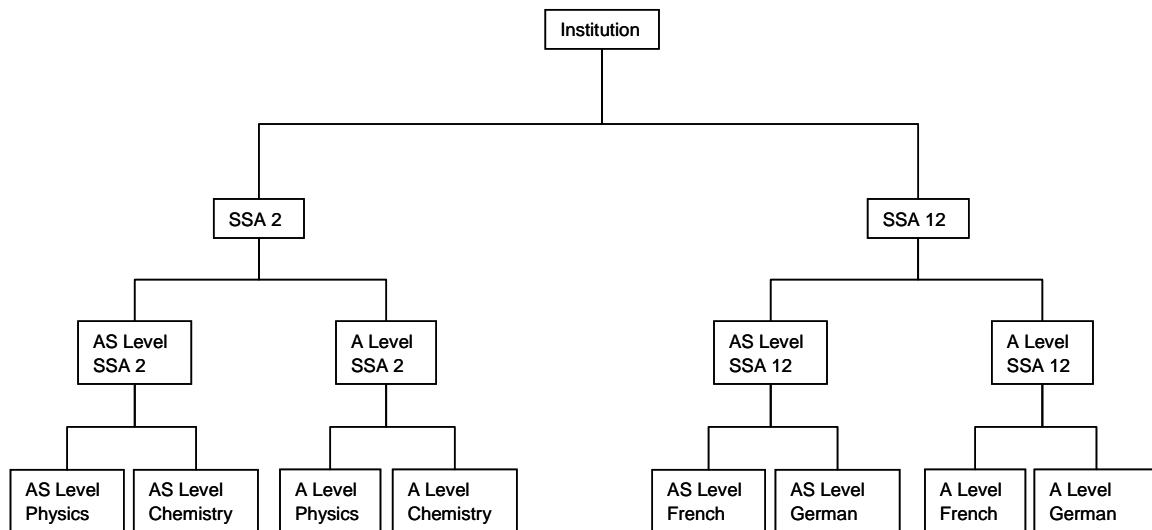
7. The corresponding standard error for $VA_{SSA_{acrossQual}}$ is given by

$$\psi_{VA_{SSA_{acrossQuals}}} = \sqrt{\sum_1^{TotalSSA} \psi_{VA_{qualSSA}}^2 \cdot \left(\frac{n_{ExamQualsSSA} \cdot Vol_{Qual}}{\sum_1^{TotalSSA} n_{ExamQualsSSA} \cdot Vol_{Qual}} \right)}$$

8. The 95% confidence intervals are then given by

$$VA_{SSA_{acrossQuals}} \pm 1.96\psi_{VA_{SSA_{acrossQuals}}}$$

9 It is very important to have a clear understanding of the way the data are organized when estimating SSA VA scores. For this, the following graph illustrates a simple example that can be useful.



10 Each box represents a VA score at different levels. The institutional VA score should **not** be calculated using an aggregation of SSA at different qualification levels. This will produce a VA score different from the score that is currently being estimated. In this sample the box of institutional VA score is presented to illustrate the organization of data levels. When producing summary charts at SSA level, the institutional VA score must be the same one that the LAT already estimates.

Section 19 - Calculation of the significance of institution VA scores

Significance at a qualification/ subject level

1. Performance of an institution significantly below the national average in a subject can be defined as

$$U < 0 \ \& \ (U + 1.96\psi) < 0$$

2. Performance of an institution significantly above the national average in a subject can be defined as

$$U > 0 \ \& \ (U - 1.96\psi) > 0$$

Significance at a qualification level

3. Performance of an institution significantly below the national average in a qualification can be defined as

$$VA_{Qual} < 0 \ \& \ (VA_{Qual} + 1.96\psi_{VA_{Qual}}) < 0$$

4. Performance of an institution significantly above the national average in a qualification can be defined as

$$VA_{Qual} > 0 \ \& \ (VA_{Qual} - 1.96\psi_{VA_{Qual}}) > 0$$

Section 20 - Specification of aggregated VA output file

Data output for aggregated VA scores

1. There are four data files containing VA scores. The one containing exam level VA scores for each subject qualification (that is, base VA scores) is specified in section 12.

The second file contains the Institution aggregate VA scores, with relevant confidence intervals:

- VA scores aggregated within subject/ qualification
- VA scores aggregated within qualifications
- VA scores aggregated across qualifications

The third file contains only the Institution aggregate VA score (that is, VA score aggregated across qualifications), with relevant confidence intervals.

The fourth file contains VA scores aggregated by SSA,

2. The format for aggregate VA score file has higher order aggregations repeated for each subject per qualification level record, as given in Table 9 in annex B.
3. The format of the Institution aggregate VA score is given in Table 10 in annex B.
4. The format of the SSA aggregate VA score is given in Table 11 in Annex B.

Annex A - Modelling rules for MLM

According to the number of observations per subject/qualification that is going to be available, certain rules may apply. It is up to the corresponding analyst to determine whether a full model can be carried on, or whether the rules would apply.

The rules are:

A. If file contains more than 5000 observations, run a full model:

$$y_{ij} = \beta_{oij} + \beta_{1j}x_{ij} + \beta_{2j}x_{ij}^2 + \beta_{3j}x_{ij}^3 + \beta_{4j}x_{ij}^4$$

$$\beta_{oij} = \beta_o + \mu_{0j} + \varepsilon_{ij}$$

$$\beta_{1j} = \beta_1 + \mu_{1j}$$

$$\beta_{2j} = \beta_2 + \mu_{2j},$$

$$\begin{bmatrix} \mu_{0j} \\ \mu_{1j} \\ \mu_{2j} \end{bmatrix} \sim N(0, \Omega_\mu): \Omega_\mu = \begin{bmatrix} \sigma_{\mu 0}^2 & \sigma_{\mu 01} & \sigma_{\mu 02} \\ \sigma_{\mu 01} & \sigma_{\mu 1}^2 & \sigma_{\mu 12} \\ \sigma_{\mu 02} & \sigma_{\mu 12} & \sigma_{\mu 2}^2 \end{bmatrix}$$

$$[\varepsilon_{ij}] \sim N(0, \Omega_\varepsilon): \Omega_\varepsilon = [\sigma_\varepsilon^2]$$

In this case the minimal number of iterations that have to be done must be equal or larger than the largest Nhat shown in the corresponding diagnostics for the variances. Also an analysis of the MCMC standard error might be useful.

B. If the file contains between 1000 and 5000 observations, then the model is:

$$y_{ij} = \beta_{oij} + \beta_{1j}x_{ij} + \beta_{2j}x_{ij}^2 + \beta_{3j}x_{ij}^3$$

$$\beta_{oij} = \beta_o + \mu_{0j} + \varepsilon_{ij}$$

$$\beta_{1j} = \beta_1 + \mu_{1j}$$

$$\begin{bmatrix} \mu_{0j} \\ \mu_{1j} \end{bmatrix} \sim N(0, \Omega_\mu): \Omega_\mu = \begin{bmatrix} \sigma_{\mu 0}^2 & \sigma_{\mu 01} \\ \sigma_{\mu 01} & \sigma_{\mu 1}^2 \end{bmatrix}$$

$$[\varepsilon_{ij}] \sim N(0, \Omega_\varepsilon): \Omega_\varepsilon = [\sigma_\varepsilon^2]$$

C. If the file contains between 500 and 1000 observations, and the total number of observations is larger or equal than 8 times the number of institutions, then the model is identical to the one above. In order to determine this it is necessary to click on [Model/ Hierarchy viewer] on MLWIN.

D. If the file contains between 500 and 1000 observations, and the total number of observations is less than 8 times the number of institutions, then the model is as follows:

$$y_{ij} = \beta_{oij} + \beta_1 x_{ij} + \beta_2 x_{ij}^2 + \beta_3 x_{ij}^3$$

$$\beta_{oij} = \beta_o + \mu_{0j} + \varepsilon_{ij}$$

$$[\mu_{0j}] \sim N(0, \Omega_\mu): \Omega_\mu = [\sigma_{\mu 0}^2]$$

$$[\varepsilon_{ij}] \sim N(0, \Omega_\varepsilon): \Omega_\varepsilon = [\sigma_\varepsilon^2]$$

E. If the file contains between 80 and 500 observations, then the model is:

$$y_{ij} = \beta_{oij} + \beta_1 x_{ij} + \beta_2 x_{ij}^2$$

$$\beta_{oij} = \beta_o + \mu_{0j} + \varepsilon_{ij}$$

$$[\mu_{0j}] \sim N(0, \Omega_\mu): \Omega_\mu = [\sigma_{\mu 0}^2]$$

$$[\varepsilon_{ij}] \sim N(0, \Omega_\varepsilon): \Omega_\varepsilon = [\sigma_\varepsilon^2]$$

F. If the file contains less than 80 observations, then it is not possible to estimate this model.

Annex B - Data tables used in VA calculations

Stage I – Data preparation

Input: Data processed by matching contractor

Processes:

- calculation of prior attainment (section 1)
- calculation of output attainment (section 1)
- selection of learners/ exams to include in calculation (section 2)
- production of data set for calculations of coefficients (section3)

Output: Table 1.

Stage II – Specification of data subsets

Input: Table 1.

Processes:

- selection of data subsets (section 4)
- specification of data for MLWIN (section 5)
- calculation of variables for MLWIN (section 5)
- storing and saving catalyst file (section 6)
- grouping of subjects (section 7)

Output: Table 2, Table 3 and Table 4.

Stage III– MLM estimation

Input: Table 3.

Processes:

- creation of master file (section 8)
- creation of subsets for subject/qualification (section 8)
- defining equations (section 9)
- running MLM regressions (section 9)
- output from MLM (section 10)

Output: Table 5.

Stage IV – Transposition of coefficients

Input: Table 4 and Table 5.

Processes:

- calculating key coefficients and variance covariance matrix (section 11)

Output: Table 6.

Stage V – Calculation of VA scores

Input: Table 2.

Processes:

- calculate exam VA scores per exam record (section 12)
- set up data file after calculating VA scores (section 13)
- calculate institution VA scores per subject qualification (section 14) (including calculation of ‘shrinkage factor’)

Output: Table 7 and Table 8.

Stage VI - Calculation of Institution VA curve

Input: Table 7 and Table 8.

Processes:

- calculation of the institution VA curve per subject/ qualification (section 15)
- calculation of the confidence intervals for the institution VA curve (section 16)

Output: LAT output charts (example attached in annex C).

Stage VII - Aggregation of VA scores

Input: Table 7 and Table 8.

Processes:

- calculation of institution VA scores at qualification and institution level (section 17)
- calculation of VA scores at SSA within qualification and SSA across qualifications (section 18)
- calculation of the significance of institution VA scores (section 19)
- specification of aggregated VA output file (section 20)

Output: Tables 9 and 10.

Output of stage I - Data preparation

Table 1: Format of the exam level file

<u>-Field Reference</u>	<u>Field Description</u>	<u>Character Position</u>	<u>Length</u>	<u>Possible Values</u>
RECORDID	Serial number of record.	1-9	9	
LEAESTAB	Latest main DfES number	10-16	7	
CANDNO	Candidate serial number	17-23	7	
EXAMNO	Exam/Assessment serial number	24-25	2	
SUBLEVNO	Qualification and assessment code	26-28	3	
GNUMBER	G number (for GNVQs only)/QAN number	29-36	8	
MAPPING	LEAP subject mapping/LDCS code	37-40	4	
SUBJECT	Numerical subject mapping for VA	41-45	5	
SURNAME	Surname	46-95	50	
FORENAME	Forename(s)	96-145	50	
MATCHREF	Matching Contractor Candidate Reference	146-156	11	
EXAMYEAR	Exam/Assessment year	157-161	5	
SEASON	Exam/Assessment season	162	1	
YEARGRP	Year group	163-164	2	01-17, 99
GENDER	Gender	165	1	M or F or blank
GRADE	Grade/Level	166-167	2	Grade
DISC2	Discount flag (two-year)	168	1	
POINTS	Point score for section 96 qualifications	169-173	5	9999.9
ASIZE	A level equivalent size	174-179	6	99.999
LEV3	Level 3 Threshold Contribution	180-182	3	999.9
AMDFLAGE	Exam Amendment flag	183-184	2	
AMDFLAGP	Learner Amendment flag	185-186	2	
AVKS4	Prior attainment score	187-193	7	to 3dp
PTS4	Total Level 2 and below prior attainment	194-200	7	to 3dp
SZEKS4	Total size of qualifications include in prior attainment	201-206	6	99.999
NATINC	National line inclusion flag	207-208	2	0 or 1
INVA	Learner is included in VA calculation	209-210	2	0 or 1
PREVYEAR	VA Qualification taken in previous year	211-212	2	0 or 1
PTRIG	Pilot trigger criteria	213-214	2	0 or 1
PNATRES	Learner in an institution that contribute to National average	215-216	2	0 or 1
PSCHRES	Learner in an institution that is published in post 16 tables	217-218	2	0 or 1
PLEARES	Learner in an institution that contribute to LEA average	219-220	2	0 or 1
PLSCRES	Learner in an institution that contribute to LSC average	221-222	2	0 or 1
ENDMARK	End of record marker	223	1	X

Output of stage II

Table 2: Fields of base data file for the LAT. (YYMMDD_VA_BASE_LAT.csv)

Field reference in Exam level File ⁶	Field Name	Description	Width	Type of Field
RECORDID	ID	Unique identifier for each record	15	Integer
CANDNO	CAND	DfES Candidate Serial Number	7	Integer
SUBLEVNO	QUAL	Qualification and Assessment Code	3	Integer
MAPPING	SUBJ	LEAP/ LDCS code	5	Integer
**	SUBJNAT	LEAP/ LDCS Subject Mapping code was used for all qualifications	5	Integer
POINTS	POINTS	Outcome attainment in QCA points	6	F6.2
AVKS4	PRIOR	Average QCA prior attainment score.	7	F7.4
GENDER	GENDER	0 is male and 1 is female	1	Integer
LEAESTAB	INST	DfES Latest institution number	7	Integer
SURNAME	SURNAME	Surname	50	Alpha – Numeric
FORNAME	FORNAME	Forename(s)	50	Alpha – Numeric
**	UPIN	Unique provider number	6	Integer
**	YEAR	Academic year of data	4	Integer

Table 3: MLM input data set (YYMMDD_VA_MLM_FILE.dat)

Exam level Title	Name	Original Width	Converted Width	Description
CANDNO	CAND	7	9	DfES Candidate Serial Number
LEAESTAB	INST	7	9	DfES Latest institution number
SUBLEVNO	QUAL	3	5	Qualification and Assessment Code
SUBJECT	SUBJ	5	7	Numerical subject mapping
GROUPING	SUBJMLM	5	7	Grouped subjects (when applicable)
***	PRIORC	7	9	Average QCA prior attainment score centred
***	PRIOR2C	***	10	Square average QCA prior attainment score centred and standardized
***	PRIOR3C	***	10	Cubic average QCA prior attainment score centred and standardized
***	PRIOR4C	***	10	Quartic average QCA prior attainment score centred and standardized
POINTS	POINTS	6	8	Outcome attainment in QCA points
RECORDID	ID	15	17	Unique identifier for each record

⁶ Field reference in Exam Level file – where the field has not originated directly from the supplied datafile (for example, it has been created or matched in) then this is shown by **

Table 4: MLM Catalyst file (YYMMDD_VA_CATA.csv)

Parameter	Description	Width	Type of Field
C	Average of <i>Prior</i>	10	F10.3
C2	Average of <i>Prior</i> ₂	10	F10.3
C3	Average of <i>Prior</i> ₃	10	F10.3
C4	Average of <i>Prior</i> ₄	10	F10.3
S2	Standard deviation of <i>Prior</i> ₂	10	F10.3
S3	Standard deviation of <i>Prior</i> ₃	10	F10.3
S4	Standard deviation of <i>Prior</i> ₄	10	F10.3
YEAR	Academic year of data	4	Integer

Output of stage III

Table 5: Structure of compiled output file (YYMMDD_VA_BETA.csv)

Field name	Field Description	Width	Type of Field
QUAL	Qualification and assessment code	3	Integer
MAPPING	LEAP subject mapping/ LDCS code	5	Integer
SUBJNAT	Numerical subject mapping used for MLM	5	Integer
VARB0	Variance of B_0	7	F7.3
COVARB01	Covariance of B_0, B_1	7	F7.3
VARB1	Variance of B_1	7	F7.3
COVARB02	Covariance of B_0, B_2	7	F7.3
COVARB12	Covariance of B_1, B_2	7	F7.3
VARB2	Variance of B_2	7	F7.3
VARERR	Variance of error term	7	F7.3
N	Number of exam records in model	6	Integer
B0	Intercept	7	F7.3
B1	Coefficient of PRIORC	7	F7.3
B2	Coefficient of PRIORC square	7	F7.3
B3	Coefficient of PRIORC cubic	7	F7.3
B4	Coefficient of PRIORC quartic	7	F7.3
YEAR	Academic year of data	4	Integer

Output of stage IV

Table 6: Structure of compiled output file of key coefficients (YYMMDD_VA_KEY_COEF.csv)

Field name	Field Description	Width	Type of Field
QUAL	Qualification and assessment code	3	Integer
MAPPING	LEAP subject mapping/LDCS code	5	Integer
SUBJNAT	Numerical subject mapping used for MLM	5	Integer
VARGAM0	Variance of γ_0	8	F8.5
COVGAM01	Covariance of γ_0, γ_1	8	F8.5
VARGAM1	Variance of γ_1	8	F8.5
COVGAM02	Covariance of γ_0, γ_2	8	F8.5
COVGAM12	Covariance of γ_1, γ_2	8	F8.5
VARGAM2	Variance of γ_2	8	F8.5
VARERR	Variance of error term	8	F8.5
N	Number of exam records in model	6	Integer
GAMMA0	Intercept γ_0	7	F7.3
GAMMA1	Coefficient of PRIORC (γ_1)	7	F7.3
GAMMA2	Coefficient of PRIORC square (γ_2)	7	F7.3
GAMMA3	Coefficient of PRIORC cubic (γ_3)	9	F9.6
GAMMA4	Coefficient of PRIORC quartic (γ_4)	9	F9.6
YEAR	Academic year of data	4	Integer

Output of Stage V

Table 7: Format of Base VA scores file (YYMMDD_VA_BASE_SCORES.csv)

Field name	Field Description	Width	Type of Field
ID	Unique identifier for each record	15	Integer
CAND	DfES Candidate Serial Number	7	Integer
QUAL	Qualification and Assessment Code	3	Integer
SUBJ	LEAP/ LDCS Subject Mapping code used for all qualifications	5	Integer
POINTS	Outcome attainment in QCA points	6	F6.2
PRIOR	Average QCA prior attainment score.	7	F7.4
GENDER	0 is male and 1 is female	1	Integer
INST	DfES Latest institution number	7	Integer
SURNAME	Surname	50	Alpha - Numeric
FORNAME	Forename(s)	50	Alpha - Numeric
VA	Raw VA score	6	F6.2
UPIN	Unique provider number	6	Integer
YEAR	Academic year of data	4	Integer

Table 8: Shrinkage factor (YYMMDD_SF.csv)

Field name	Field Description	Width	Type of Field
QUAL	Qualification and assessment code	3	Integer
SUBJ	Numerical subject mapping used for MLM	5	Integer
INST	DfES Latest institution number	9	Integer
UPIN	Unique provider number	6	Integer
LAMBDA	Shrinkage factor per institution/ subject/ qualification	10	F10.8
YEAR	Academic year of data	4	Integer

Output of stage VII

Table 9: Format of LAT aggregated output file (YYMMDD_VA_AGG_SCORE.csv)

Field name	Field Description	Width	Type of Field
INST	DfES Latest institution number	7	Integer
SUBJ	LEAP/ LDCS Subject Mapping code used for all qualifications	5	Integer
QUAL	Qualification and Assessment Code	3	Integer
SUBVA	Aggregate VA score for subject/ qualification combination	6	F6.2
SUBSD	Standard Deviation of Aggregate VA score for subject/ qualification combination	6	F6.2
SUBNAIMS	Number of exam records per subject/ qualification combination	6	Integer
UCISUB	Upper confidence interval per subject/ qualification combination at 95%	6	F6.1
LCISUB	Lower confidence interval per subject/ qualification combination at 95%	6	F6.1
SUBPRIOR	Average prior attainment per subject/ qualification combination	7	F7.4
QUALVA	Aggregate VA score for qualification	6	F6.2
QUALSD	Standard Deviation of aggregate VA score for qualification	6	F6.2
QUALNAIM	Number of exam records per qualification	6	Integer
NSUBS	Number of subjects per qualification	6	Integer
UCIQUAL	Upper confidence interval per qualification at 95%	6	F6.1
LCIQUAL	Lower confidence interval per qualification at 95%	6	F6.1
QUAPRIOR	Average prior attainment per qualification	7	F7.4
INSTVA	Aggregate VA score for institution	6	F6.2
INSTSD	Standard Deviation of aggregate VA score for institution	6	F6.2
INSTNAIM	Number of exam records per institution	6	Integer
NSUBSINS	Number of subjects per institution	3	Integer
NQUALS	Number of qualifications per institution	3	Integer
UCIINST	Upper confidence interval per institution at 95%	8	F8.4
LCIINST	Lower confidence interval per institution at 95%	8	F8.4
INSPRIOR	Average prior attainment per institution	7	F7.4
UPIN	Unique provider number	6	Integer
YEAR	Academic year of data	4	Integer

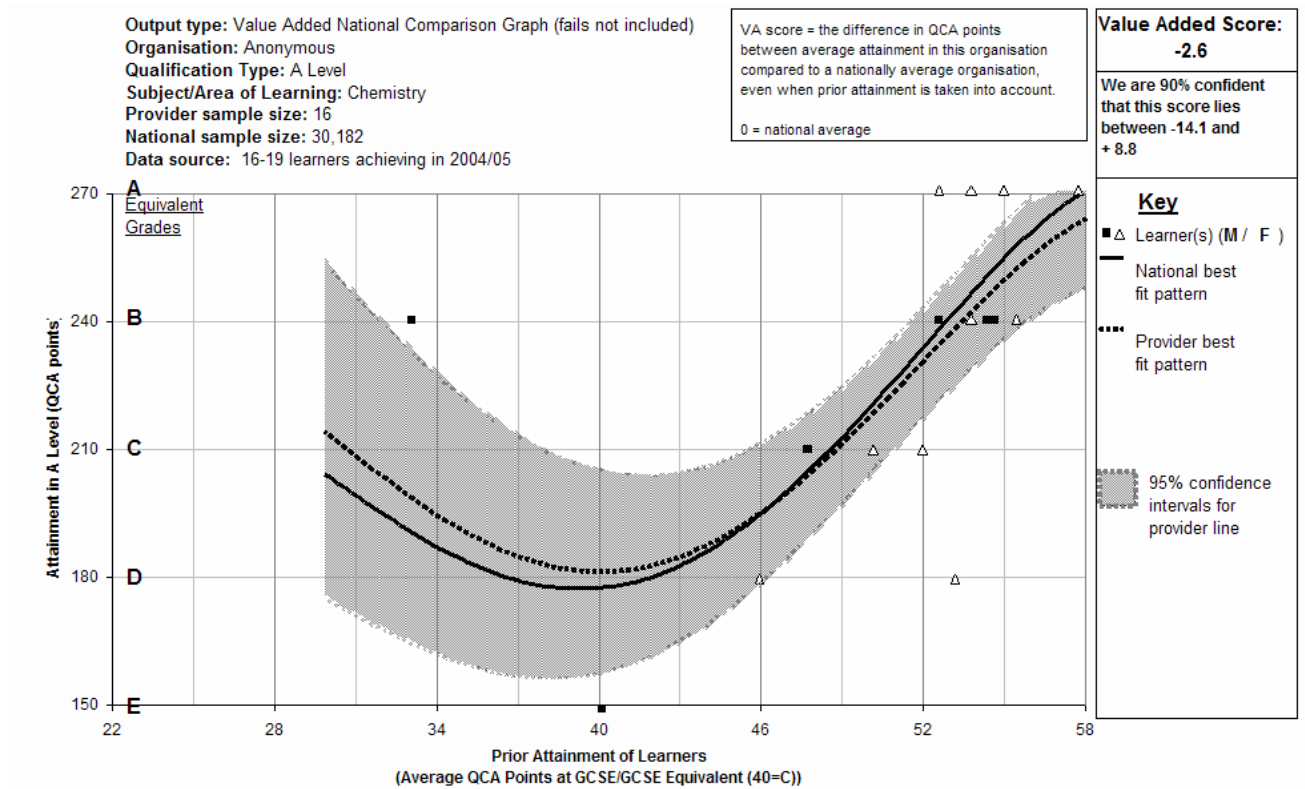
Table 10: Format of LAT aggregated output for inclusion in DfES Attainment and Achievement Tables (YYMMDD_VA_SCAAT.csv)

Variable	Name	Character Position	Length	Description	Type of Field
PROVDRID	Institution identifier	1-9	9	DfES Latest institution number	Numeric
INSTVA	inst VA	30-35	6	Aggregate VA score for institution	Numeric
UCIINST	Upper CI per inst	16-22	7	Upper confidence interval per institution at 95%	F7.1
LCIINST	Lower CI per inst	23-29	7	Lower confidence interval per institution at 95%	F7.1
INSTNAIM	N exam records per inst	10-15	6	Number of exam records per institution	Numeric
INSPRIOR	Avg prior per inst	36-42	7	Average prior attainment per institution	Numeric

Table 11: Format of LAT aggregated SSA output file (YYMMDD_VA_AGG_SSA.csv)

Field name	Field Description	Width	Type of Field
INST	DfES Latest number	7	Integer
QUAL	Qualification and Assessment Code	3	Integer
SSA	SSA Mapping code	5	Integer
SSAQVA	Aggregate VA score per SSA within qualification	6	F6.2
SSAQSD	SD agg VA score per SSA within qualification	6	F6.2
SSAQNAIM	Number of exam records per SSA within qualification	6	Integer
SSAQNSUB	Number of subjects per SSA within qualification	6	Integer
UCISSAQ	Upper confidence interval per SSA within qualification at 95%	6	F6.1
LCISSAQ	Lower confidence interval per SSA within qualification at 95%	6	F6.1
SSAQPRIOR	Average prior attainment per SSA within qualification	6	F6.2
SSAVA	Aggregate VA per SSA across qualifications	6	F6.2
SSASD	SD of agg VA per SSA across qualifications	6	F6.2
SSANAIM	Number of exam records per SSA across qualifications	6	Integer
SSANSUB	Number of subjects per SSA across qualifications	3	Integer
SSANQUAL	Number of qualifications per SSA across qualifications	3	Integer
UCISSA	Upper confidence interval per SSA across qualifications at 95%	6	F6.1
LCISSA	Lower confidence interval per SSA across qualifications at 95%	6	F6.1
SSAPRIOR	Average prior attainment per SSA across qualifications	6	F6.2
UPIN	UPIN Number	6	Integer

Annex C - Example of LAT Output Chart



© LSC March 2006
 Published by the Learning and Skills Council.

Extracts from this publication may be reproduced for non-commercial educational or training purposes on condition that the source is acknowledged and the findings are not misrepresented. This publication is available in electronic form on the Learning and Skills Council website: www.lsc.gov.uk

If you require this document in an alternative format or language, please contact the LSC Helpdesk.

LSC Helpdesk: 0870 900 6800
 Publication reference: LSC-P-NAT-060130