

July 2006/23

Core funding/operations

Assurance and funding

This report is for information and guidance

This document is a guide to HEFCE's web facility which generates several types of analysis from Higher Education Statistics Agency (HESA) student data. Use of the web facility will help higher education institutions to return accurate data to HESA, and to identify errors and forecasting discrepancies in HEFCE funding returns.

2005-06 statistics derived from HESA data

Guide to HEFCE web facility

Contents

Executive summary		4
Introduction		6
Annexes		
Annex A	Summary of changes since publication of '2004-05 statistics derived from HESA data: Guide to HEFCE web facility' (HEFCE 2005/32)	13
Annex B	Using the HEFCE web facility	16
Annex C	Comparison of HESES05 and the HESES05 re-creation	26
Annex D	Comparison of the HESES05 re-creation and the HESES05 re-creation based on cost centre sector norms for subjects	29
Annex E	Comparison of RAS05 and the RAS05 re-creation	33
Annex F	Derived statistics likely to inform the 2007-08 widening participation allocations	36
Annex G	Information for leads and members of HEFCE-recognised funding consortia	41
Annex H	Research degree qualification rates (RDQRs)	45
Annex I	HEFCE statistics used for publication and to inform policy	50
Annex J	HESES06 non-completion toolkit	53
	List of abbreviations	55
Technical appendices (see separate downloads)		
Appendix 1	HESES05 re-creation algorithms	57
Appendix 2	Troubleshooting the differences between HESES05 and the HESES05 re-creation	87
Appendix 3	Problems of fit with the HESES05 re-creation algorithms	100
Appendix 4	HESES05 re-creation based on cost centre sector norms for subjects algorithms	103
Appendix 5	Troubleshooting the differences between the HESES05 re-creation and the HESES05 re-creation based on cost centre sector norms for subjects	110
Appendix 6	Problems of fit with the HESES05 re-creation based on cost centre sector norms for subjects algorithms	113
Appendix 7	RAS05 re-creation algorithms	114
Appendix 8	Troubleshooting the differences between RAS05 and the RAS05 re-creation	129
Appendix 9	Problems of fit with the RAS05 re-creation algorithms	135
Appendix 10	2007-08 Widening participation (WP) allocation algorithms	138
Appendix 11	2005-06 RDQR algorithms	148
Appendix 12	Guidance for completing and submitting RDQR action plans	159
Appendix 13	Example RDQR action plan (see separate Excel download)	
Appendix 14	Submitting amendments to 2004-05 HESA data for the purpose of correcting RDQRs	164

Appendix 15	Submitting corrections to derived fields for the purpose of amending RDQRs	166
Appendix 16	HEFCE statistics used for publications and to inform policy algorithms	172
Appendix 17	HESES06 non-completion toolkit algorithms	191
Appendix 18	Index of derived fields	193

2005-06 statistics derived from HESA data

Guide to HEFCE web facility

To Heads of HEFCE-funded higher education institutions
Heads of universities in Northern Ireland

Of interest to those responsible for Student data, Audit, Finance

Reference 2006/23

Publication date July 2006

Enquiries to For enquiries regarding the HESES05 comparison, cost centre sector norms comparison or RAS05 comparison contact:

Lauren Humphrey tel 0117 931 7373

Peter McMahon tel 0117 931 7338

e-mail hesa_heses_stats@hefce.ac.uk

For enquiries regarding the use of HESA data to inform the 2007-08 widening participation allocations contact:

Anthony Ryan tel 0117 931 7297

e-mail hesa_heses_stats@hefce.ac.uk

For enquiries regarding the RDQRs contact:

Ben Grassby tel 0117 931 7493

e-mail hesa_heses_stats@hefce.ac.uk

For enquiries regarding the use of HESA data in HEFCE publications and to inform policy contact:

Lisa Readdy tel 0117 931 7492

e-mail hesa_heses_stats@hefce.ac.uk

Executive summary

Purpose

1. This document is a guide to HEFCE's web facility which institutions can use to check their 2005-06 Higher Education Statistics Agency (HESA) student data. It generates the following derived statistics:
 - a. Higher Education Students Early Statistics Survey 2005-06 (HESES05) re-creation.
 - b. HESES05 re-creation based on cost centre sector norms for subjects.
 - c. Research Activity Survey 2005 (RAS05) re-creation.
 - d. Derived statistics that are likely to be used to inform the 2007-08 widening participation allocations.
 - e. Research degree qualification rates (RDQRs).
 - f. Derived statistics that will be used in HEFCE publication.
 - g. Higher Education Students Early Statistics Survey 2006-07 (HESES06) non-completion toolkit.

Key points

2. The use of the web facility is strongly encouraged as it will assist institutions to:
 - a. Return accurate data to HESA.
 - b. Reduce the likelihood of selection for the derived statistics exercise for 2005-06 (being selected involves considerable additional work and potential funding adjustments).
 - c. Identify discrepancies between forecasts made in HESES05 and the outturn position for 2005-06.
 - d. Identify errors in HEFCE funding returns.
 - e. Make informed estimations of non-completion rates for HESES06.
 - f. Ensure RDQRs and regional statistics derived from the data returned to HESA are suitable for publication.
3. This document (including the technical appendices which need to be downloaded separately) provides:
 - a. Guidance for using the web facility.
 - b. Algorithms used in the production of the derived statistics.
 - c. Guidance on troubleshooting the differences between the HEFCE funding returns and the re-creations.
 - d. Details of problems of fit with algorithms.
 - e. Guidance for using the HESES06 non-completion toolkit.

Outputs for publication

4. This will be the last opportunity for institutions to verify the accuracy of the data underlying the RDQRs before they are asked to sign off the rates as accurate and as being

suitable for publication. Also, as in previous years, this facility will be the only opportunity to ensure that the regional statistics are suitable for publication.

Action required

5. Use of this web facility is optional, but we strongly encourage institutions to use it as part of their data quality processes. A review of the web facility has confirmed that it will be an essential element of the majority of higher education institutions' 2005-06 HESA data quality processes. This review also revealed that institutions that do not use the facility are more likely to be selected in the monitoring exercise described in paragraph 6.

Relationship with 'Statistics derived from HESA data'

6. Annually we launch the 'Statistics derived from HESA data for monitoring and allocation of funding' exercise in which we monitor institutions' HESES and Research Activity Survey (RAS) returns using HESA student data. Our funding allocations are informed by the data provided by institutions. If we find, either through reconciliations with HESA data, or any data audit, that data do not reflect the outturn position for the year, and that this has resulted in institutions receiving incorrect funding allocations, we will adjust their funding accordingly (subject to the appeals process and the availability of our funds). Any funding adjustments arising from the reconciliation of a re-creation of HESES05 from HESA 2005-06 student data (the HESES05 re-creation) with HESES05 or from the comparison of cost centre assignments with the sector norms for subjects, are likely to affect the funding previously announced for 2005-06 and all subsequent years. In recent years the vast majority of institutions selected for the monitoring exercise have had their funding adjusted because we have identified that their data returned to HEFCE did not reflect the outturn position for the year, and that this had resulted in them receiving incorrect funding allocations.

7. Any funding adjustments arising from the reconciliation of RAS05 with the re-creation of RAS05 from HESA 2005-06 student data (the RAS05 re-creation) are likely to affect the funding previously announced for 2006-07.

8. While the web facility is provided to complement the monitoring exercise it does not replace it.

Introduction

9. This document provides guidance on using the HEFCE's web facility and its outputs. The primary purpose of the web facility is to help higher education institutions to return accurate data to the Higher Education Statistics Agency (HESA). The web facility offers institutions an opportunity to identify, and therefore rectify, any errors in fields that affect the outputs generated by the web facility prior to the final submission of data.

10. Use of this facility, prior to submission of the HESA return, is strongly encouraged by both HESA and HEFCE as both organisations regard it as an essential element of all higher education institutions' data quality processes. The HEFCE web facility for 2005-06 statistics derived from HESA data is a separate utility to the HESA data collection system.

11. We believe that the introduction of the web facility has contributed to an improvement in HESA student data returns. As a result of this improved data quality we have been able to significantly reduce the size of the selection thresholds for the 'statistics derived from HESA data for monitoring and allocation of funding' exercise. We have found that institutions that use the web facility are less likely to be selected for the statistics derived from HESA data exercise.

12. Use of the web facility can also help identify errors and discrepancies between the forecasts made in the Higher Education Students Early Statistics Survey 2005-06 (HESES05) and the outturn position for 2005-06. Where discrepancies occur, we expect institutions to take full account of the outputs from the web facility when preparing future HESES returns. We encourage institutions to analyse the web facility outputs as part of their planning and audit processes.

13. Changes to the web facility since '2004-05 statistics derived from HESA data: Guide to HEFCE web facility' (HEFCE 2005/32) are described in Annex A. Annex B describes how to access and use the web facility.

14. The web facility generates the following types of output:

- re-creations of HEFCE funding returns
- derived statistics likely to inform HEFCE funding
- derived data for verification by institutions
- data quality tools.

Using the web facility

15. Typically institutions submit their data to the web facility and retrieve the resultant derived statistics several times before all identified errors are remedied. Therefore, we will not restrict institutions' use of the web facility. However, as resources at HEFCE are limited, at times when there is high demand, response times may be slower.

16. Institutions should include adequate time within their timetable to allow them to make full use of the web facility without jeopardising HESA quality arrangements or timetables.

Re-creations of HEFCE funding returns

HESES05 re-creation

17. The web facility generates a re-creation of HESES05 from HESA 2005-06 student data. This includes a re-calculated grant adjustment report produced by applying the same formulae that were used to calculate the grant adjustment report for HESES05. The HESES05 re-creation is generated using the methods detailed in Annex C.

18. We use the algorithms detailed in Appendix 1 to generate the HESES05 re-creation. These are the same algorithms that we intend to use in the 2005-06 statistics derived from HESA data exercise. We may however make changes to improve the algorithms.

19. We advise that institutions use this opportunity to identify the cause of all discrepancies between their HESES05 return and the HESES05 re-creation, so that where errors in HESA data are a cause, these can be corrected before submission to HESA. The removal of such errors in HESA data reduces the likelihood of selection for the 2005-06 statistics derived from HESA data exercise.

20. The HESES re-creation is generated using the assignment of cost centres that are returned by individual institutions. In addition to the HESES re-creation, the web facility generates a re-creation of HESES05 that uses a sector norm mapping rather than the cost centres returned on the HESA return by individual institutions. This sector norm mapping is based on how we have observed all institutions to have assigned cost centres to activity on their HESA 2004-05 student data. Further details on how we create the sector norm mapping and other information about this re-creation are contained in Annex D.

21. The 'HESES05 re-creation based on cost centre sector norms for subjects' includes a re-calculated grant adjustment report that is produced by applying the same formulae that were used to calculate the grant adjustment report for HESES05.

22. We use the algorithms detailed in Appendix 4 to generate the HESES05 re-creation based on cost centre sector norms for subjects. These are the same algorithms that we intend to use in the 2005-06 statistics derived from HESA data exercise, except that the sector norm mapping of subjects to cost centres will be revised on receipt of full HESA 2005-06 student data. In the absence of these data, and for the purpose of the web facility, we will use the mapping of subjects to cost centres generated from HESA 2004-05 student data, and we do not expect the mapping to change significantly.

RAS05 re-creation

23. The web facility generates a re-creation of Research Activity Survey 2005 (RAS05) from HESA 2005-06 student data. This includes a re-calculation of the 2006-07 research degree

programme (RDP) supervision fund produced by applying the same formulae that were used to calculate the research grant from RAS05. The RAS05 re-creation is generated using the methods detailed in Annex E.

24. We use the algorithms detailed in Appendix 7 to generate the RAS05 re-creation. These are the same algorithms that we intend to use in the 2005-06 statistics derived from HESA data exercise, except that the assignment of units of assessment (UOAs) to subject groups will be revised on receipt of full HESA 2005-06 staff data. In the absence of these data and for the purpose of the web facility we will use the assignment of UOAs to subject groups generated from HESA 2004-05 staff data; we do not expect the assignment to change significantly.

25. We advise institutions to use this opportunity to determine the cause of discrepancies between their RAS05 and the RAS05 re-creation, so that where errors in HESA data are a cause, these can be corrected before submission to HESA. The removal of such errors in HESA data reduces the likelihood of an institution being selected for the 2005-06 statistics derived from HESA data exercise.

Derived statistics likely to inform HEFCE funding

Derived statistics likely to inform 2007-08 WP allocations

26. It is likely that HESA 2005-06 student data will be used to inform the following widening participation (WP) allocations for 2007-08:

- widening access from disadvantaged backgrounds: full-time and part-time students
- improving retention: full-time
- students with disabilities

The derived statistics used to inform the 2007-08 WP allocations are likely to be generated using the methods detailed in Annex F.

Derived statistics used in HEFCE statistical publications

RDQRs

27. We will derive, from the HESA student data submitted to the web facility, RDQRs for students undertaking research degree programmes. In 2007 we intend to publish RDQRs for full-time students that commenced their research degree programmes in 1999-2000. We expect to publish RDQRs for later cohorts in future years, therefore institutions should verify the entire dataset. We will write to institutions in August 2006 to give further information about the process for signing off the RDQRs that we intend to publish in 2007.

28. This web facility provides institutions with a final opportunity to verify that their RDQRs are fit for publication. Annex H describes the methods used to produce RDQRs and outlines the steps institutions must take to remedy any issues with their RDQRs resulting from

underlying errors in the HESA student data. This process will vary depending on the year of the HESA return that contains the errors.

29. In order for institutions to remedy their RDQRs they must first provide an RDQR action plan. This action plan must be submitted to us at HEFCE by 8 September 2006. Appendix 12 gives details on completing and submitting this plan. We expect to make action plan templates available on the extranet for download from 1 August 2006.

30. Errors in HESA student data for the years 1999-2000 to 2003-04 will require correction through the submission of a historical data error file to enable us to correct the derived fields for the individual students affected. Appendix 15 gives details on producing and submitting historical data error files. The deadline for submitting historical data error files is 26 September 2006.

31. Errors in 2004-05 HESA student data will require institutions to submit a revised return to HESA. Appendix 14 gives details on submitting revised returns to HESA. The deadline for submitting amendments to 2004-05 HESA data for the purpose of correcting RDQRs is 26 September 2006.

32. Institutions are invited to comment on the sector rates that are provided as part of the RDQR outputs. Institutions should e-mail their comments to hesa_heses_feedback@hefce.ac.uk by 16 September 2006. The subject line of the e-mail should be 'RDQR: comments on sector rates'.

HEFCE regional statistics

33. We intend to publish regional statistics derived from 2005-06 HESA student data. As was the case last year, the web facility is the sole opportunity for institutions to verify the following derived statistics that will be used:

- students registered at one institution and taught by another institution
- campuses
- distance learning
- provision by location.

These data may be used for regional analysis and to inform policy decisions.

34. Details of the methods used to derive these statistics are given in Annex I.

HESES06 non-completion toolkit

35. The web facility will generate an Excel spreadsheet that is designed to assist institutions in estimating non-completion rates ready for inclusion in their HESES06 return. Annex J explains how to use it.

36. At this time it is unclear whether industrial action by academic staff unions over student assessment in 2005-06 has affected students' 2005-06 completion status. Where disruption has occurred, institutions are advised to use the outputs of the toolkit with caution as the non-completion rates for some courses may be inaccurate and atypical. For such courses we would expect institutions to base their HESES06 forecasts on the rates applied to HESES05, adjusting for any known errors in these rates.

Amendments to 2005-06 HESA data

37. Each institution's procedures for the quality assurance of HESA data must take place before it signs off the HESA data as correct. The web facility will complement any other quality assurance processes the institution undertakes. However, we recognise that HESA returns are necessarily complicated, and, exceptionally, we will allow corrections of errors identified after the HESA data collection has closed to enable the recalculation of funding.

Data confidentiality

38. The web facility is a tool for institutions to use. Therefore, HEFCE staff will not access any data submitted to, or derived by, the web facility except where an institution explicitly gives permission. Only those fields listed in the technical appendices will be retained on our system. We will remove individualised files and any outputs from the web facility by 1 May 2007, and the submitted data files will be deleted immediately after the outputs have been generated. Details about our commitments to handling personal data can be found on the HEFCE web-site, www.hefce.ac.uk, under About us/Standards of service/How we do it – our commitments/Release guidelines for personal data.

HEFCE-recognised funding consortia

39. In July 2006, we made available 'HEFCE web facility for 2005-06 statistics derived from ILR data' (HEFCE 2006/22) to all further education colleges that provide higher education provision. The primary purpose of that facility is to assist colleges in the return of accurate 2005-06 Individualised Learner Record (2005-06 ILR) data to the Learning and Skills Council (LSC). We encourage member colleges of HEFCE-recognised funding consortia to use this facility prior to submission of the 2005-06 ILR to the LSC.

40. Member colleges' 2005-06 ILR F04 data will form part of the HESES05 re-creation for the lead institution of the HEFCE-recognised funding consortium when the 'Statistics derived from HESA data' exercise is launched later in the year. We have introduced extra functionality into the HESA data web facility, for the first time this year, so that when a higher education institution (HEI) is a lead of a HEFCE-recognised funding consortium and its member colleges have uploaded their ILR data to their own web facility, then this data will be incorporated into the lead institution's HESES re-creation.

Northern Irish higher education institutions

41. For Northern Irish higher education institutions we produce the following outputs:

- HESES05 re-creation
- HESES05 re-creation based on cost centre sector norms for subjects
- RAS05 re-creation
- Derived statistics likely to inform 2007-08 WP allocations
- RDQRs
- HESES06 non-completion toolkit.

Guidance

Instructions on how to use the web facility and the outputs available

42. Instructions on how to get started with the web facility and how to access the outputs are at Annex B.

FAQs

43. Frequently asked questions (FAQs) regarding the web facility are on the HEFCE web-site under Learning & teaching/Data collection. We encourage institutions to refer to the FAQs for guidance in the first instance. We will only use our e-mail list of HESES or RAS contacts to notify institutions of significant changes or updates.

Troubleshooting

44. Appendices 2, 5 and 8 will help identify the causes of discrepancies between the 2005-06 HESA student data and the HESES05, HESES05 re-creation based on cost centre sector norms for subjects, and RAS05 comparisons respectively.

HEFCE contacts

45. Support is available for the web facility at HEFCE via telephone and e-mail (see page 2 for contact details).

Annexes and technical appendices

46. The annexes provide guidance on using the web facility and its outputs. The technical appendices describe the algorithms used to generate the derived statistics. The appendices are downloadable separately from the HEFCE web-site with this document, under Publications.

Comments

47. All institutions are invited to comment on any of the methods described in this publication. Comments should be e-mailed to hesa_heses_feedback@hefce.ac.uk for the attention of Ben Grassby.

Annex A

Summary of changes since publication of '2004-05 statistics derived from HESA data: Guide to HEFCE web facility' (HEFCE 2005/32)

Purpose

1. This annex describes the changes that have been made since the release of the equivalent publication from last year: '2004-05 statistics derived from HESA data: Guide to HEFCE web facility' (HEFCE 2005/32).

Information for lead institutions and members of HEFCE-recognised consortia

2. We have enhanced both the HESA web facility and the equivalent facility that is provided to further education colleges (FECs) so that where data from member institutions have been submitted to the ILR web facility these will be incorporated in the appropriate outputs generated for the lead institution when it submits its data to the web facility (including the re-creation of Table 6: HEFCE-recognised funding consortia 2005-06).

3. Annex G provides lead institutions and members of HEFCE-recognised funding consortia with details of specific information relating to the availability and provision of data.

RDQRs

4. We monitor, using HESA student data, the time research students take to obtain their qualifications. This year the web facility will produce one qualification rate; this will be for full-time students that commenced their research degree in 1999-2000. Institutions will be required to verify that the rates produced are suitable for publication and that the underlying data used is fit for purpose. Unlike previous years, where institutions identify errors in HESA student data for the years 1999-2000 to 2003-04 we will accept historical data error files to enable us to correct the derived fields for the individual students affected.

5. Unlike previous years, we include an additional context statistic on the number of students who transfer to another institution. The methods we will use to identify such students are described in Appendix 11.

Transfer of information to the TDA

6. Last year the web facility generated performance profile data for providers of initial teacher training which were passed to the Training and Development Agency for Schools (TDA). This function has now been removed from the web facility as it is now offered directly by HESA.

RAS05 Re-creation comparison tables

7. To reflect recent changes in the way we calculate the funding associated with the Research Activity Survey, in the summary comparison of RAS05 and the provisional RAS05

re-creation by broad subject group we now only include in this comparison full-time equivalent (FTE) from Tables R1a and R1b. In addition we have only included FTE from years 1-3 from the R1a (full time) tables and we have only included years 1-6 from the R1b (part time) tables.

Index of derived fields

8. Appendix 18 contains an alphabetical list of all algorithms that are included in the publication. They are also listed in the order they appear in their appropriate appendix.

Changes to outputs

9. Tables 1, 2 and 3 no longer have Column 5 containing separate new entrant data.

10. Institutions that are the lead of a HEFCE-recognised funding consortium will also have Table 6 in the HESR05XXXX.xls output, which is a subset of Columns 1 and 2 of Tables 1, 2 and 3, disaggregated by consortium member and level.

11. All home and EC postgraduate research students are now treated as non-fundable in the HESES05 re-creation. They are still included in the Research Activity Survey in the same way as previous years.

12. It is now possible for institutions to record foundation degree students in Table 2 as long as they meet appropriate definitions of sandwich year-out.

Changes to algorithms

13. The changes to algorithms listed here are those that have changed since '2004-05 statistics derived from HESA data: guide to HEFCE web facility' (HEFCE 2005/32).

HESTYPE

14. The algorithm for allocating students to level of study has changed to reflect the change in definition for home and EC postgraduate research students in HESES05, who should now be returned as non-fundable. The algorithm for HESTYPE is given in paragraph 23 of Appendix 1.

EQGRP

15. The algorithm for the entry-qualification risk group that the student was assigned to has been modified. Young unknowns have been separated from the young, low risk category and mature unknowns have been separated from the mature low risk category. The algorithm for EQPOP is given in paragraph 26 of Appendix 10.

EARLYLVE

16. Previously where DATELEFT or COMDATE have changed between HESA records, the earliest DATELEFT and latest COMDATE (STARTDTE) were used respectively. The algorithm has now changed to use the latest DATELEFT and the earliest COMDATE.

RDQRTUIT

17. We now define the source of fees for research degree students as 'research council funded' where the major source of tuition fee (MSTUFEE) has been returned on the HESA record as being from a research council (values 11,12,13,14,15,16,17 or 19) and the student has been returned as active in any academic year. Please see paragraph 38 of Appendix 11 for further details.

RASTYPE

18. The algorithm for allocating students to level of study has changed to reflect the change in definition for home and EC postgraduate research students in HESES05, who should now be returned as non-fundable. The algorithm for RASTYPE can be found in paragraph 36 of Appendix 7.

RSTUEXCL

19. We have added an additional, institutional specific, exclusion code (RSTUEXCL=32) which applies where an institution has identified to us students on the RAS re-creation that are also staff and generate one staff FTE. Such students should be excluded from the re-creation to avoid inflation of numbers.

RDQR subject

20. We have revised the RDQR subject assignments to more closely reflect those used in publishing Teaching Qualification Information (TQI) summaries and by HESA in their publications. The changes affect the RDQR subject (RDSBJXX) derived field, but in addition, there are two new fields that inform the derivation of RDSBJXX that categorise the subject areas for the HESACODE (RDHCODXX) and JACS (RDJACSXX) subject classifications for 1999-2000 to 2001-02 and 2002-03 to 2005-06 respectively. Please see paragraphs 32-24 of Appendix 11 for further details.

Annex B

How to use the HEFCE web facility

Purpose

1. This annex contains detailed instructions on how to use the web facility.

Data preparation

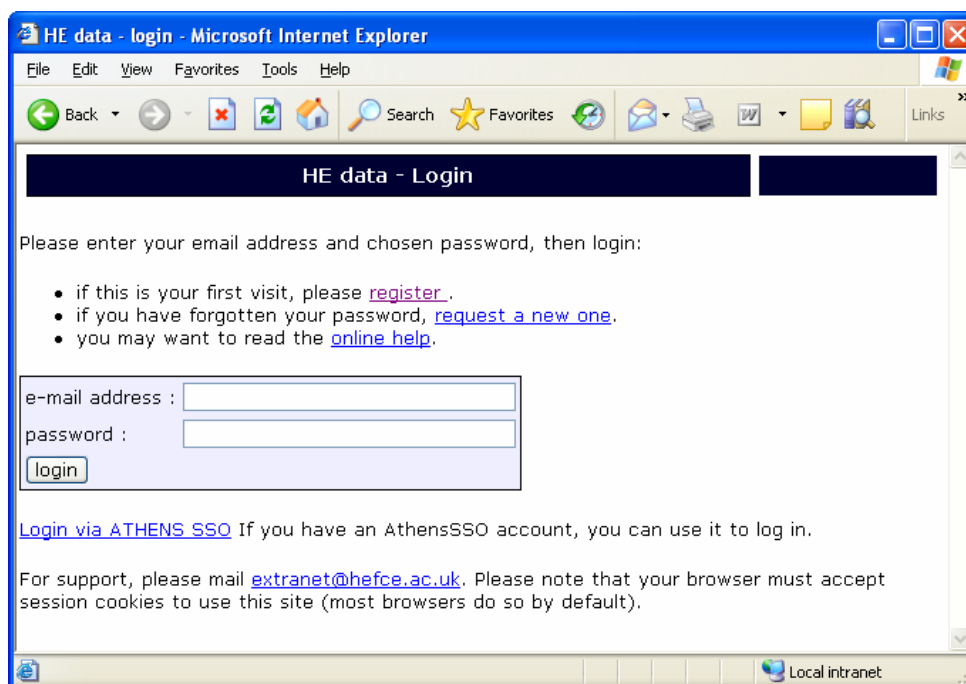
2. The web facility can process data files structured in the formats described in the HESA Student Record Coding Manual 2005/06. This includes both fixed-width and comma-separated formats, as well as both the student/module and combined records.
3. It is currently possible to submit HESA data to the web facility that is invalid (in other words HESA data that has not passed all the validation rules). It should be noted however that although we believe the facility to be robust, it may not function reliably when the input files contain invalid data. In order to ensure reliable outputs are generated, institutions should first correct any errors identified by running the HESA validation software on their data before submission. HESA's validation software can be found at www.hesa.ac.uk under Data Collection/Download Validation Kits.
4. In addition HESA publishes a number of commit-level validation checks. If data fail these checks, the web facility may generate unreliable results. In particular, if the combination of HUSID, INSTID and NUMHUS is not unique for any records for students with higher education qualification aims, the re-creations will generate incorrect results.

Format of files

5. In each institution's HESA file(s), all fields, for example RECID, should retain their leading zeros; otherwise the web facility will not be able to read the information correctly. Users may find that saving files in Excel may mean the leading zeros are lost. To save the file(s) with leading zeros we recommend the use of a text editor such as Notepad.
6. If an institution returns student HESA data using the student and module structure, both the student and module files should be submitted to generate the re-creation.
7. The web facility can process zip files, provided that the contents are in the format specified by HESA.

Access to the facility

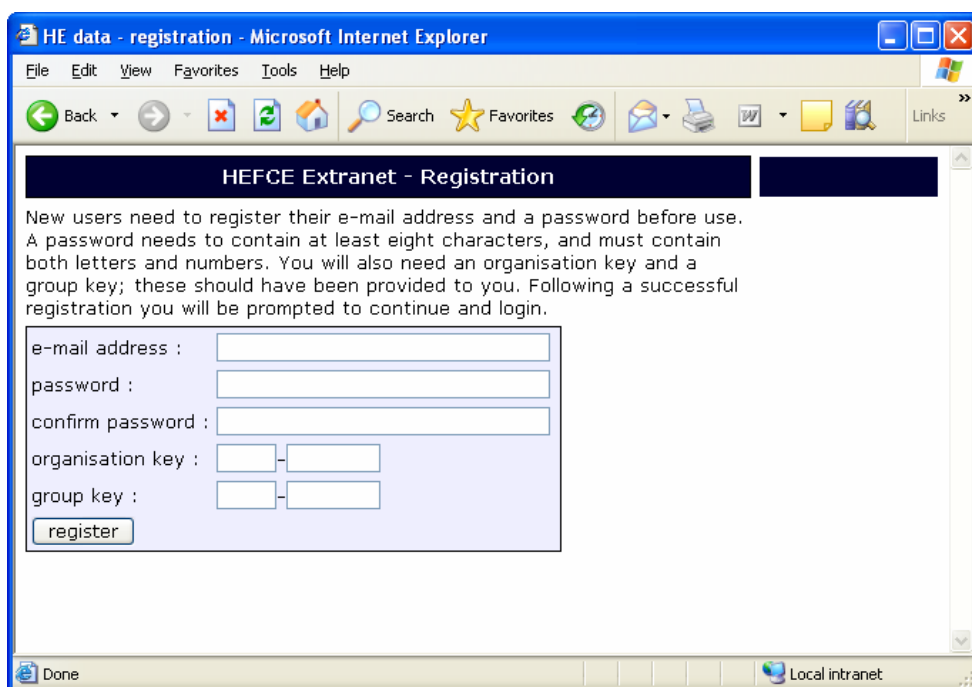
8. The web facility can be accessed via the HEFCE extranet at <https://extranet.hedata.ac.uk>. On viewing the screen below you will be prompted to either log in or register for access to the HEFCE extranet as a new user.



Registering a new account

9. New users of the HEFCE extranet will first need to register an e-mail address and extranet password before use. This can be done by clicking the 'Register' link on the login screen. In order to register, you will require an 'organisation key' and a 'group key', details of which have been provided in the letter titled 'HEFCE web facility for 2005-06 statistics derived from HESA data' sent to your head of institution by Ben Grassby in July 2006. This information cannot be sent electronically for security reasons so please retain the information sent in the letter. We can re-issue the information by post if necessary.

10. Once registered you should be able to log in by entering the e-mail address you used during registration, and the password that you created.



Existing users of the extranet

11. If you have used the HEFCE extranet for other HEFCE returns (for example HESES05) you will be required to log in and join the HEFCE web facility for 2005-06 statistics derived from HESA data group. Follow the log-in procedure by entering your e-mail address and password. You will be redirected to a page for 'HEFCE extranet – All resources'; under 'Applications' click 'Join a group'. Enter the group key supplied in Annex A of the letter titled 'HEFCE web facility for 2005-06 statistics derived from HESA data' sent to your head of institution by Ben Grassby in July 2006, and select 'Join group'.

12. If you have registered in the past but your account has expired, you will be required to refresh your account using the organisation key referred to in paragraph 0.

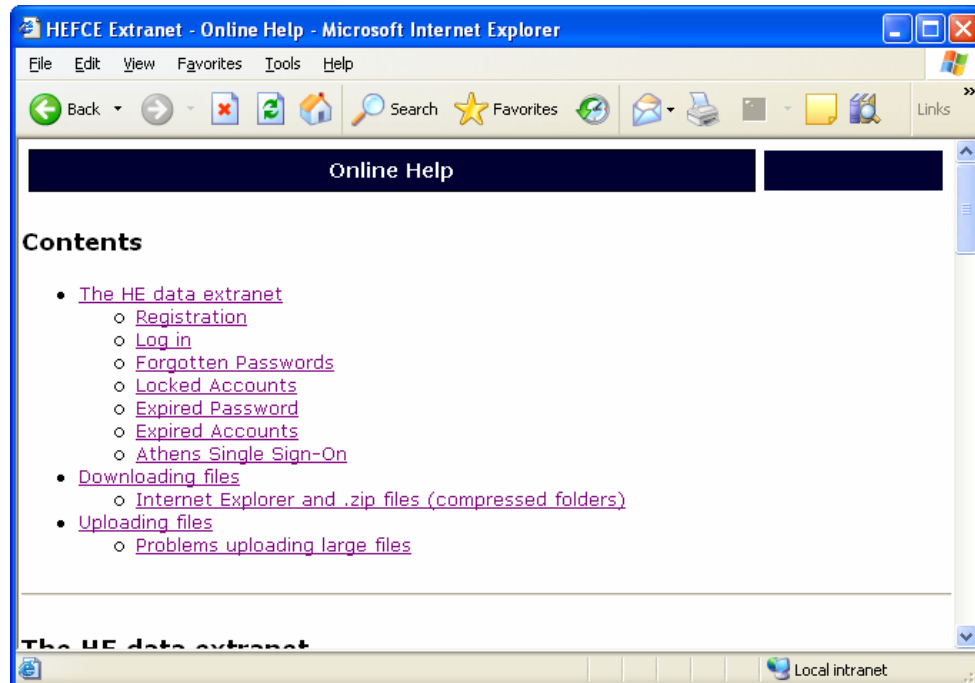
Athens Single Sign On Account

13. You can also log in to the HEFCE extranet using Athens Single Sign On account (if this is available at your institution).

- a. Follow the 'Log in via Athens SSO' link on the login page.
- b. Log in to Athens as normal (if not already done).
- c. When Athens has authenticated you, your browser will be directed to the 'HEFCE extranet - All resources' page.
- d. You will be required to join the group. Under 'Applications' click 'Join a group'. Enter the group key sent to the head of your institution in a letter titled 'HEFCE web facility for 2005-06 statistics derived from HESA data' by Ben Grassby in July 2006, and select 'Join group'.

Help

14. An information page can be accessed by clicking the 'online help' link, located above the log-in box, or on the right of the 'HEFCE extranet - All resources' page, under 'Navigation'.



Forgotten password

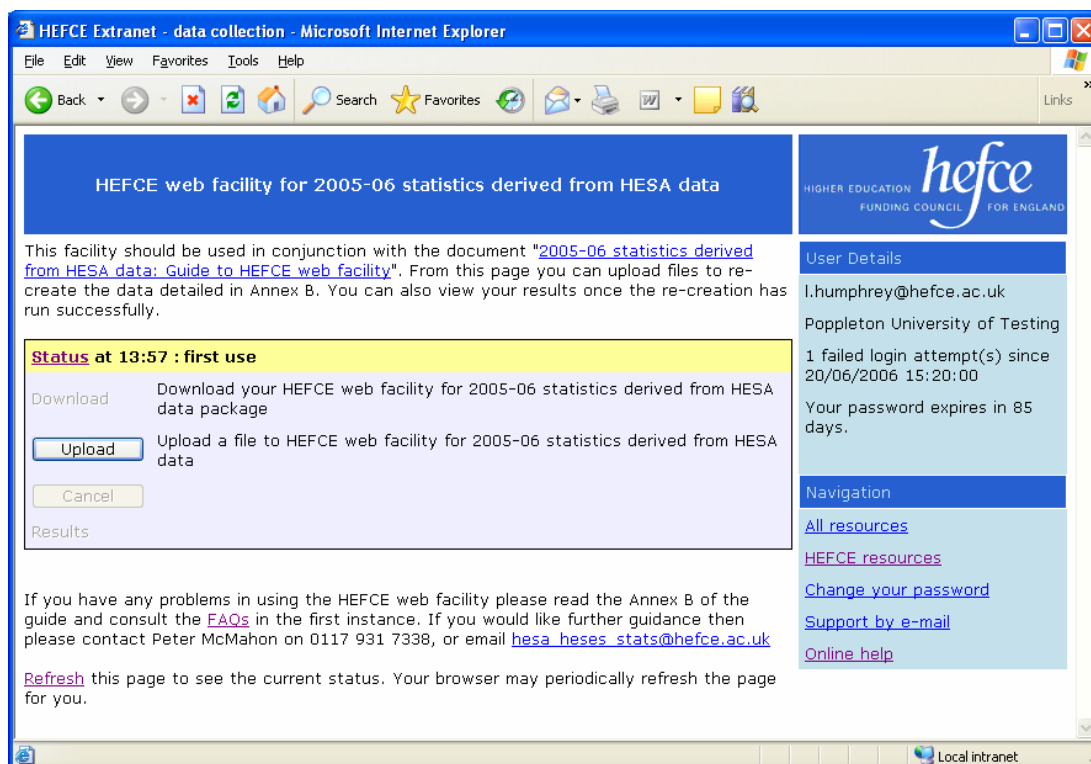
15. If you forget your password, select 'request a new one' on the log-in page. Enter your e-mail address in the space provided and click 'request password'. A new password will be generated and sent to you. You will be required to change this password before being able to access the HEFCE extranet.

16. Each user has three attempts to log in before the account is locked. If all three attempts are used and access to the web facility is still not allowed, contact Lauren Humphrey or Peter McMahon (e-mail hesa_heses_stats@hefce.ac.uk).

All resources

17. After verifying the username and password, your browser will be directed to the 'HEFCE extranet - All resources' page. Click 'HEFCE resources'.

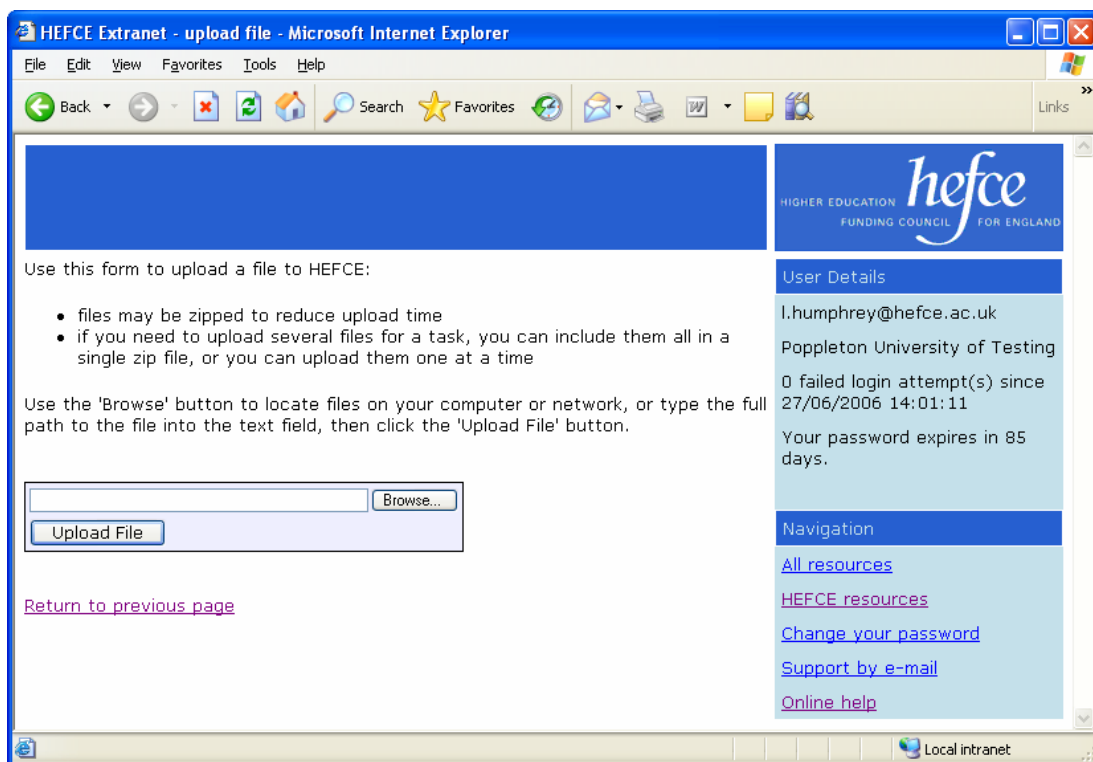
18. From the 'Applications' list click on the 'HEFCE web facility for 2005-06 statistics derived from HESA data' link.



19. This option will allow HESA file(s) to be uploaded one at a time. Use the 'Browse' button to reach the location of the data file and once the correct file is located, select 'Upload File'. The browse facility does not allow selection of multiple files.

20. Continue to add files until all HESA files have been uploaded. All files should be uploaded by repeating the process in paragraph 19.

21. Files which have been uploaded to the web facility are listed under 'Current files'. An uploaded file may be deleted by selecting the 'delete' link next to the appropriate file.



Process file(s)

22. Once all of the files have been added, users may then select 'Process'. This places the file(s), added in the previous step, into the processing queue.

Processing

23. You can leave the extranet while the derived statistics are being generated. To do this, exit the internet in the usual way, or select a new address in the address bar. You may be prompted to log in again when returning to the page, but no information will be lost.

24. Each time a collection of files is submitted for re-creation, you should refresh the page using either the icon in the tool bar or the link on the page. Refreshing the page ensures that the 'History of HEFCE web facility for 2005-06 statistics derived from HESA data submissions' reflects the current stage of processing, as the page may not update automatically.

25. Our systems only allow one submission to be processed at a time; therefore if more than one institution uploads data, the submissions will be placed in a queue. A submission by an average sized institution typically takes 15 minutes to process (once it reaches the front of the queue). The total time taken to complete the re-creation, including waiting time, will depend on the number of institutions in the queue.

26. If the uploaded files are waiting in the queue it is possible to cancel the submission by selecting 'Cancel'.

27. If you do not use the extranet for 60 minutes, you will be automatically logged out of the system. To regain entry, select any option and re-enter the e-mail address and extranet password.

Output

28. Once the files have been processed, the outputs can be accessed by selecting 'Results'.

29. If the outputs were not generated, and the reason is known, it will be given under 'History of HEFCE web facility for 2005-06 statistics derived from HESA data submissions'. If the outputs were not generated because an unspecified error occurred, this will typically be a result of invalid HESA data.

Accessing results

30. The results are made available as a zip archive (HESR05XXXX, where XXXX denotes the HESA institution identifier) containing the following files:

- HESR05XXXX.xls – This is an Excel workbook containing the HESES05 re-creation tables. For leads of HEFCE-recognised funding consortia this includes the combined 2005-06 HESA data uploaded by the lead and 2005-06 ILR data uploaded by member colleges. (If member colleges have not uploaded these data they will not be included. We expect institutions within HEFCE-recognised funding consortia to liaise with each other to coordinate submission of data to the web facilities.)
- SNCC05XXXX.xls – This is an Excel workbook containing the HESES05 re-creation based on cost centre sector norms for subject tables.
- RASR05XXXX.xls – This is an Excel workbook containing the RAS05 re-creation tables.
- WP05XXXX.xls – This is an Excel workbook containing the derived statistics that are likely to be used to inform the 2007-08 WP allocations (WPC05YYYYYY.xls for members of consortia). For leads of HEFCE-recognised funding consortia this includes the combined 2005-06 HESA data uploaded by the lead and 2005-06 ILR data uploaded by member colleges. (If member colleges have not uploaded these data they will not be included.)
- RDQR05XXXX.xls – This is an Excel workbook containing the RDQRs.
- PUB05XXXX.xls – This is an Excel workbook containing the derived statistics that will be used in HEFCE regional statistics.
- NCTK05XXXX.xls – This is the HESES06 non-completion toolkit pivot table.

- HESR05XXXX.ind – This is the HESES05 re-creation individualised student data file – for leads of HEFCE-recognised funding consortia this file does not include data for the member colleges.
- SNCC05XXXX.ind – This is a comma-separated file containing the derived fields that generate the HESES05 re-creation based on cost centre sector norms for subjects.
- RASR05XXXX.ind – This is a comma-separated file containing the derived fields that generate the RAS05 re-creation.
- WP05XXXX.ind – This is the 2007-08 widening participation individualised student data file – for leads of HEFCE-recognised funding consortia this file does not include data for the member colleges.
- RDQR05XXXX.ind – This is a comma-separated file containing the derived fields that generate the RDQRs.
- PUB05XXXX.ind – This is a comma-separated file containing the derived fields that generate the derived statistics that will be used in HEFCE regional statistics.

31. For leads of HEFCE-recognised funding consortia, in addition to the outputs above the following outputs are also available:

- LEAD05XXXX.xls – This is an Excel workbook containing the HESES05 re-creation tables for the lead institution. These tables are created from the lead institution's 2005-06 HESA data that have been submitted to the web facility.
- HEIRC05YYYYYY.xls – These are Excel workbooks containing the HEIFES05 re-creation tables for each of the member colleges (YYYYYY denotes the UPIN provider number for the college). These tables are created from the 2005-06 ILR data that the member college(s) have submitted to the ILR web facility. These tables will only become available after member colleges have submitted data to the web facility.
- HHWPLEAD05XXXX.xls – This is an Excel workbook containing the widening participation allocation tables for the lead institution. These tables are created from the lead institution's 2005-06 HESA data that have been submitted to the web facility.
- IHWPC05YYYYYY.xls – These are Excel workbooks containing the widening participation allocation tables for each of the member colleges of a HEFCE-recognised funding consortium that have submitted ILR data to the ILR web facility.

Using the individualised files

32. The individualised files show the allocation of students to cells within the equivalent re-creation tables and, where relevant, details of why they were excluded. For institutions with individualised files that do not contain greater than 65,530 records, the following

guidance will assist them in the troubleshooting process. All records with a specific value in a specific field can be found in the following way:

- a. Open an individualised file, from the list above in Microsoft Excel, and click <File>, <Open>. Users will need to specify 'All Files' in the 'Files of type' box before the individualised file will appear in the file listing. Once the file is selected, the 'Text Import Wizard' will appear. Ensure that 'Delimited' is selected near the top of the window, then click 'Next'. On the next page, uncheck 'Tab' and check 'Comma'. Click 'Finish' to open the file.
- b. Select the row containing the field headings.
- c. Select <Filter> from the <Data> menu and then <Autofilter>.
- d. Click on the arrow in the column containing the data which needs filtering.
- e. Either select a specific value or select <Custom> to apply a comparison operator other than equality.
- f. To select records using multiple fields, repeat steps d and e for each field.

33. The RDQR action plan template (APRDQR05XXXX.xls) is available from the HEFCE extranet – 2005-06 Statistics derived from HESA data page. To access the action plan follow the instructions in paragraph 17. You will need to click the download link to download the template.

34. APRDQR05XXXX.xls is an Excel workbook containing the action plan template for RDQRs. This workbook is available for all institutions.

35. RDQR action plans will be available for download from the HEFCE extranet from **1 August 2006**.

Using the output

36. The tables are best viewed in Microsoft Excel. When you open the individualised file, some fields (such as HUSID) may look like '9.64E+08'. To obtain the 13-digit code as returned to HESA follow these steps:

- a. Highlight the column containing the field you wish to format.
- b. Click on <Format>, then <Cells>.
- c. In the <Number> tab, select the <Custom> category.
- d. In the <Type> text box, enter 13 zeros.
- e. Click <OK> to get back to the spreadsheet.

These files will be overwritten each time new data are submitted. Note that it is not possible for institutions to see one another's files except for members of HEFCE-recognised funding consortia.

37. At this point it may be useful to check whether the tables reflect the number of students expected. If it does not, view the files uploaded and check that they are the files that were

intended to be submitted. We are not responsible for the accuracy of the files submitted to the web facility.

Technical issues

38. While there are no limits on the size of files that can be uploaded to the facility, it is common for individual institutions to have limits imposed by their network administrators. Attempts to upload a file beyond the size limit imposed within the institution are unlikely to generate a meaningful error message; it will however cause the browser to 'stall' or 'hang'.

39. If difficulties are encountered uploading files due to file size or any other issue relating to connectivity we will normally be able to advise. We recommend that you involve your network administrator in such discussions.

40. Please note that the internet protocol (IP) address for this year's web facility is 195.194.167.114.

Annex C

Comparison of HESES05 and the HESES05 re-creation

Purpose

1. This annex describes the methods we use to generate a provisional HESES05 re-creation via the web facility, for use by institutions prior to their data submission to HESA. It also describes how we are likely to use the comparison of HESES05 and the HESES05 re-creation in the '2005-06 statistics derived from HESA data' exercise.

Web facility outputs

2. The HESES05 re-creation is generated from HESA 2005-06 student data submitted to the web facility. Annex B describes how to access the provisional HESES05 re-creation Excel workbook (HESR05XXXX.xls – where XXXX denotes the HESA institution identifier). The provisional HESES05 re-creation workbook contains the worksheets listed in Table 1 below. The outputs are provisional because they are generated before submitting final data to HESA.

Table 1 Excel workbook 'HESR05XXXX'

Page number	Worksheet (see tabs on spreadsheet)	Description
1	Coversheet	Title page
2	Summary	Summary comparison of HESES05 and the provisional HESES05 re-creation
3	SummaryPG	Summary comparison of price group activity between HESES05 and the provisional HESES05 re-creation
4	Excl	Students excluded from the provisional HESES05 re-creation
5	FTS	Provisional HESES05 re-creation Table 1a: Full-time and sandwich years of programme of study
6	MED	Provisional HESES05 re-creation Table 1b: Medical and dental full-time and sandwich years of programme of study
7	SWOUT	Provisional HESES05 re-creation Table 2: Sandwich year-out years of programme of study
8	PT	Provisional HESES05 re-creation Table 3: Part-time years of programme of study and load
9	FEE	Provisional HESES05 re-creation Table 4: Home and EC fees
10	CONS	Provisional HESES05 re-creation Table 6: HEFCE-recognised funding consortia 2005-06
11	HBK	Provisional HESES05 re-creation grant adjustment report

12	STD	Provisional HESES05 re-creation standard resource table
13	F05	Provisional HESES05 re-creation assumed fee income table
14	hFTS	HESES05 Table 1a: Full-time and sandwich years of programme of study
15	hMED	HESES05 Table 1b: Medical and dental full-time and sandwich years of programme of study
16	hSWOUT	HESES05 Table 2: Sandwich year-out years of programme of study
17	hPT	HESES05 Table 3: Part-time years of programme of study and load
18	hFEE	HESES05 Table 4: Home and EC fees
19	hCONS	HESES05 Table 6: HEFCE-recognised funding consortia 2005-06
20	hHBK	HESES05 grant adjustment report
21	hSTD	HESES05 standard resource table
22	hF05	HESES05 assumed fee income table

Note: Only institutions that are the lead of a HEFCE-recognised funding consortium will be able to view data in Table 6 of the HESR05XXXX.xls output, which is a subset of Columns 1 and 2 of Tables 1, 2 and 3, disaggregated by consortium member and level.

The workbook also contains a version number that will increment by one following each successful submission of data to the web facility, as well as contact details for any queries relating to data contained in the workbook.

3. The information contained in the HESES05 re-creation tables can be re-built from the individualised file, HESR05XXXX.ind, which the web facility produces. The file contains details of how each student was classified in the re-creation and hence how the student is represented in the tables or, where relevant, details of why they were excluded. A full description of the data contained in the individualised file is given in Appendix 1.

Methods

4. The algorithms used to generate the provisional HESES05 re-creation are described in Appendix 1.

5. Appendix 2 contains troubleshooting guidance on the causes of differences between HESES05 and the provisional HESES05 re-creation.

6. All known problems of fit with the provisional HESES05 re-creation algorithms are described in Appendix 3.

2005-06 statistics derived from HESA data exercise

4. In late 2006 or early 2007 we will derive a HESES05 re-creation from 2005-06 student data submitted to HESA. We intend to use the algorithms detailed in Appendix 1 to generate the HESES05 re-creation; however, we may change them if we identify any improvements that can be made.
5. We will compare the HESES05 re-creation to HESES05. This comparison takes place after the 2005-06 student data have been finalised with HESA and passed to HEFCE.
6. We will re-calculate a grant adjustment report for the HESES05 re-creation by applying the same formulae that were used to calculate the grant adjustment report for HESES05.
7. It is likely that we will select institutions to explain discrepancies between their HESES05 and HESES05 re-creation on the basis of thresholds. The thresholds are likely to be set in terms of funding differences identified from the comparison of the HESES05 and the HESES05 re-creation grant adjustment reports. Notwithstanding the thresholds, we may also ask for further information from any institution in respect of this comparison. This may result ultimately in adjustments to grant, where appropriate.
8. Each institution requested to make a response will be asked to reconcile the two data sources. After reconciling the two data sources and confirming that the HESES05 re-creation reasonably reflects the outturn position for 2005-06, the HESES05 re-creation will supersede HESES05 and any consequent grant adjustments will be made (subject to the appeals process and the availability of our funds).

Annex D

Comparison of the HESES05 re-creation and the HESES05 re-creation based on cost centre sector norms for subjects

Purpose

1. In addition to generating a provisional HESES05 re-creation, the web facility produces a provisional HESES05 re-creation based on cost centre sector norms for subjects using the HESA data that the institution submits via the web facility. This is essentially a HESES05 re-creation, but the cost centre allocations are replaced with a sector norm mapping for subjects. This sector norm mapping of cost centre for subjects is created by observing the way the sector has assigned cost centres for subjects where the total student FTE assigned across the subject is greater than 100.
2. This annex describes the methods we use to generate this provisional HESES05 re-creation based on cost centre sector norms for subjects via the web facility.
3. Due to the absence of full HESA 2005-06 data, the web facility uses the cost centre sector norms for subjects that we generated from HESA 2004-05 student data. We do not expect the sector norms for 2005-06 to differ significantly from the mapping we are using.
4. This annex also describes how we will use the comparison of HESES05, the HESES05 re-creation and the HESES05 re-creation based on cost centre sector norms for subjects in the 2005-06 statistics derived from HESA data exercise, when it is launched in late 2006 or early 2007.

Background

5. We expect, based on HEFCE guidance, that student load has been allocated to cost centres based on the cost centre of the member of staff most directly associated with it. In general, academic staff should be assigned to the cost centre that best represents the majority of their academic activity.
6. Further guidance on assigning academic departments to cost centres is in HEFCE Circular letter 32/2005, 'Assignment of departments to academic cost centres: 2005-06' which was published in December 2005. This is available for download from the HEFCE web-site under Publications/Circular letters.

Web facility outputs

7. For each institution, the HESES05 re-creation based on cost centre sector norms for subjects is generated from the HESA 2005-06 student data that it submits to the web facility together with the sector norm mapping we create. Annex B describes how to access the relevant Excel workbook (SNCC05XXXX.xls – where XXXX denotes the HESA institution identifier), which contains the worksheets in Table 2 below.

Table 2 Excel workbook 'SNCC05XXXX.xls'

Page number	Worksheet (see tabs on spreadsheet)	Description
1	Coversheet	Title page
2	snSummary	Summary comparison of the provisional HESES05 re-creation and the provisional HESES05 re-creation based on cost centre sector norms for subjects
3	snSummaryPG	Summary comparison of price group activity between the provisional HESES05 and the provisional HESES05 re-creation based on cost centre sector norms for subjects
4	snExcl	Students excluded from the provisional HESES05 re-creation based on cost centre sector norms for subjects
5	snFTS	Provisional HESES05 re-creation based on cost centre sector norms for subjects Table 1a: Full-time and sandwich years of programme of study
6	snMED	Provisional HESES05 re-creation based on cost centre sector norms for subjects Table 1b: Medical and dental full-time and sandwich years of programme of study
7	snSWOUT	Provisional HESES05 re-creation based on cost centre sector norms for subjects Table 2: Sandwich year-out years of programme of study
8	snPT	Provisional HESES05 re-creation based on cost centre sector norms for subjects Table 3: Part-time years of programme of study and load
9	snFEE	Provisional HESES05 re-creation based on cost centre sector norms for subjects Table 4: Home and EC fees
10	snHBK	Provisional HESES05 re-creation based on cost centre sector norms for subjects grant adjustment report
11	snSTD	Provisional HESES05 re-creation based on cost centre sector norms for subjects standard resource table
12	snF05	Provisional HESES05 re-creation based on cost centre sector norms for subjects assumed fee income table
13	FTS	Provisional HESES05 re-creation Table 1a: Full-time and sandwich years of programme of study
14	MED	Provisional HESES05 re-creation Table 1b: Medical and dental full-time and sandwich years of programme of study
15	SWOUT	Provisional HESES05 re-creation Table 2: Sandwich year-out years of programme of study
16	PT	Provisional HESES05 re-creation Table 3: Part-time years of programme of study and load
17	FEE	Provisional HESES05 re-creation Table 4: Home and EC fees

18	HBK	Provisional HESES05 re-creation grant adjustment report
19	STD	Provisional HESES05 re-creation standard resource table
20	F05	Provisional HESES05 re-creation assumed fee income table

Note: The workbook also contains a version number that will increment by one following each successful submission of data to the web facility, as well as contact details for any queries relating to data contained in the workbook.

8. The information in the HESES05 re-creation based on cost centre sector norms tables can be re-built from the individualised file provided. The file (SNCC05XXXX.ind) contains details of how each student was classified in the re-creation, and where relevant, details of why they were excluded. A full description of the data in the individualised file is given in Appendix 4.

Methods

9. The algorithms used to generate the provisional HESES05 re-creation based on cost centre sector norms for subjects are described in Appendix 4.

10. Appendix 5 contains troubleshooting guidance to identify the causes of differences between the provisional HESES05 re-creation and the provisional HESES05 re-creation based on cost centre sector norms for subjects, especially where they are attributable to errors in HESA data.

11. All known problems of fit with the provisional HESES05 re-creation based on cost centre sector norms for subjects algorithms are described in Appendix 6.

2005-06 statistics derived from HESA data exercise

12. This web facility enables an institution to view re-creations based on its HESA student data prior to committing its data to HESA. Once we have received a finalised version of all institutions' data from HESA, in December 2006/early 2007, we will derive a HESES05 re-creation from HESA 2005-06 student data. We intend to apply the algorithms detailed in Appendix 1 to generate the HESES05 re-creation.

13. At this time we will also derive a HESES05 re-creation based on cost centre sector norms for subjects. We intend to use the algorithms detailed in Appendix 4 to generate the HESES05 re-creation based on cost centre sector norms for subjects, except that the sector norm mapping of subjects to cost centres will be revised on receipt of full HESA 2005-06 student data. The web facility uses the cost centre sector norms generated from HESA 2004-05 student data, and we do not expect the sector norms for 2005-06 to differ significantly from them. However, if we identify an error in our algorithms, we will correct it.

14. We will calculate a grant adjustment report for the HESES05 re-creation and the HESES05 re-creation based on cost centre sector norms for subjects by applying the same formulae that were used to calculate the grant adjustment report for HESES05.

15. When the data exercise is launched, it is likely that we will select institutions to explain discrepancies between their cost centre assignments and the sector norm cost centre assignments on the basis of thresholds. The thresholds are likely to be set in terms of funding differences identified from the comparison of the grant adjustment reports produced by the HESES05 re-creation and the HESES05 re-creation based on cost centre sector norms for subjects. Notwithstanding these thresholds, we may also ask for further information from any institution in respect of any part of this comparison. This may result ultimately in adjustments to grant where appropriate.

16. During this comparison, we will incorporate previous approvals for the way an institution has assigned subject activity to cost centres that differs from the sector norm.

17. Each institution that has been requested to make a response will be asked to reconcile its HESES05, HESES05 re-creation and the HESES05 re-creation based on cost centre sector norms for subjects. After reconciling all three and confirming that the HESES05 re-creation reasonably represents the outturn position for 2005-06, the HESES05 re-creation will supersede HESES05 and any consequent grant adjustments will be made (subject to the appeals process and the availability of our funds).

Annex E

Comparison of RAS05 and the RAS05 re-creation

Purpose

1. This annex describes the methods we use to generate a provisional RAS05 re-creation via the web facility. It also describes how we are likely to use the comparison of RAS05 and the RAS05 re-creation in the 2005-06 statistics derived from HESA data exercise. The outputs are provisional because they are generated before submitting final data to HESA.

Web facility outputs

2. The provisional RAS05 re-creation is generated from HESA 2005-06 student data submitted to the web facility. Annex B describes how to access the provisional RAS05 re-creation Excel workbook (RASR05XXXX.xls – where XXXX denotes the HESA institution identifier). The provisional RAS05 re-creation workbook contains the worksheets listed in Table 3 below.

Table 3 Excel workbook 'RASR05XXXX.xls'

Page number	Worksheet (see tabs on spreadsheet)	Description
1	Coversheet	Title page
2	Summary	Summary comparison of RAS05 and the provisional RAS05 re-creation (fundable home fee paying students, years 1-3 FT, 1-6 PT).
3	Summary by subject groupings	Summary comparison of RAS05 and the provisional RAS05 re-creation (fundable home fee paying students, years 1-3 FT, 1-6 PT).
4	R1a	Form R1a: Full-time home and EC research students by year of programme
5	R1b	Form R1b: Part-time home and EC research students by year of programme
6	Funds	RAS05 calculations for 2006-07 RDP supervision funding using HESA 2005-06 and other HEFCE research funding data
7	rR1A	RAS05 Form R1a: Full-time home and EC research students by year of programme
8	rR1B	RAS05 Form R1b: Part-time home and EC research students by year of programme
9	rFunds	RAS05 calculations for 2006-07 RDP supervision funding using RAS 2005 and other HEFCE research funding data
10	Broad subject groups	Mapping of units of assessment to broad subject groups used for comparison tables
11	Research cost bands	Assignment of units of assessment to subject cost bands

Note: The workbook also contains a version number that will increment by one following each successful submission of data to the extranet, as well as contact details for any queries relating to data contained in the workbook.

3. The information contained in the RAS05 re-creation tables can be re-built from the individualised file provided. The file (RASR05XXXX.ind) contains details of how each student was classified in the re-creation, and, where relevant, details of why they were excluded. A full description of the data in the individualised file is given in Appendix 7.

Methods

4. The algorithms used to generate the provisional RAS05 re-creation are described in Appendix 7.

5. Appendix 8 contains troubleshooting guidance for the causes of differences between RAS05 and the provisional RAS05 re-creation, especially where they are attributable to errors in HESA data.

6. All known problems of fit with the provisional RAS05 re-creation algorithms are described in Appendix 9.

2005-06 statistics derived from HESA data exercise

7. The '2005-06 statistics from HESA data' exercise will contain a comparison of RAS05 and the RAS05 re-creation. This will be achieved by our deriving a RAS05 re-creation of forms R1a and R2b, by applying the algorithms detailed in Appendix 7 to HESA 2005-06 student data. This exercise is likely to be launched in late 2006 or early 2007.

8. The assignment of UOAs to subject groups which will be used in the December re-creation may be different from the one used in this web facility as it will be created using HESA 2005-06 staff data. For the purpose of the web facility, we will use the assignment of UOAs to subject groups generated from HESA 2004-05 staff data. However, we do not expect the assignment to change significantly. If we identify an error in our algorithms, we will correct it.

9. We will assign each UOA to one of eight subject groups. The assignments of UOAs to subject groups are given in paragraph 3 of Appendix 7.

10. We will compare the RAS05 re-creation to RAS05 after the 2005-06 student data have been finalised with HESA. We are likely to select institutions to make a response to the derived statistics exercise based upon the calculations for 2006-07 RDP supervision funding calculated from the RAS and the RAS re-creation. We will re-calculate the 2006-07 RDP supervision fund from the RAS05 re-creation by applying the same formulae that were used to calculate it from RAS05.

11. In the summary comparison of RAS05 and the provisional RAS05 re-creation by broad subject group we have included in this comparison only FTE from tables R1a and R1b. In addition we have only included FTE from years 1-3 from the R1a tables. Similarly we have only included years 1-6 from the R1b tables.

12. . Further information about research funding is provided in Appendix 7.

13. The 2005-06 statistics from HESA data exercise involves selection of institutions to explain discrepancies between their re-creations and the returns. It is likely that we will select institutions to explain discrepancies between their RAS05 and RAS05 re-creation on the basis of thresholds.

14. These thresholds are likely to be set in terms of differences in the FTEs at the broad subject group level arising from the RAS05 and RAS05 re-creation. We are likely to use FTE only from fundable home and EC fee paying research students that are in years 1-3 full-time or years 1-6 part-time. These FTE differences will have consequent funding adjustments associated with them. Notwithstanding the thresholds, we may also ask for further information from any institution in respect of this comparison.

15. Each institution requested to make a response will be asked to explain the reasons and causes of differences between RAS05 and the RAS05 re-creation. Where we are satisfied with these explanations, and where it is appropriate, we will ask the institution to confirm that the RAS05 re-creation reasonably reflects the outturn position for 2005-06. Upon this confirmation, the RAS05 re-creation will supersede RAS05 and any consequent grant adjustments will be made (subject to the availability of our funds).

Annex F

Derived statistics likely to inform the 2007-08 WP allocations

Purpose

1. This annex describes how we propose to use 2005-06 HESA data to inform the widening participation allocations for 2007-08. Further details of the algorithms are provided in Appendix 10.
2. The web facility offers institutions the opportunity to check and, where errors would affect these allocations under current proposals, correct their data prior to submission to HESA.
3. We expect HESA 2005-06 student data to inform the following widening participation allocations for 2007-08:
 - widening access for students from disadvantaged backgrounds
 - improving retention for full-time students
 - widening access for disabled students.

Web facility outputs

4. The web facility generates the derived statistics likely to inform 2007-08 WP allocations based on the provisional HESA data submitted to it.
5. Annex B describes how to access the derived statistics likely to be used to inform the 2007-08 WP allocations in an Excel workbook (WP05XXXX.xls – where XXXX denotes the HESA institution identifier).
6. The WP derived statistics can be re-built from the individualised file which we provide. The file (WP05XXXX.ind) contains details of how each student was classified in the tables and, where relevant, details of the reasons why they were excluded from the tables. A full description of the data in the individualised file is given in Appendix 10.

Indicative funding calculations

7. The web facility will generate an indicative 2007-08 funding calculation for each of the WP allocations. The calculations use 2006-07 allocation rates applied to 2006-07 assumed FTEs. They will not incorporate 2007-08 additional student numbers, transfers or mergers.
8. The indicative allocations for 2007-08 are provided solely to highlight potential HESA student data errors. They should not be considered to be any kind of funding commitment by us and are without prejudice to what our Board may agree to be the final allocations for any institution. The final allocations for 2007-08 may be higher or lower than the illustrations given in the output from the web facility as a result of changes to data by the institution or to the data provided by any other institution, or to the total sum available for allocation.

Widening access for students from disadvantaged backgrounds

9. This is a formula-based allocation of funding for teaching to recognise the extra costs associated with recruiting and supporting undergraduate students from disadvantaged backgrounds who are currently under-represented in higher education. The funds will be allocated on the basis of higher education participation rates and average educational achievement for census wards. They will be calculated for each institution using its HESA 2005-06 student data. The proposed method is as follows.

10. Firstly, using postcode information from HESA 2005-06 data, each student is mapped to a ward. Information about the mapping we use is available by following the link <http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=7426&Pos=&ColRank=1&Rank=272>

11. These wards are then ranked in terms of their higher education participation rate (for young, under 21, full-time students – FTS) or average educational achievement (for mature full-time students and all part-time students), split into quintiles, and weighted as follows:

Quintile	Weighting
1 Lowest HE participation (young FTS) or lowest average educational achievement (mature FTS and part-time)	2
2	1
3, 4, 5	0

12. Part-time and mature students who already hold a higher education (HE) qualification at the same level as, or higher than, their current qualification aim are given a weighting of zero, irrespective of their postcode.

13. We calculate a 'widening access average weight' (separately for full-time and part-time) as follows:

$$\frac{\text{Total weight for all students in the population}}{\text{Total students in the population}}$$

The population is defined as:

- for young full-time students: HEFCE-funded UK domiciled entrants that generate a Column 4 countable year in the provisional HESES05 re-creation
- for mature full-time students: HEFCE-funded English and Welsh domiciled entrants that generate a Column 4 countable year in the provisional HESES05 re-creation
- for part-time students: HEFCE-funded English and Welsh domiciled entrants that generate a Column 4 countable year in the provisional HESES05 re-creation.

Some students are excluded from the population:

- those whose postcode is in a district that includes a school or other such institution, as there is greater uncertainty about the participation rates in these areas
- full-time mature and part-time students in Scotland and Northern Ireland
- other EU students.

These students are excluded from both numerator and denominator, and therefore receive an average weight for the purpose of allocating funds.

14. The average weight derived from paragraph 13 is London weighted (8 per cent for inner London and 5 per cent for outer London) and applied to the assumed undergraduate (UG) (including foundation degree) FTE for 2007-08 (HESES06 column 4 HEFCE-funded FTE plus adjustments for additional student numbers and mergers/transfers).

Improving retention

Full-time

15. As well as allocating funding to widen access, we will allocate funding to improve retention. For full-time undergraduate students, this is likely to be based on their entry qualifications and age. The allocations are likely to be calculated as follows.

16. Firstly, using age and entry qualification information from HESA 2005-06 data, full-time undergraduate entrants are assigned to one of six categories which are then weighted as shown below. Students are only part of the population if they generate a HEFCE-fundable Column 4 countable year in the HESES05 re-creation.

	Young	Mature
Low risk	0	0
Medium risk	1	1.5
High risk	1.5	2.5

17. For this allocation, mature students are those aged 21 or over on entry. The assignment of students to risk categories based on entry qualifications is as shown in Table 4.

18. We calculate an 'improving retention average weight' as:

$$\frac{\text{English-domiciled full-time and sandwich undergraduate entrants, weighted according to age and entry qualification}}{\text{All English-domiciled full-time and sandwich undergraduate entrants}}$$

19. The average weight derived from paragraph 18 is given a London weighting and applied to the assumed FTS UG (including foundation degree) FTE for 2007-08.

Table 4 The assignment of students to risk categories based on entry qualifications

	Young	Mature
Low risk	A-levels/Highers/vocational A-levels with more than 260 or 0* Tariff points Baccalaureate Degree or higher Unknown†	A-levels/Highers/vocational A-levels with more than 320 or 0* Tariff points Degree or higher Unknown†
Medium risk	A-levels/Highers/vocational A-levels with between 161 and 260 tariff points Foundation course Vocational A-levels only Other HE qualification	A-levels/Highers/vocational A-levels with 320 or fewer tariff points Other HE qualification Foundation course Access course Vocational A-levels only
High risk	A-levels/Highers/vocational A-levels with between 1 and 160 tariff points BTEC Access course Others None	BTEC Baccalaureate Others None

* Entrants whose highest qualification on entry is A-levels but who did not enter via UCAS, so do not have tariff points recorded, are allocated to medium risk.

† Entrants with 'unknown' entry qualifications or unknown UCAS tariff points have been assigned to the 'low risk' category. Institutions should ensure that entry qualifications are fully recorded if students are to be weighted appropriately in the allocation method for this stream of funding.

Part-time

20. This allocation is likely to be distributed pro rata to London weighted part-time undergraduate (including foundation degree) assumed FTE for 2007-08 (HESES06 Column 4 HEFCE-funded FTE plus adjustments for additional student numbers and mergers/transfers). Therefore, the web facility does not produce derived statistics for the part-time improving retention allocation.

Widening access for students with disabilities

21. This allocation is likely to be calculated using HESA 2005-06 data as follows.

22. Firstly, we calculate the proportion of eligible students who were in receipt of the Disabled Student's Allowance (DSA). Students are only part of the population if they generate a HEFCE-fundable Column 4 countable year in the HEIFES05 re-creation.

23. Each institution is assigned to one of four quartiles, according to the proportion of students in receipt of the DSA, although this is smoothed to ensure that no institution falls by

more than one quartile since the previous year. Separate weightings are attached to each of the four quartiles, as follows:

Quartile	Weighting
A (lowest proportion)	1
B	2
C	3
D (highest proportion)	4

24. Institutions' share of the allocation is pro rata to the assumed FTE for 2007-08 (HESES06 Column 4 HEFCE-funded FTE plus adjustments for student numbers and mergers/transfers), weighted according to the quartile in which they fall and a London weighting, although a minimum allocation of £10,000 for each HEI applies.

Annex G

Information for leads and members of HEFCE-recognised funding consortia

Purpose

1. This annex provides institutions that are leads and members of HEFCE-recognised funding consortium agreements with information about the outputs that are available to them via the web facility. It also contains more detailed information about the worksheets within the HESES05 re-creation outputs that are available to them.

Data collection arrangements for HEFCE-recognised funding consortia

2. Under HEFCE-recognised funding consortia arrangements, each student is recognised as a student of the appropriate consortium member, rather than of the lead. Therefore lead institutions of HEFCE-recognised funding consortia must submit data to the HESA only for students that are registered at their institution. Likewise students that are registered at a member further education college will be returned on that college's 2005-06 ILR.

3. All member colleges' data for students funded under the consortium arrangement should have been returned by the lead institution on its HESES05 return. Therefore, for the lead institution of a HEFCE-recognised funding consortium, the HESES05 re-creation will incorporate the 2005-06 HESA data that it submits to the web facility together with the ILR data that each member college has submitted to our web facility. A similar arrangement is made for the 2007-08 widening participation allocations.

4. Annex C provides further details about the HESES05 re-creation. Annex F provides further details about the 2007-08 widening participation allocations.

Web facility outputs

HESES05 and the HESES05 re-creation

5. The web facility generates a provisional HESES05 re-creation. This is described in Annex C. Annex B contains a description of how to access the HESES05 re-creation Excel workbook from the HEFCE extranet. The information contained in the HESES05 re-creation will vary depending on whether or not the institution is the lead of a HEFCE-recognised funding consortium.

6. Lead institutions of a HEFCE-recognised funding consortium will be able to access the following files:

- LEAD05XXXX.xls – This is an Excel workbook containing the HESES05 re-creation tables for the lead HEI, which is constructed from the 2005-06 HESA data that the lead has uploaded to the web facility.

- HESR05XXXX.xls – This is an Excel workbook containing the HESES05 re-creation tables that is constructed from the 2005-06 HESA data that is uploaded by the lead institution as well as ILR data that is uploaded by member colleges to the ILR web facility.
- HEIRC05YYYYYY.xls – These are Excel workbooks containing the HEIFES05 re-creation tables for each of the member colleges (where YYYYYY denotes the UPIN provider number for the college). These workbooks contain the 2005-06 ILR data that the member college has submitted to our web facility. Note that these tables will only become available after member colleges have submitted data to the web facility.
- HESR05XXXX.ind – This is the individualised file that supplements the provisional HESES05 re-creation tables for the lead institution (LEAD05XXXX.xls). All of the information contained in the HESES05 re-creation tables for the lead institution can be re-built by categorising and aggregating the data in this file. It contains details, in the form of 2005-06 HESA fields and derived fields, of how each student was classified in the re-creation. A full description of the data contained in the individualised file is given in Appendix 1.

7. Member colleges of a HEFCE-recognised funding consortium will be able to access the following files from their ILR derived statistics web facility:

- HEIFERC05YYYYYY.xls – These are Excel workbooks containing the HEIFES05 re-creation tables for the member college.
- HEIFER05YYYYYY.ind – This is the individualised file to supplement the provisional HEIFES05 re-creation tables for the member college (HEIFERC05YYYYYY.xls). All of the information contained in the HEIFES05 re-creation tables for the member college can be re-built by categorising and aggregating the data in this file. It contains details, in the form of 2005-06 ILR F04 and derived fields, of how each student was classified in the re-creation. A full description of the data contained in the individualised file is given in Appendix 1.

Derived statistics likely to inform the 2007-08 widening participation allocations

8. The web facility generates illustrative 2007-08 widening participation allocations. These are described in Annex F, together with a description in Annex B of how to access the HHWP05 re-creation Excel workbook from the HEFCE extranet. The information contained in the HHWP05 re-creation will vary depending on whether or not the institution is the lead of a HEFCE-recognised funding consortium.

9. Lead institutions of a HEFCE-recognised funding consortium will be able to access the following files:

- HHWPLEAD05XXXX.xls – This is an Excel workbook containing the widening participation tables for the lead institution, derived from the 2005-06 HESA data that the institution has uploaded.

- HHWP05XXXX.xls – This is an Excel workbook containing the widening participation tables built from the HESA data uploaded by the lead institution. It also contains the 2005-06 ILR data that has been uploaded by member colleges to the web facility.
- IHWPC05YYYYYY.xls – These are Excel workbooks containing the widening participation tables for each of the member colleges built from the ILR data they have uploaded to the ILR web facility. Note that these tables will only become available after member colleges have submitted data to the web facility.
- HHWP05XXXX.ind – This is the individualised file that supplements the provisional 2007-08 widening participation allocations (HHWPLEAD05YYYY.xls). All of the information contained in the WP05 re-creation tables for the lead college can be re-built by categorising and aggregating the data in this file. It contains details, in the form of 2005-06 HESA fields and derived fields, of how each student was classified in the re-creation. A full description of the data contained in the individualised file is given in Appendix 10.

10. Member colleges of a HEFCE-recognised funding consortium will be able to access the following files from the ILR web facility:

- IHWPC05YYYYYY.xls – This is an Excel workbook containing the WP05 re-creation tables for the member college.
- IHWPC05YYYYYY.ind – This is the individualised file that is intended to supplement the provisional 2007-08 widening participation allocations (IHWPC05YYYYYY.xls). All of the information contained in the WP05 re-creation tables for the member college can be re-built by categorising and aggregating the data in this file. It contains details, in the form of 2005-06 ILR F04 and derived fields, of how each student was classified in the re-creation. A full description of the data contained in the individualised file is given in Appendix 10.

Access to outputs

11. Lead institutions will be given automatic access to all outputs except the separate HEIFES05 re-creation and WP individualised files for each of the member colleges. Each member college will receive access to its own HEIFES05 re-creation, and a separate HEIFES05 re-creation individualised file containing only the 2005-06 ILR data that the college has uploaded. This arrangement will also apply for the WP outputs.

12. Where a member college gives us permission to grant the lead institution with access to its HEIFES05 re-creation or WP05 re-creation individualised files, we will write to the lead to describe how it can access the individualised files via the HEFCE extranet.

Worksheets

13. The provisional HESES05 re-creation workbook for lead institutions contains the following worksheets:

Page number	Worksheet (see tabs on spreadsheet)	Title
1	Coversheet	Title page
2	FTS	Provisional HESES05 re-creation Table 1a: Full-time and sandwich years of programme of study
3	MED	Provisional HESES05 re-creation Table 1b: Medical and dental full-time and sandwich years of programme of study
4	SWOUT	Provisional HESES05 re-creation Table 2: Sandwich year-out years of programme of study
5	PT	Provisional HESES05 re-creation Table 3: Part-time years of programme of study and load
6	FEE	Provisional HESES05 re-creation Table 4: Home and EC fees
7	CONS	Provisional HESES05 re-creation Table 6: HEFCE-recognised funding consortia 2005-06
8	Excl	Students excluded from the provisional HESES05 re-creation

14. The provisional HEIFES05 re-creation workbook for member colleges, HEIRC05YYYYYY.xls, contains the following worksheets:

Page number	Worksheet (see tabs on spreadsheet)	Description
1	Coversheet	Title page
2	FTS	Provisional HEIFES05 re-creation Table 1: Full-time and sandwich years of programme of study for the member college
3	SWOUT	Provisional HEIFES05 re-creation Table 2: Sandwich year-out years of programme of study for the member college
4	PT	Provisional HEIFES05 re-creation Table 3: Part-time years of programme of study for the member college
5	FEE	Provisional HEIFES05 re-creation Table 4: Home and EC fees for the member college
6	Excl	Provisional HEIFES05 re-creation exclusion table for the member college
7	Credibility	Identifies areas for recognised HE qualification aims on 2005-06 ILR F04 where data is potentially inaccurate

Annex H

Research degree qualification rates (RDQRs)

Purpose

1. This annex provides details of how we will use HESA student data to measure and report information relating to the rates of qualification obtained by students undertaking research degree programmes. It is expected that institutions will use this opportunity to check that the HESA data used to produce RDQRs are fit for publication. Full details of the algorithms we have used to produce these statistics are provided in Appendix 11.

Background

2. In September 2004 we wrote to institutions announcing new arrangements for the quality assurance of research degree programmes at institutions that receive HEFCE funding (HEFCE Circular letter 18/2004). These arrangements comprise two parts. The first concerns alignment with section 1 of the Quality Assurance Agency (QAA) code of practice on academic standards. The second part is monitoring, through RDQRs, the time that research students take to obtain their qualifications.

3. '2004-05 statistics derived from HESA data: Guide to HEFCE web facility' (HEFCE 2005/32) invited institutions to comment on a proposed methodology for the calculation of RDQRs. We also asked institutions to comment on whether their underlying HESA data were fit for publication.

4. In June 2006 we announced the outcomes from this consultation (HEFCE Circular letter 10/2006 titled 'Developments in HEFCE's use of research degree qualification rates') and in response to concerns raised we announced both a further opportunity for data verification and changes to the methodology. We also confirmed both our intention to pass RDQRs to the QAA to inform its audit work, and publish RDQRs for full-time students during 2007. Paragraphs 17 and 18 below describe the changes to the methodology that have resulted from the consultation.

Data verification

5. Use of the web facility will be the final opportunity for institutions to verify RDQRs for full-time students ahead of publication. We provide the outputs described below to enable institutions to identify and remedy any issues with the accuracy of their RDQRs.

Action plans

6. Institutions that identify that their RDQRs are not suitable for publication should determine the cause of issue and submit an action plan that describes the cause and explains how it will be resolved. Guidance for completing and submitting an action plan is in Appendix 12. We expect to make action plans available on the extranet for download from 1 August 2006.

7. We expect any issues with RDQRs to fall into either or both of the following categories:

- errors in HESA 2004-05 student data
- errors in HESA student data for years between 1999-2000 and 2003-04.

8. If an institution wishes to correct its 2004-05 HESA data for incorporation in the 2005-06 RDQRs for full-time students to be published during 2007 it should submit an RDQR action plan by 8 September 2006.

9. Errors in 2004-05 HESA student data should be resolved by submitting to HESA a revised 2004-05 HESA student return. HESA will not accept amendments to 2004-05 HESA student data until it has received confirmation that we are content with the institution's action plan. Amendments should be submitted by 26 September and subsequently signed off by 9 October 2006 to ensure incorporation in the 2005-06 RDQRs. Appendix 14 describes the process for making amendments to 2004-05 HESA student data.

10. If an institution identifies errors in its HESA student data for years between 1999-2000 and 2003-04 and it wishes to correct the derived fields for the individual students affected, it should submit an RDQR action plan by 8 September 2006.

11. Errors in HESA student data for years between 1999-2000 and 2003-04 should be resolved by submitting an historical data error file to enable us to correct the derived fields for the individual students affected. Corrections should be submitted by 26 September and subsequently signed off by 9 October 2006 to ensure incorporation in the 2005-06 RDQRs. Appendix 15 describes the process for submitting historical data error files.

RDQR statistics outputs

12. Paragraph 28 of Annex B describes how to access the derived statistics, in an Excel workbook, that we intend to publish during 2007 (RDQR05XXXX.xls, where XXXX is the HESA identifier for the institution). This workbook contains the following worksheets:

- Cover page
- LEVEL1_ALL – RDQRs after seven years for 1999-2000 full-time students
- LEVEL1_SECTOR – sector RDQRs after six years for 1999-2000 full-time students (2005-06 HESA data for the sector is still unavailable)
- LEVEL2_ALL – RDQRs for all students in all cohorts
- LEVEL2_HOMEEC – RDQRs for home and EC students in all cohorts
- LEVEL2_ISOVI – RDQRs for island and overseas students in all cohorts
- LEVEL2_SECTOR – sector RDQRs for all students in all cohorts (2005-06 HESA data for the sector is still unavailable).

13. We intend to publish the statistics presented in the worksheet 'LEVEL1_ALL'. The remaining worksheets allow comparison of RDQRs against the sector (note that 2005-06 is not available for the sector so the latest sector tables are for 2004-05 data), and further verification of the entire dataset so that we can look to publish RDQRs for more cohorts in future years.

14. The information contained in the RDQR tables can be re-built from the individualised file provided by the web facility. The file (RDQR05XXXX.ind) contains details of how each student was classified in the RDQR tables, and where relevant, details of why they were excluded from the population. A full description of the data contained in the individualised file is given in Appendix 11.

15. This is also an early opportunity for institutions to verify data that will be published in the future. While we do not expect to publish data for part-time students until 2009 (and for students that switch from full-time to part-time study, and vice versa mid-study), we intend to publish RDQRs for full-time students each year from 2007 onwards.

RDQR sign-off

16. In August 2006 we will write to institutions giving details of the process for confirming that their RDQRs for full-time students are suitable for publication.

Table 5 Indicative timetable for RDQR verification process

Indicative date/deadline	Process
August 2006	We write to institutions giving details of the process for confirming that their RDQRs for full-time students are suitable for publication
8 September 2006	Deadline for submitting an RDQR action plan
26 September 2006	Deadline for submission of amendments to 2004-05 HESA student data or historical data error files to correct derived fields for incorporation in 2005-06 RDQRs
9 October 2006	Deadline for signing off amendments to 2004-05 HESA student data or historical data error files to correct derived fields for incorporation in 2005-06 RDQRs
August 2006 - October 2006	Institutions sign-off their 2005-06 RDQRs
December 2006	We receive 2005-06 student data from HESA and produces 2005-06 RDQRs
January/February 2007	We pass 2005-06 RDQRs to QAA to inform its audit activities
Summer/Autumn 2007	We publish 2005-06 RDQRs

Changes to methodology

17. In HEFCE Circular letter 10/2006 we announced the outcomes from last year's consultation on the RDQR methodology. This resulted in the following changes to the methodology, which otherwise is the same as described in HEFCE 2005/32. An overview of the RDQR methodology is given below.

Student commencing their studies before 1 August 1999

18. The consultation highlighted that most errors in RDQRs were for 1995-96 to 1997-98 HESA student data. Therefore we decided to restrict RDQRs to students commencing their studies after 31 July 1999. This will limit RDQRs to the period after 1998-99, after between-years validation was introduced to the HESA record. The result will be increased consistency

of data that span several academic years. This will enable us to link research degree programmes between years using the HUSID x INSTID x NUMHUS (HIN) fields on the HESA record.

Transfers to other institutions

19. We will include an additional context statistic on the number of students who transfer to another institution. The methods we will use to identify such students are described in Appendix 11.

Overview of the RDQR methodology

RDQR population

20. For each year, starting with 1999-2000, we have identified the population of students that started aiming for a postgraduate research degree (MPhil, PhD and comparable qualifications). Each population is tracked across academic years, up to 2005-06, in order to derive cumulative RDQRs across time up to 2005-06. We removed students that left in the first year of the qualification aim. Students who die are excluded in the year that they die.

RDQR countable

21. Students in the RDQR population become countable towards RDQRs when they start aiming for a doctorate research degree (PhD and comparable qualifications). For example, if a student is aiming for an MPhil they will only become countable towards RDQRs if they later change their qualification aim to a doctorate research degree. In this example they will be in the RDQR population of the year that they started aiming towards the MPhil.

Qualifiers

22. Students that become countable towards the RDQR are regarded as qualifiers when they obtain their doctorate research degree.

RDQR calculation

23. RDQRs are calculated by dividing the number of qualifiers by the number of students that become countable towards the RDQR.

Groupings

24. For each year the rates have been grouped into the following categories depending on their mode:

- full-time students
- part-time students
- students that switch mode during the qualification aim.

25. Similarly the students have been grouped as either home and EC, or overseas, depending on their fundability and residential status.

Context statistics

26. Four context statistics are presented alongside the RDQRs: the number of active students, the number on science-based courses, the number receiving funding from the UK Research Councils and the number that transfer to another institution.

27. Active students are defined as those that are countable towards RDQRs who have not been awarded a doctorate research degree, but who are actively pursuing studies at the point of calculating the RDQRs.

28. For this purpose, science-based courses are defined as medicine, veterinary science, subjects allied to medicine, biological sciences, physical sciences, engineering, mathematics, and agriculture.

Annex I

HEFCE statistics used for publication and to inform policy

Purpose

1. This annex describes the outputs we intend to publish, or use to inform policy decisions during 2007 once these data are submitted to the Higher Education Statistics Agency.
2. Institutions should note that the web facility offers institutions the only opportunity to check and correct the derived statistics that will be used in HEFCE statistical publications during 2007.
3. We will generate data from HESA 2006-07 student data covering four broad areas:
 - students registered at one institution and taught by another institution (also referred to as 'franchised')
 - campus data
 - distance learning data
 - provision by location.

These data may also be used for regional analysis and to inform any allocations of growth in student numbers.

4. Details of the algorithms used in generating the data for publication are given in Appendix 12. Where appropriate we have included derived fields on the individualised student data file, to allow institutions to better understand our derivations made from the data.

Web facility outputs

5. The web facility generates the derived statistics that will be used in HEFCE statistical publications, in Excel file PUB05XXXX.xls. Annex B describes how to access these results.
6. The information contained in the published tables can be re-built from the individualised file which we provide via the web facility. The PUB05XXXX.ind file contains details of how each student was classified in the re-creation, and, where relevant, details of why they were excluded. A full description of the data contained in the individualised file is given in Appendix 12.

Students registered at one institution and taught by another institution

7. In HEFCE 2006/05 we used the 2005-06 HESA student data to publish details of students registered at one institution but taught by another. We intend to generate the equivalent data for 2006-07 to enable institutions to check that they are accurate and suitable for publication. We expect to publish this information during 2007.

8. Where students are registered at one institution and taught by more than two institutions, only two teaching institutions are shown in the tables in addition to the registering institution.

Campus data

9. Campus data returned on the HESA 2006-07 student record will be used in regional analysis. We are providing summary campus information to enable institutions to check that it is accurate and suitable for publication. In particular, institutions must ensure campuses that are a significant distance apart are separately identified on the HESA student record. Campuses are defined as being a significant distance apart if they are in different cities or are over 25 miles apart.

10. We encourage institutions to identify separate campuses wherever they are geographically distinct. Where institutions have merged since 31 July 2002 we have created a separate entry for each campus of the merged institution. We would expect institutions to continue to make this distinction in future years.

Distance learning data

11. We intend to publish data on distance learning in the 'Learning in the region' page of the annual HEFCE regional profiles publication. This page shows the extent to which the number of students registered at institutions in the region differs from those that are learning in the region. We assume that distance learners study in their region of domicile. We wish to publish both the numbers registered in the region but domiciled outside it, and those that are registered outside the region but domiciled inside it.

Provision by location

12. We use a database of provision by location, derived from the HESA student record and the LSC's ILR, for regional analyses. It is likely that we will continue to use similar data in the future. We expect to publish this information during 2007. Therefore, it is essential that we have accurate data about where students are taught.

13. The location where students are taught is derived from either the data described in paragraphs 7 and 8 above or the campus at which the student is taught, as described in paragraph 9 above. Where neither campus information, nor information about students that are registered at one institution and taught by another, is given, students are assumed to be taught at the institution's main campus.

14. The following geographical categories based on the teaching institution's postcode are included in the database: region, county, county/unitary authority, local education authority and the LSC area.

15. Students who study via distance learning or are on a sandwich year out are not included in the database of provision by location.

16. The database also contains details of the student's mode and level of study as well as broadly grouped subjects of study.

17. To improve the quality of the data, we intend to exclude records that are duplicated. This may result from duplications between HESA and the LSC's ILR returns as well as duplicates within the HESA return. As we cannot remove duplicates until all institutions have completed their returns this will have to be done after submission. The removal of duplicates may result in differences between tables created by the web facility and those that we eventually publish.

Annex J

HESES06 non-completion toolkit

Purpose

1. This annex contains details on how to use the HESES06 non-completion toolkit as part of an institution's estimation of non-completion rates on HESES06. This annex must be read in conjunction with and used in accordance with the forthcoming HESES06 publication in which full guidance relating to the return of non-completion forecasts will be given.

Industrial action affecting student assessment

2. At this time it is unclear whether industrial action by academic staff unions over student assessment in 2005-06 has affected students' 2005-06 completion status. Where disruption has occurred, for example where assessments have not been set and consequently students will not take the usual final assessment, institutions are advised to use the outputs of the toolkit with caution as the non-completion rates for some courses may be inaccurate and atypical. In such cases we would expect institutions to base their HESES06 forecasts on the rates applied to HESES05, adjusting for any known errors in these rates. The forthcoming HEFCE publication HESES06 will make clear that estimates of non-completion informed by historical data should not be skewed by exceptional circumstances such as industrial action affecting exams.

Description of the non-completion toolkit

3. The non-completion toolkit is a spreadsheet that allows institutions to generate estimated non-completion rates for all students by course and year for 2005-06, from the submitted HESA 2005-06 student data using fields that they consider influential at the forecast level. The estimated values generated may be used by institutions to inform the forecasts of non-completions returned in HESES06.

Reliability of estimated non-completion rates

4. Institutions should note that the toolkit is provided purely as an aid in generating non-completion rate estimates in HESES06. We expect institutions to use their own judgement in assessing whether the rates generated are applicable.

5. Estimates may be unreliable if the numbers of students involved are small, or if the underlying HESA data contain errors or is not completed in line with the HEFCE guidance on non-completion status. Annex E of 'HESES05' (HEFCE 2005/39) gives definitive guidance on non-completions.

Using the non-completion toolkit

6. Users should follow the instructions in Annex B to access the file NCTK05XXXX.xls, then open in Excel 2000 (or a later version). Due to the technology used the toolkit will not

function correctly in earlier versions. The creation of tables uses macros, so these should be enabled to use the facility. The worksheet called 'NCdata' contains the raw data used to generate the tables, and the 'Pivot Table' worksheet displays the results. When you open the file you will be prompted to select the fields you want to use to generate the non-completion rates – we advise summarising by course in the first instance. To create the table click 'OK', and the screen will close, enabling you to view the table. Click on 'Create a new table' to repeat the process – this will delete the previous table from the worksheet.

7. Non-completion rates based on small numbers of students may be unreliable, so any rate based on a population of less than 100 is coloured red and noted as unreliable. In these cases we advise using less detailed data in order to obtain a more reliable estimate.

8. In order to save a table you have created, open the <File> menu and select <Save As>.

List of abbreviations

ASNs	Additional student numbers
CFTE	Contract full-time equivalent
DSA	Disabled Student's Allowance
EC	European Community
FAQ	Frequently asked questions
FE	Further education
FEC	Further education college
FTE	Full-time equivalent
FTS	Full-time and sandwich
HE	Higher education
HEFCE	Higher Education Funding Council for England
HEI	Higher education institution
HESA	Higher Education Statistics Agency
HESES	Higher Education Students Early Statistics
HIN	HUSID X INSTID X NUMHUS
INSET	In-service education for teachers
ILR	Individualised learner record
ITT	Initial teacher training
JACS	Joint Academic Coding System
LSC	Learning and Skills Council
PGCE	Postgraduate certificate of education
QAA	Quality Assurance Agency for Higher Education
QTS	Qualified teacher status
RAS	Research Activity Survey
RDP	Research degree programme
RDQR	Research Degree Qualification Rate
UOA	Unit of assessment
WP	Widening participation