

HEFCE – Issued April 2008

Further guidance on evaluation planning

Introduction

1. Partnerships have indicated at cluster meetings that they would welcome further guidance on evaluation and that a 'template' for plans would be helpful. The complexity of the evaluation task and the differences in scale, resources and circumstances of partnerships as well as differences in approach mean that we remain reluctant to be overly prescriptive. However, we are now commissioning a substantial research project into the impact of Aimhigher/WP (which a number of partnerships have helped us refine). We think there is a real possibility of bringing national research and local evaluation together in complementary ways: we want to bring area level evidence into a national framework and give it additional credibility:
 - a. The first phase of the national research will help to test and refine some of the new categories and methods set out in this guidance for local evaluation;
 - b. To the extent that local evaluation is based on data that is at least *broadly* comparable (participant characteristics, roughly comparable Aimhigher/WP experiences, participant outcomes related to attainment and progression) such data can contribute to a *national* as well as local evaluation.
2. This will involve more consistency of data and methodology. To achieve this we will need partnerships to adopt a standard approach to (i) participant data and (ii) participant attainment /progression outcomes (iii) to use structured inputs from professionals and learners that link Aimhigher/WP experience and outcomes. We discuss this in more detail below.
3. We also think that this approach to evaluation requires us to categorise activities (similar in type) for data collection purposes, and to distinguish 'levels' of experience of the Aimhigher programme so that we are more confident that we are comparing 'like with like' when we ask about outcomes for Aimhigher participants. We acknowledge, of course, that the definitions of 'categories' of activity, or the difference between levels of experience will have to be tested by partnerships and by the research. Again this is discussed in more detail below.
4. This further guidance does not claim to be comprehensive. It is more prescriptive than the guidance that we usually offer but it is prescriptive about a limited number of core things. Partnerships are encouraged to ask complementary research questions and use the evidence base along with ancillary evidence to answer them. You will note that the 'template' below says that 'Partnerships will wish to define a range of outcomes but these must include outcomes for attainment and progression'. The partnership's evaluation plan will therefore

still involve decisions about research foci/questions which address other issues such as embedding, culture shift, strength of partnership, performance of subgroups of the learner cohort, and so on provided that there is also key information on the validity of the activity/programme in relation to the learners and that there is good participant data.

5. The template offered below is the basis for a plan, not the plan itself. The plan will link evaluation priorities with the strategic plan and its aims and objectives and performance measures, will set out the scope and scale of the evaluation to be undertaken (including resource commitment), how layers of evidence will be built up, and may articulate with other evaluation/research programmes. We expect the CSET evaluation 'toolkit' to offer a similarly broad view. This further guidance sets out the core of what partnerships must do; it does not limit what partnerships can do.

6. This further guidance is in two sections. First, there is a summary of what is required in respect of participant data, data on outcomes for participants, and the evaluative questions and processes that link them. Secondly, there is a template that can be used to write (or revise, or develop) an evaluation plan. A brief discussion of the issues that have led us to this guidance/template are given in the Annex.

Summary

7. There are three essential components of evaluation
 - a. participant data
 - b. data on outcomes for participants
 - c. the evaluative questions and processes that link them

Participant data

8. There can be no evaluation of Aimhigher/WP without *good* data about participants. That data must include
 - a. Name, DoB (standard format dd/mm/yyyy), gender, ethnicity, disability,
 - b. Post code
 - c. Occupational background (mum, dad/carers and who is the chief wage earner; with code for NS-SEC)
 - d. Parent/carers experience of higher education
 - e. School/college/training provider

9. Participant data clearly also needs to record details of involvement in activity (activity type/date/duration). These data can be stored in a data base (or spreadsheet), with the individual as the key field. The difficulty is that data cannot be collected about *all* activity, and almost certainly therefore, not on *all* participants.

10. In some areas schools/colleges have defined an 'Aimhigher cohort'. Where this is the case it should be straightforward to collect the data above (a. to e.). If there is an 'open group' rather than a cohort but the core participants have been identified, it should be possible to record data about members in a similar way. When partnerships record data about participants in *activities* it should be possible to flag those who are part of a cohort/core.

11. We propose that partnerships record details of participants/activities as follows:

a. **All** participants for between **10% and 20%** of large group/short duration activity which we will call **category 1** (for more detail see Annex para.4)

b. **All** participants in between **30% and 50%** of high intensity activity (linked series of events over time, relatively smaller numbers) which we call **category 2** (for more detail see Annex para.5)

12. We specify 10%-30%, and 50% to 60% to take account of differences between partnerships. The important thing is to be as clear as it is possible to be (and in practice it will only be approximate) about what the percentage is.

13. All data collected, whether directly from participants or from parents/carers, should have the appropriate permissions for data protection purposes, and include where it is possible to obtain it, permission for sharing data for research and evaluation purposes.

Data on outcomes for participants

14. Aimhigher/WP activities/programmes are designed to achieve a purpose (raising aspirations, building confidence, closer engagement with learning, improved attainment, improved access to HE). Outcomes may be defined for specific activities (e.g. mentoring), or for the programme as a whole at key transition points (14, 16, exiting post-16 learning or progression). Outcomes should be measurable and data available to measure them. Among the outcomes partnerships define, improved attainment (particularly differences between predicted and actual outcomes) and progression must be prominent.

15. Attainment and progression outcomes are discussed further below. At this point we note that some partnerships are reaching data sharing arrangements so that outcomes, at least for a defined cohort, can be ascertained. Where there are no arrangements of this sort

recorded outcomes will be more limited and depend on good partnership links with key schools and colleges.

16. Attainment and progression data must be about the participants who have been involved in activity. Data about attainment and progression for the area/school/local authority is useful as context but is no substitute for data about participants.

17. *Partnerships should regard good participant data as the key. Where this exists links with administrative data sets might be possible through a national research project or service.*

Evaluation

18. The first two steps above are about data and linking data. The evaluation task is to state and explain the link between the experience of participants and the learner outcomes being explored. We would expect the evaluation to use:

- a. consistency and scale of association between participant involvement and outcomes, in particular differences between predicted and more positive actual attainment (where/if this exists);
- b. the professional judgment of teachers (senior staff and head teachers) where these judgments are adequately supported and evidenced;
- c. evidence from participants themselves where this is related to outcomes (separated from the delivery of specific activities) and is also rigorously interrogated and evidenced.

19. It is important, so far as possible, to compare like-with-like. Where the outcome evaluated relates to a specific activity we can set aside other circumstances and ask about e.g. the specific outcomes of a mentoring programme for mentees. Where the outcomes relate to the programme at e.g. age 16, some account must be taken of the extent to which participants have been involved in a number of activities, and the *category* of activity. We propose two levels of participation as a basis for establishing 'groups' of participants with broadly comparable Aimhigher experiences against which learner outcomes and explanations can be mapped. These will link back to the categories of activity defined above.

- a. **Experience 1:** those whose involvement is confined to category 1 activity (low intensity/short duration) and no more than one category two activity.
- b. **Experience 2:** those whose involvement includes category 1 activity (any number), and a minimum of three category 2 activities over two to three years (not necessarily consecutive) e.g. including a summer school and mentoring.

Template

Participant data	The data that will be collected about individuals (whether for a cohort/open group);	The minimum data collected must include Name, DoB, gender, ethnicity, disability Post code Occupational background Parental HE experience School/college/training provider
	what proportion of category 1 activity (useful to specify activity) category 2 activity (useful to specify activity)	
	The target number of learners that the evaluation will involve	
Participant outcomes	Partnerships will wish to define a range of outcomes but these must include outcomes for attainment and progression.	For each of the outcomes specified state how the data will be obtained
<i>For specified activities</i>	indicate the proportion of the activity type where outcomes are specified	
<i>For the programme as a whole</i>	Specify the 'level' (experience 1 and 2) of involvement for the group	
Evaluation methods	How data on participants and outcomes will be analysed	Whether outcomes can follow the learner trajectory

		from e.g. age 14 (using assessment on entry to secondary school and predictions as well as actual outcomes)
	How the professional judgement of teachers will be accessed	For teachers and learners how e.g. semi structured interviews and focus groups can move beyond before/after feedback from activities
	How the learner voice is to be accessed	

Partnership judgement

20. Partnerships must use their judgment about *how much* of this they can do. All partnerships must collect data in the ways set out. This provides the consistency in the potential evidence base that we must have in place, but the scale on which participant data is matched to outcomes, and evaluated *is for partnerships to decide*. Partnerships will not necessarily use all the participant/outcome data in their evaluation. It is much better for partnerships to carry out a very small scale evaluation, for example in a small number of schools where partnership relationships are strongest, and for a small number of discrete activities (e.g. mentoring, master classes), or a small group, and do it well than attempt a larger scale exercise for which the resource and expertise is not available.

Annex: Guidance Discussion

The problems of identifying impact

1. There are a number of problems associated with identifying the impact of WP, the most obvious of which is separating the influence of WP from a range of other influences, and identifying the specific contribution of WP to outcomes for learners. However, there are others. The data on WP participants is difficult to gather and manage in a coherent way, and even when data on participants is available it is not easy to match it with outcomes for those specific learners.

Data collection and analysis

2. There can be no evaluation of WP without data. Practice across partnerships is variable but the ability to match participant data with measurable outcomes for more than a relatively small cohort is uncommon. That said, this guidance acknowledges the real difficulties involved in this. Standardization is very difficult indeed:
- a. Some partnerships have agreed Aimhigher “cohorts” in schools (much less often in colleges) but *actual participants* in some activities might involve a much wider group than the cohort, or only involve a small number from the cohort. The cohort is not *necessarily* a guide to participants.
 - b. There is a huge range of WP activity, some of it limited exposure for whole class or even whole year groups to a particular activity where it would not be feasible or sensible to attempt data capture
 - c. Even a ‘standard list of activities’ actually covers a wide range of experiences. Master classes vary widely in content, challenge, depth, extent; summer schools are from 2 to 10 days in length, residential and non residential. They are in effect broad categories
 - d. Even with standardized data about individuals and activities, actual experience of the programme has to be factored in: how extensive and intensive has participants’ involvement been?
3. We have attempted to simplify to the extent that we propose *categories* of activity and *levels* of experience.
4. Category 1 encompasses large group/short duration activities, typically whole class/whole year group campus visits; student ambassador (and other) talks to schools, and generic WP activities.

5. Category 2 encompasses longer events involving a linked series of experiences over a number of days/weeks and relatively smaller numbers. They include master classes, mentoring (more than a single session), summer schools, one-day (or half-day) events involving specific, more intensive and focused work.

6. As noted above, it would make little sense to compare the participation of a learner in a single Aimhigher activity with someone who has been regularly involved in activity over time and treat them equally when considering the question whether Aimhigher activity had a positive influence on attainment and progression. We propose two levels of participation (to be tested by the research) as a basis for establishing 'groups' of participants with broadly comparable Aimhigher experiences against which participant outcomes and explanations can be mapped. These will link back to the categories of activity defined above. It is acknowledged that there would be a strong case for at least three levels of involvement but this might be developed later or reviewed in the light of the research.

7. **Experience 1:** those whose involvement is confined to category 1 activity (low intensity/short duration) and no more than one category two activity.

8. **Experience 2:** those whose involvement includes category 1 activity, and a minimum of three category 2 activities over two to three years (not necessarily consecutive) e.g. including a summer school and mentoring.

Participant outcomes

9. The availability of data on outcomes for participants has always been limited by the need to avoid 'burden' for schools. Partnerships are now more mature and we need to revisit the question of how data on participant outcomes are obtained. Without it, no serious evaluation is possible.

10. Some partnerships are entering into data sharing agreements with local authorities that will enable partnerships to track the attainment results of individual Aimhigher participants; in the best case this will be done automatically (i.e. through linked data bases). Others have excellent partnership relationships with schools that enable them to do this manually in-school for 'cohorts' through Aimhigher coordinators. This includes, potentially, the assessment of learners on entry to secondary school as well as predicted and actual outcomes at key stage 3 and 4 as well as attainment at age 18/19. A major outcome of the research we wish to commission will be to establish best practice in this area and assist partnerships in adopting it more generally.

11. It may be possible to do more. To the extent that participant data sets, standardised and comparable in the ways described above, are available from *all* partnerships it would be possible to commission a *national* analysis of those data sets for the purposes of attainment and progression. This would obviate the need for a great deal of work locally. On the other hand it removes the scrutiny of results from the direct involvement of local partners and this

could weaken important local links. Evaluation is part and parcel of a process that refines aims and objectives, reviews and revises activities, and in the best cases, is incorporated in school and college improvement plans.

12. However, it may be possible to combine local management of evaluation through data sharing and shared analysis with local authorities and others through a national service to partnerships who wish to take advantage of it. This depends on good participant data, capable of being presented in a common format.

13. For post 16 learners in colleges the match will be made using the ILR, again depending on arrangements made with the local LSC. In some cases partnerships are working with Connexions who have flagged Aimhigher learners on their data base but Connexions do not hold information on predictions and this could form an important part of the analysis.

14. Progression is more difficult to measure. At an aggregate level Connexions can report the proportion of Aimhigher learners who continue in full time learning and for some 'cohorts' this will be a significant measure of achievement.

Drawing conclusions

15. All the discussion above about data, and the steps necessary to relate that data to participant outcomes only sets the scene for the key evaluation question: what did the participation in Aimhigher/WP contribute to those outcomes i.e. were those outcomes better than they might otherwise have been. We would expect that connection to be established in three ways

16. First, the extent to which attainment (and progression) of learners participating in Aimhigher is 'better than predicted' in comparison to CAT scores at age 11, key stage 3 and so on.¹ We are particularly interested in outcomes for cohorts of Aimhigher learners at key stage 4 for whom there are participant data from year 9, but in broad terms for learner outcomes at any age for which there is a two to three year record of participation. The consistency and scale of such an association would be significant and could be compared to national data for all learners and/or those with similar characteristics.

17. Secondly, we would expect to evaluate the link between participation and outcome through the professional judgement of teachers. In particular, we would expect senior teachers with some responsibility for overall school performance (particularly head teachers, obviously) such as heads of year, subject leaders, heads of 6th form and so on to be able to make an informed judgement of the value of Aimhigher activity directly for participants and

¹ Some partnerships use 'current actual' performance at a given point rather than predictions which are regarded as unreliable. This is a question the research should clarify.

for the school/college. We would expect much more than a canvassing of opinion. **Semi** structured interviews (possibly focus groups) would aim to establish not only the judgement made but the reasons for it.

18. Thirdly, we would expect to be able to evaluate the link by asking participants themselves. We do not mean participant responses before/after particular activities. We would expect semi structured interviews and focus groups with, for example, learners now in post 16 study about the past and current value of *the programme* and of specific activities. As with teachers the research would aim to establish the reasons for judgements made as well as record the consistency with which they are made.

19. Finally, for some groups of Aimhigher participants it should be possible to track numbers in higher education and, ideally, compare the proportion of the sample entering HE with *all* of those with similar characteristics nationally. HEFCE will do this with summer school participant data. We could do it for another data set e.g. for students who had experienced a given number/type of activities over a period of time. The difficulties associated with this are not to do with the 'tracking' but in getting good participant data.

The specific contribution of WP

20. The key evaluation questions often seem too difficult to answer: did the WP programme lead to these outcomes or was it due to other influences? If WP did make a contribution, how much of the outcome can be attributed to the WP programme? Aimhigher, coupled with the complementary commitments of HEIs, is a national programme. There seems to be no scope for setting up a social science experiment in which the experiences of a 'WP group' is compared with a control group, similar in all other respects. This might be feasible on a very small scale but not with significant numbers. It seems impossible to separate the specific contribution of Aimhigher/WP.

21. However, we may have made too much of the problem. It may be *good enough* for the practical purposes of policy makers and politicians if we can make an evidence-based *judgement* on the *balance of probabilities* that a given programme is producing reasonable results. And if this is all that is available to us the choices are clear.