

Prepared for: Heritage Lottery Fund

Heritage Grants: A Review of the Self-Evaluations and Outcomes of 326 completed projects funded during Strategic Plan 3

Final Report

June 2018

CONTENTS

- 1 EXECUTIVE SUMMARY 3**
- 2 PROJECT BACKGROUND AND APPROACH 7**
- 3 PROJECTS REVIEWED 9**
 - 3.1 SPREAD OF PROJECT REPORTS EVALUATED 9
- 4 PERFORMANCE AGAINST EVALUATION CRITERIA..... 11**
- 5 FACTORS WHICH INFLUENCED THE QUALITY OF REPORTS..... 13**
 - 5.1 OVERVIEW 13
 - 5.2 EVALUATION WAS EXTERNAL 14
 - 5.3 PLANNED SHARE OF GRANT INITIALLY ALLOCATED TO EVALUATION..... 16
 - 5.4 REPORT LENGTH..... 18
 - 5.5 HLF OUTCOMES MAPPED ON TO THE 326 REPORTS 19
 - 5.5.1. Overview 19*
 - 5.5.2 Correlation between number of outcomes met and quality of reports 20*
 - 5.5.3 Correlation between categories of outcomes met and quality of reports..... 20*
 - 5.5.4 Investigating links between mapped outcomes and other project characteristics: grant size, heritage area, region..... 21*
- 6 UNDERSTANDING THE INCIDENCE OF MAPPED OUTCOMES 23**
 - 6.1 OUTCOMES FOR HERITAGE 24
 - 6.2 OUTCOMES FOR PEOPLE 25
 - 6.3 OUTCOMES FOR COMMUNITIES..... 26
- 7 CONCLUSIONS 29**
- 8 RECOMMENDATIONS 30**
 - 8.1 RECOMMENDATIONS: IMPROVING THE OUTCOMES FRAMEWORK..... 30
 - 8.2 RECOMMENDATIONS: IMPROVING THE QUALITY OF SELF-EVALUATION REPORTS..... 31
- APPENDIX CONTENTS 33**
- APPENDIX 1: PERFORMANCE AGAINST EVALUATION CRITERIA 34**
 - 1. CRITERION 1: TELLING THE PROJECT STORY..... 34
 - 2. CRITERION 2: COUNTING, INVOLVING, CHOOSING INDICATORS THAT MATTER..... 34
 - 3. CRITERION 3: BEYOND COUNTING 37
 - 4. CRITERION 4: AVOIDING BIAS 39
 - 6. CRITERION 6: IMPROVE NOT JUST PROVE 41
- APPENDIX 2: ANALYSIS OF HERITAGE AREA, REGION AND FINANCIAL SIZE OF PROJECT 42**
- APPENDIX 3: OVERALL SUMMARY OF SELF-EVALUATION REPORT STRENGTHS AND WEAKNESSES 45**
- APPENDIX 4: HLF ADOPTED APPROACH TO REVIEWING GRANTEE EVALUATIONS..... 47**
- ANNEX A – CRITERION SCORING 50**



1 Executive Summary

1. The Heritage Lottery Fund (HLF) commissioned RF Associates to undertake a comprehensive review of the self-evaluation process for 326 of the completed Heritage Grants projects funded under HLF's 3rd Strategic Plan (SP3) which ran between 2008 and 2013.
2. This review was conducted in two distinct stages: a review of 200 self-evaluation reports in 2017 (which built on an initial review of 93 reports), followed by a review of a further 126 self-evaluation reports in 2018. This report brings together the findings from these reviews and includes a comparative appraisal of the quality, scope and methodology of the self-evaluated reports against six evaluation criteria and the type, range and quality of activities and outcomes achieved by completed projects. The review also includes a mapping of the project outcomes onto HLF's current framework of 14 outcomes for heritage, people and communities.
3. Overall just over a third, 37%, of the 326 self-evaluation reports were graded as good or excellent with just under two thirds, 63%, falling within the adequate or poor categories. 10% of reports were assessed to be excellent and 23%, were considered poor.
4. The aggregated findings in this report show that the quality criteria scores have significant dependence on several characteristics, confirmed by a range of statistical tests:
 - **Evaluation was external:** external consultants/organisations tended to write better quality reports.
 - **Planned share of grant initially allocated to evaluation:** where higher amounts of expenditure had been originally allocated for evaluation, the reports tended to be of better quality.
 - **Report length:** longer reports which contained more explanation and more data, tended to be of better quality than the shorter reports.
 - **The number of HLF outcomes assessed as being met:** better quality reports tended to record a greater number of outcomes being met.
5. No statistical dependence could be found for heritage type, region, financial size of the project or whether the assessment was from the 2017 or 2018 studies assessed by RF Associates.
6. The reviews of the self-evaluations have included an assessment of which of the 14 outcomes set out in the HLF Outcomes Framework can be attributed to each project. The statistical analysis finds that there is an apparent strong link between the assessment scores and the sum of outcomes achieved. There is, though, not such a clear rationale for this link as for the other significant factors. A plausible explanation for the strong relationship is that the high degree of correspondence derives from the review process. Criteria scores and outcomes - determined by the same assessors - both act as indicators of self-evaluation quality.
7. As we found in our prior report, there is an inherent tension within self-evaluation between a project telling its own story to an external audience, whilst also avoiding bias. Many reports seemed to be written primarily as project history documents, rather than documents which sought to objectively evaluate their activities and identify openly where things did not go well and could be improved in the future.

8. Some of the individuals tasked with writing the evaluation reports were experts in research/evaluation and their reports tended to be good or excellent. However, this was only the case if they were involved sufficiently early in the project to be able to collect data systematically throughout the entirety of the project, or where the grantee put in place robust evaluation systems from the outset. Few evaluation reports were the product of robust systematic approaches.
9. The diversity of content and evaluation approaches, despite the existence of the HLF guidance, shows that many grantees did not focus on what the HLF wanted from the project self-evaluation report. Our review would suggest that the evaluation of projects to HLF standards was not necessarily a key priority for many grantees who were more focused on meeting deliverables, rather than systematically collecting evidence of their impact throughout.
10. It is clear that some projects found it difficult to go beyond a discussion of outputs, and that to track even short-term outcomes was a challenge. This may be due to the inherent nature of the project / the scale of funding / activity achieved with it. It may also be due to the varying levels of expertise and understanding within each individual project team about how to measure and evaluate outcomes effectively.
11. Judging by the quality of the reports, organisations are generally aware of the need to count relevant elements in relation to the activity they have undertaken such as number of event attendees, but their understanding of the wider rationale and value of undertaking more enhanced evaluation activities is mixed.
12. The current 14 outcomes overlap in places and include a range of short and long-term outcomes and impacts. This makes it challenging for grantees to demonstrate whether the outcome has been met in their reporting. The most commonly mapped outcomes were all concerned with outcomes for people and for heritage. As before, there was evidence in over three-quarters of reports that:
 - **People will have volunteered time** (87%)
 - **Heritage will be better interpreted and explained** (79%)
 - **Heritage will be in better condition** (79%)
 - **People will have learnt about heritage** (78%)
 - **People will have developed skills** (77%)
 - **People will have had an enjoyable experience** (76%)
13. Outcomes for communities stood out as the group where there was the least amount of evidence available in the reports to support an interpretation that each outcome had been met. They are the most complex outcome area to map as they require the most consideration as well as involving complex measurement. At least in part, this is because they reach furthest outside the specific heritage project activity to its widest impact. As we found before, less than half of the reports contained evidence for any of these outcomes being met.

Recommendations: Improving the outcomes framework

1. It seems that the following outcomes for heritage are likely to be at the heart of what most projects set out to do:
 - Heritage will be in better condition
 - Heritage will be identified and recorded
 - Heritage will be better interpreted and explained
2. However, the remaining heritage outcome 'Heritage will be better managed' is wider and perhaps more subjective than the other heritage outcomes and it also overlaps to an extent with the outcome for communities - 'Your organisation will be more resilient'. We suggest that these two outcomes should be reviewed. They could be replaced with one single outcome that focuses on the positive impact that HLF funding has on an organisation.
3. There is also an overlap between the outcome for heritage 'Heritage will be better interpreted and explained' and the outcome for people 'People will have learnt about heritage'. The first outcome may lead to the second. We recommend that these could also be reviewed and replaced with one single outcome.
4. We suggest also that the outcomes for people could also be simplified and reduced in number. 'People will have volunteered time' is the most frequent outcome recorded in our review and the easiest for many grantees to measure. Connected to this outcome is 'People will have developed skills', 'People will have had an enjoyable experience' and 'People will have changed their attitudes and/or their behaviour'. HLF could review and rationalise these by considering which is the most important outcome it wants to achieve in terms of the impact that projects have on the people who get involved with them. Indeed, projects could be encouraged to focus on elements particularly pertinent to their project rather than all of the outcomes for people.
5. The outcome for communities 'More people and a wider range of people will have engaged with heritage' is a challenging outcome as it requires grantees to measure two elements over time: who and how many people were engaging before and who and how many have engaged as a result of the project. To demonstrate they have met this outcome grantees would need to conduct baseline research which provides data on their current audience and reach and then measure this again over time and at the end of the project to compare the two. The intent appears to be that HLF funding should deliver the positive impact of reaching new audiences who would not have previously engaged with that specific heritage. HLF may wish to consider simplifying this outcome so that it focuses on either more people or a wider range of people – or separating the two out to allow for more focus when outcomes are about volume, or diversity.
6. The other outcomes for community: 'Environmental impacts will be reduced', 'Your local area will be a better place to live, work or visit' and 'Your local economy will be boosted' all overlap to an extent. They are also challenging impacts for projects to measure and prove as they require measurement over the long-term and the expertise to be able to identify project impacts in isolation from other factors which may be happening in the local community at the same time. We suggest that these three outcomes could be replaced by one positive outcome for communities.

Recommendations: Improving the quality of self-evaluation reports

1. It is essential that projects are able to demonstrate good value for money. They will only be able to do this by writing good quality evaluation reports which include evidence of positive outcomes being met. As we found before, projects would benefit from more clarity and support from HLF about how to meet and measure the revised outcomes if they are to manage this successfully.
2. There are a number of different models deployed in the field of evaluation consultancy. One of the most commonly used basic tools is the idea of a 'logic model' that lays out the project inputs and activities and considers their impact. We believe that it would benefit most projects to have to lay out the logic model of their project – and that this would help with the development and undertaking of on-going measurement activities.
3. It is clear that some projects find it difficult to go beyond a discussion of outputs, and that to track even short-term impacts is a challenge. This may be due to the inherent nature of the project / the scale of funding / activity achieved with it. It may also be due to the varying levels of expertise and understanding within each individual project's team about how to measure and evaluate outcomes effectively
4. In summary, it is important to consider the scale of evaluation appropriate to the type and nature of the project. For example, larger projects in general terms are perhaps more likely to be engaged in driving towards longer term outcomes. Outcomes should of course be tracked specific to an individual project. It is unrealistic to expect projects with small budgets and low expertise to measure and track longer-term impacts. Perhaps projects could be encouraged to consider their outcomes at the various levels of activity and therefore understand more how they build on one another over time, but also depend on the focus of the project.
5. Since our 2017 review, the Heritage Lottery Fund has taken a number of actions to seek to improve the quality of evaluation and to support organisations to work with the outcomes framework. These have included:
 - Updating the evaluation guidance available to projects. The guidance now includes examples to help projects understand how to apply best practice approaches.
 - Providing training on evaluation to HLF staff. The central evaluation team at HLF has been delivering training across the HLF regions.
 - A more efficient and better developed feedback loop between HLF operational teams and the evaluation team to support programme learning.
 - Increasing the budget thresholds projects can attribute to evaluation as part of their grant application.
 - Producing more evaluation related content within HLF's online communities to encourage knowledge exchange between grantees and potential applicants.
 - A simplification of the outcomes framework based on this report and our prior report which will support the next Strategic Funding Framework.

2 Project Background and Approach

The Heritage Lottery Fund (HLF) commissioned RF Associates to undertake a comprehensive review of the self-evaluation process for 326 completed Heritage Grants projects funded under HLF's 3rd Strategic Plan (SP3) which ran between 2008 and 2013. Organisations awarded funding under SP3 were required to undertake a self-evaluation of their activity after the completion of their project. The evaluations were required to address the effectiveness of the outcomes of each project against the initial objectives (e.g. why there was a need for the project, what the applicant intended to do with the grant funds and what difference the applicant expected to make for both heritage and people).

This review includes a comparative appraisal of the quality, scope and methodology of the self-evaluated reports and the type, range and quality of activities and outcomes achieved by completed projects, including a mapping of the impact of their work onto HLF's current framework of 14 outcomes for heritage, people and communities. It combines, as far as possible, the findings from an initial review of 92 reports from an earlier evaluation, which was completed in 2015, with RF Associates' evaluation of an additional 108 reports in 2017, followed by RF Associates' most recent review of an additional 126 reports in 2017/18.

The quality of the submitted self-evaluation reports was assessed on a four-point scale (excellent, good, adequate or poor) using six criteria. These six criteria focused on the extent to which the evaluation:

1. Provided a logical framework
2. Included appropriate and methodological ways of providing robust evidence
3. Demonstrated that data was subject to robust analysis and provided evidence on outcomes
4. Was objective and free from bias
5. Presented the results clearly
6. Included sufficiently clear conclusions and recommendations to enable stakeholders to apply any lessons learned.

Further details about these criteria are provided in **Appendix 4**.

The HLF currently uses an outcome framework, which sets out 14 potential desired outcomes from its investments. The framework has 4 outcomes for heritage; 5 outcomes for people and 5 outcomes for communities. For each potential outcome, the framework provides basic examples of success to aid consideration of whether or not the outcome has been achieved. This framework was not in place at the time of application of the projects under review, therefore mapping outcomes to the framework has involved considerable interpolation. It was not always possible to map an outcome against the framework, this does not mean that the outcome was not achieved, rather that the evaluation report does not provide clear evidence that can be interpolated.

We adapted our approach for this final review of 126 reports in terms of the extent to which the outcomes had been met. Previously we had qualitatively mapped the outcomes using a basic three-point scale: yes, no and possibly. For this most recent review we removed the 'possibly' category, so that outcomes were mapped as a simple yes or no.

As before, some projects were given the ‘benefit of the doubt’ as to whether an outcome had been met based on the contents of the report, as it was not possible for us to test the evidence. For example, a project may have reported that they added interpretation panels to their heritage site but not provided evaluation evidence related to the impact of these i.e. whether these meant that the information was actually better interpreted. In this case we said that the outcome ‘heritage will be better interpreted and explained’ had been met, as we assumed that this would be the case. Our review has identified a number of issues with the interpretation of the various elements of the outcomes framework, which will need to be resolved to help clarify the requirements of future evaluations.

This report is divided into the following key sections:

- Section 3 focuses on the nature of the projects being considered within this review.
- Section 4 focuses on performance against evaluation criteria.
- Section 5 considers factors influencing evaluation scoring.
- Section 6 focuses on the outcomes framework.
- Sections 7 and 8 draw conclusions and provide recommendations.

3 Projects reviewed

3.1 Spread of project reports evaluated

The grants awarded by HLF to fund these 326 projects adds up to a total of £254,728,024. Figure 1 highlights how the project reports were distributed across a range of funding bands.

Figure 1: Percentage of evaluation reports by HLF funding bands¹



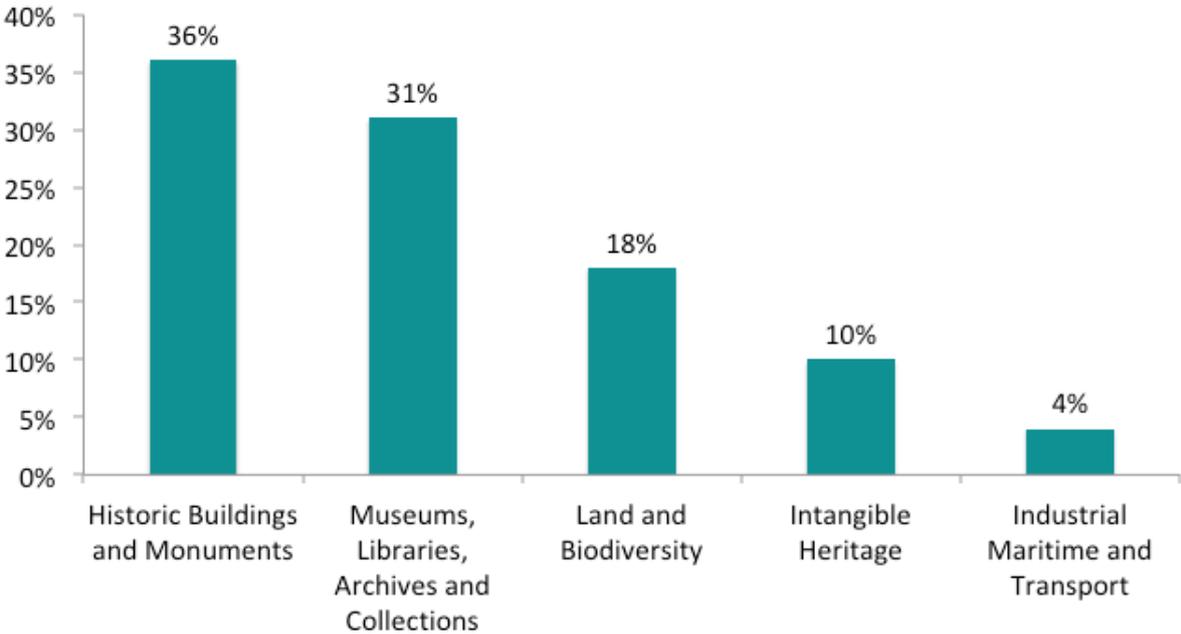
Source: HLF Heritage Programme Database Analysis (Base 326 self-evaluation reports)

The total amount allocated for evaluation on the original grant applications submitted across all 326 projects was £1,677,572. This represents less than 0.6% of the total funding awarded to these 326 projects which was £254,728,024.

HLF categorise each project into one of five sector types: Industrial Maritime and Transport, Museums and Collections, Land and Biodiversity, Historic Buildings and Intangible Heritage. Intangible Heritage includes oral and community history projects. The proportion of assessed reports within each sector type is shown in Figure 2.

¹ Throughout the report some percentages and values are rounded as appropriate, figures may therefore total 99% , 100% or 101% according to how decimals have been rounded.

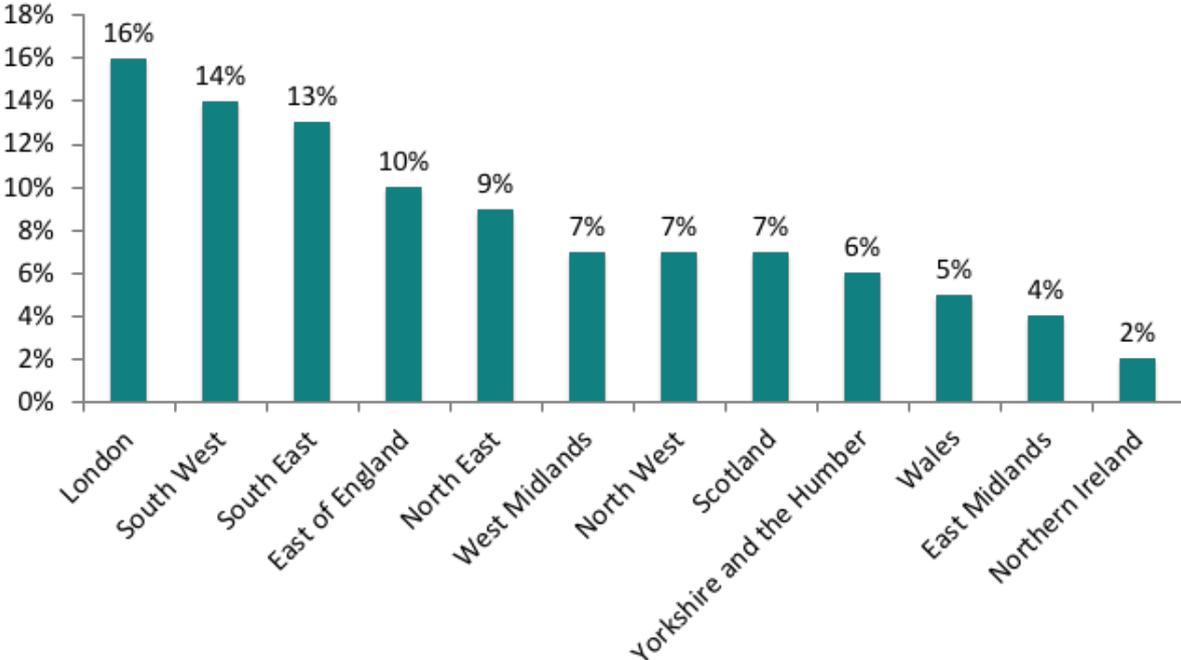
Figure 2: Percentage of evaluation reports by sector type



Source: HLF Heritage Programme Database Analysis (Base: 326 self-evaluation reports)

Figure 3 shows the regional spread of the projects included in this review.

Figure 3: Percentage of evaluation reports reviewed by region



Source: HLF Heritage Programme Database Analysis (Base: 326 self-evaluation reports)

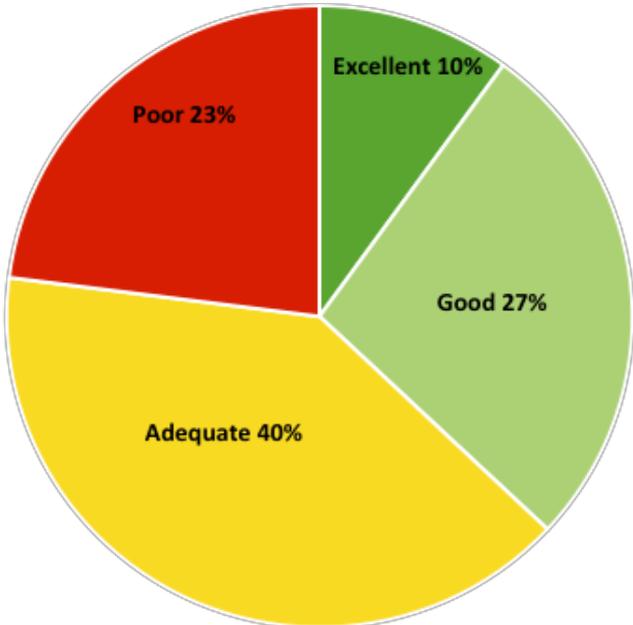
4 Performance against evaluation criteria

This review has compared the overall report scores for the first 200 reports evaluated with the next 126 reports evaluated and then the total 326 reports evaluated.

The scores are remarkably consistent. The review of the first 200 reports found that 12% of reports were excellent, 25% were good, 39% were adequate and 25% were poor.

Figure 4 shows the quality of all 326 reports evaluated. Consistent with the last report’s findings, just over a third, 37% of reports were graded as good or excellent, with just under two thirds, 63%, graded as adequate or poor.

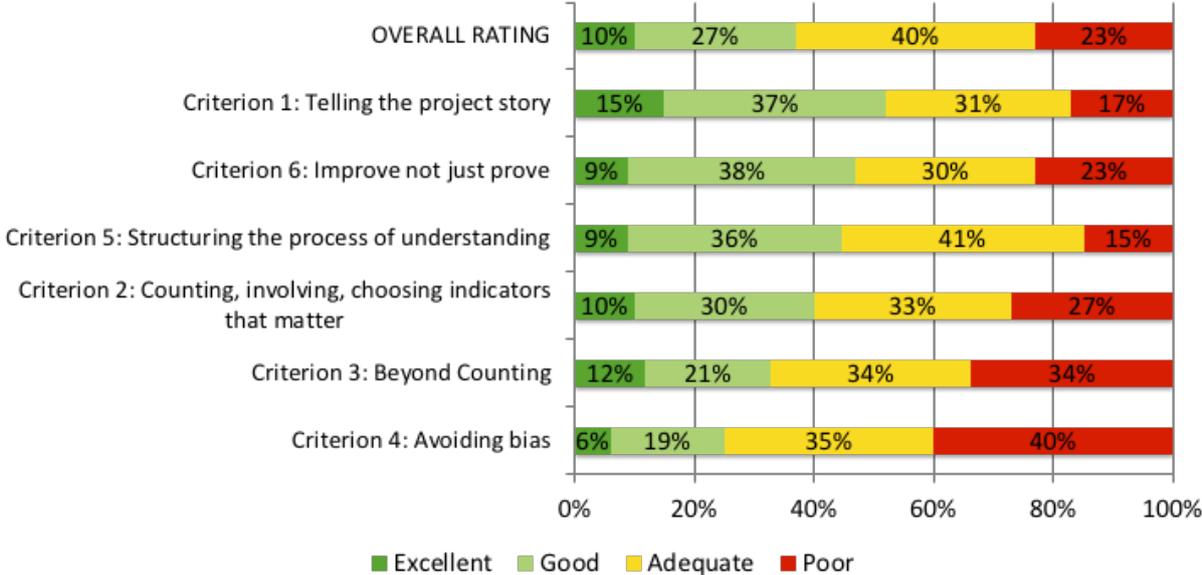
Figure 4: Quality of self-evaluation reports



Source: HLF Heritage Programme Database Analysis (Base 326 self-evaluation reports)

Figure 5 ranks the six evaluation criteria in descending order in terms of the proportion achieving excellent/good scores. There has been little movement between the review of the first 200 and the next 126. The ranking of the criteria for all 326 reports is in the same order as before.

Figure 5: Criteria in rank order of quality score



Source: HLF Heritage Programme Database Analysis (Base 326 self-evaluation reports)

As we found before, the criteria with the highest proportion of excellent/good scores, were **Criterion 1: Telling the project story – The evaluation provides a logical framework setting out linkages between activities, expected outputs and outcomes for all elements of the project** with 52% and **Criterion 6: Improve not just prove - the conclusions and recommendations are sufficiently clear to enable stakeholders to identify and apply any lessons learned**, with 47%.

As before, report authors were strongest at telling the project story and drawing out conclusions and recommendations than in any other aspects of evaluation. These areas require the least technical understanding when undertaking the evaluation process.

The criteria with the lowest proportion of excellent/good scores, were the same as previously: **Criterion 3 Beyond counting – data was subject to robust analysis to provide evidence on outcomes** and **Criterion 4 Avoiding bias – the evaluation is objective and free from bias**. Fewer report authors were able to analyse the data robustly and to provide evidence of objectivity. This indicates that grantees need help with these aspects particularly, if the quality of self-evaluation reports is to improve in future.

Detailed commentary is available for each of the criteria in **Appendix 1**.

5 Factors which influenced the quality of reports

5.1 Overview

This review of 326 reports confirms that the criteria scores show significant dependence on several characteristics, confirmed by a range of statistical tests.

- **Evaluation was external:** external consultants/organisations tended to write better quality reports.
- **Planned share of grant initially allocated for evaluation:** where higher amounts of expenditure had been originally allocated for evaluation, the reports tended to be of better quality.
- **Number of outcomes recorded:** the higher quality reports were more likely to have more HLF outcomes ascribed to them.
- **Report length:** longer reports which contained more explanation and more data were of better quality than the shorter reports.

Regression analysis indicates that:

- The use of an external evaluator has a large positive effect on the probability of achieving a “good” score and a smaller but still significant positive effect on achieving an “excellent” score.
- The number of outcomes achieved has a similar sized but more statistically significant relationship with achieving a “good” score and again a smaller but significant link with achieving an “excellent” score.
- The planned share of grant for evaluation has a significant but smaller effect on the probabilities of “good” or “excellent” scores.

These factors do not guarantee better criteria scores but promote their likelihood.

We have analysed a range of other factors which might have been expected to have some impact on the quality of reports but none of these show any statistical dependence:

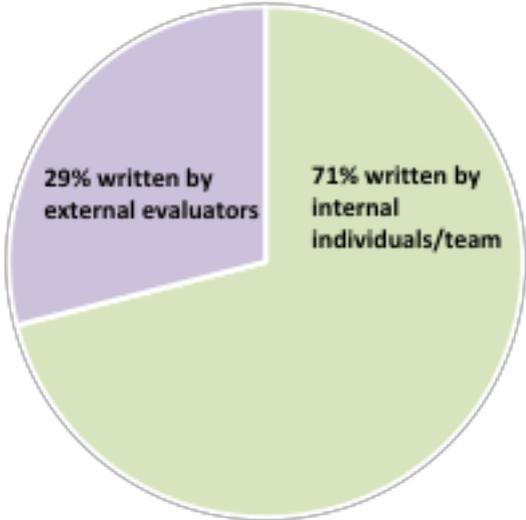
- Heritage area
- Region
- Financial size of grant awarded by HLF for the project
- Type of organisation awarded the grant
- Whether the assessment was from the 2017 or 2018 reviews conducted by RF Associates

The data for this is included in **Appendix 2**.

5.2 Evaluation was external

External evaluators were employed in 93 of the 326 evaluations reviewed. There are 8 reports where it is unclear whether they were written internally or externally. Figure 6 shows the proportion of these as a percentage of the whole; the base size for this calculation is 318.

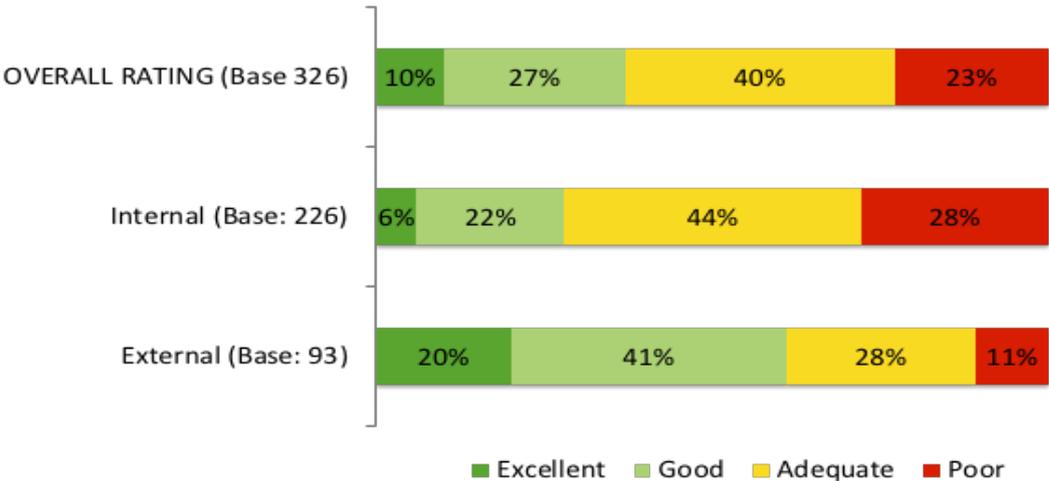
Figure 6 Proportion of reports written by external evaluators or written internally



Source: HLF Heritage Programme Database Analysis (Base 318 self-evaluation reports)

On average, reports which were written by external evaluators show a strongly and statistically significant higher quality than those undertaken entirely internally. Figure 7 shows the cross tabulation.

Figure 7: Overall score by type of evaluator used



Source: HLF Heritage Programme Database Analysis (Base 318 - 326 self-evaluation reports)

The findings in Figure 7 are consistent with the previous review. Just under three quarters of all

326 reports evaluated were written internally by a project staff member or members. Of these, just under three quarters, 72%, scored adequate/poor. Only a very few, 6%, scored excellent.

The base size of reports written by external evaluators has now increased. In the previous review, 49 reports were written by external evaluators. Now the base size is 93.

These numbers are more robust for analysis purposes, and there has been some change, with a smaller proportion of externally evaluated reports scoring excellent/good than in the last review. In the first 200 report review, 72% of the small base size of 49 reports scored excellent/good. In this review of 326 reports in total, 61% scored excellent/good. There has been an increase of reports scoring adequate, from 18% in the first 200 review, to 28% in total for all 326 reports. However, the proportion scoring poor has remained consistent, at 10% (previously 11%).

Despite this adjustment, overall, the contrast in quality between the internally written reports and those written by external evaluators shows that the use of external evaluators has a positive impact on the quality of the evaluation. This is likely to be due to the expertise that external consultants bring in collecting, analysing and reporting on data. Again, this finding needs to be considered with the following caveat: it was not possible to know what internal expertise and resource was available within each project team to help provide further support to the external evaluator and to help with the planning and implementation of the ongoing evaluation during the lifetime of the project. This finding should therefore be taken as indicative, rather than a definitive indicator of quality.

The conclusion bears out the reasonable assumption that the involvement of external expertise in project evaluation can often add value. This result does not imply that external sourcing is essential for good evaluation but suggests that it is an option to be carefully considered in each case.

As in the previous review, it is helpful to understand why the 11% of poor reports, which were written by external evaluators, scored poor. There are a range of reasons for this:

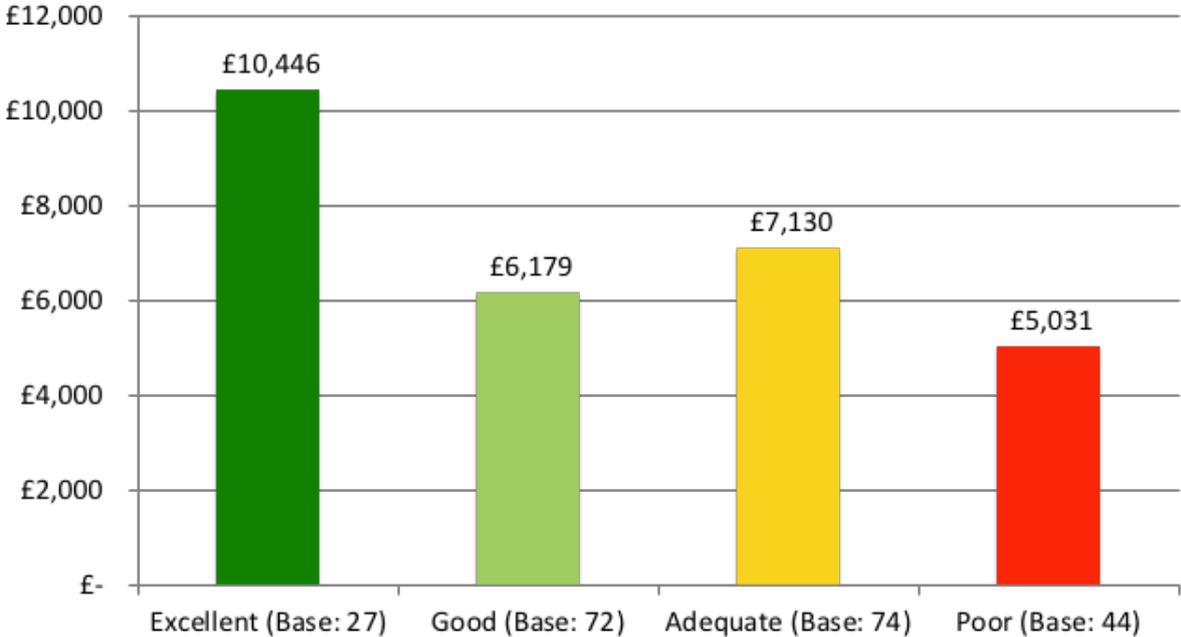
- The report was not an actual complete project evaluation report – it was research about one aspect of the project (e.g. education programme, volunteer programme) which was then submitted to HLF as a final evaluation report.
- The report was written as a short summary final review for the organisation itself and this organisation then submitted it to HLF in place of a detailed evaluation report.
- The report was written by the construction contractors rather than a research specialist, and thus did not act as a full evaluation report for the project.
- Insufficient data had been collected during the course of the project and the external consultants appeared to have been contracted to summarise the project rather than conduct any primary research into achievements and outcomes.

5.3 Planned share of grant initially allocated to evaluation

217 projects out of the 326 reviewed, 67%, had allocated specific amounts of funding for evaluation in their original bid documents. The proportion allocated varied from less than 0.1% to over 6% of the total grant award to the project. The amount allocated varied from £60 to £266,200.

Figure 8 shows the average amount allocated for evaluation for the projects whose evaluations were assessed as falling within the four quality grades. The average amount allocated for evaluation of reports assessed excellent was more than £10,000. In contrast, the average amount allocated for reports assessed as poor was just over £5,000.

Figure 8: Average amount originally allocated for evaluation by score of final report

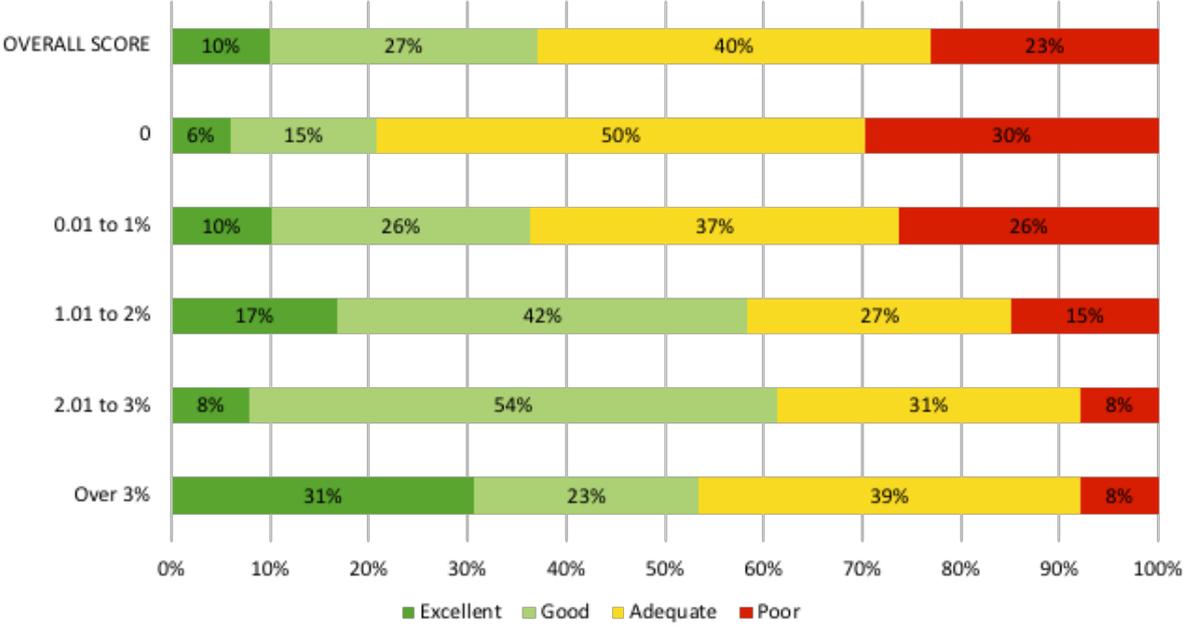


Source: HLF Heritage Programme Database Analysis (Base 217 self-evaluation reports)

It is plausible to suppose that planning resource for evaluation can increase the probability of an effective analysis. This is borne out by the statistical analysis - there is a tendency for plans for more evaluation expenditure to be associated with higher average criteria assessment scores, and statistical tests indicate that this is statistically significant.

Figure 9 shows how the overall quality of reports increases in line with increases to the proportion of grant funding originally allocated for evaluation.

Figure 9: Overall score for amount allocated to evaluation as a proportion of grant awarded

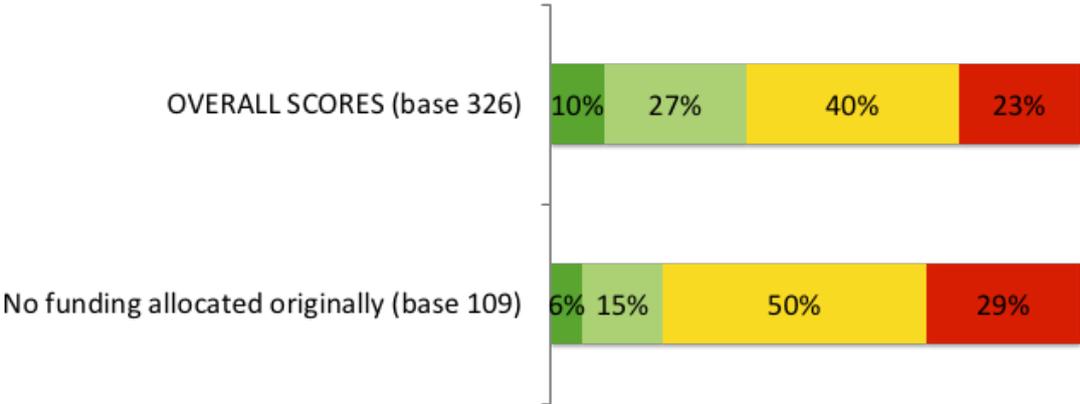


Source: HLF Heritage Programme Database Analysis (Base 217 - 326 self-evaluation reports)

However, this finding needs to be considered with the following caveat: the amount allocated for evaluation was identified in data submitted at the original bid stage. It was not possible to know whether this amount was actually spent on evaluation as had been intended; it may be the case that a different amount was spent on evaluation.

109 reports were submitted by projects which had not allocated any funding for self-evaluation in their original documents. Figure 10 shows that almost four-fifths of these, 79%, scored adequate or poor and just over a fifth, 21%, scored good or excellent. These findings are consistent with the last review, where the proportions of the 72 reports were within +/-2 of the percentages in this latest review of 326 reports.

Figure 10: Score for reports with no funding originally allocated for evaluation



Source: HLF Heritage Programme Database Analysis (Base 109 – 326 self-evaluation reports)

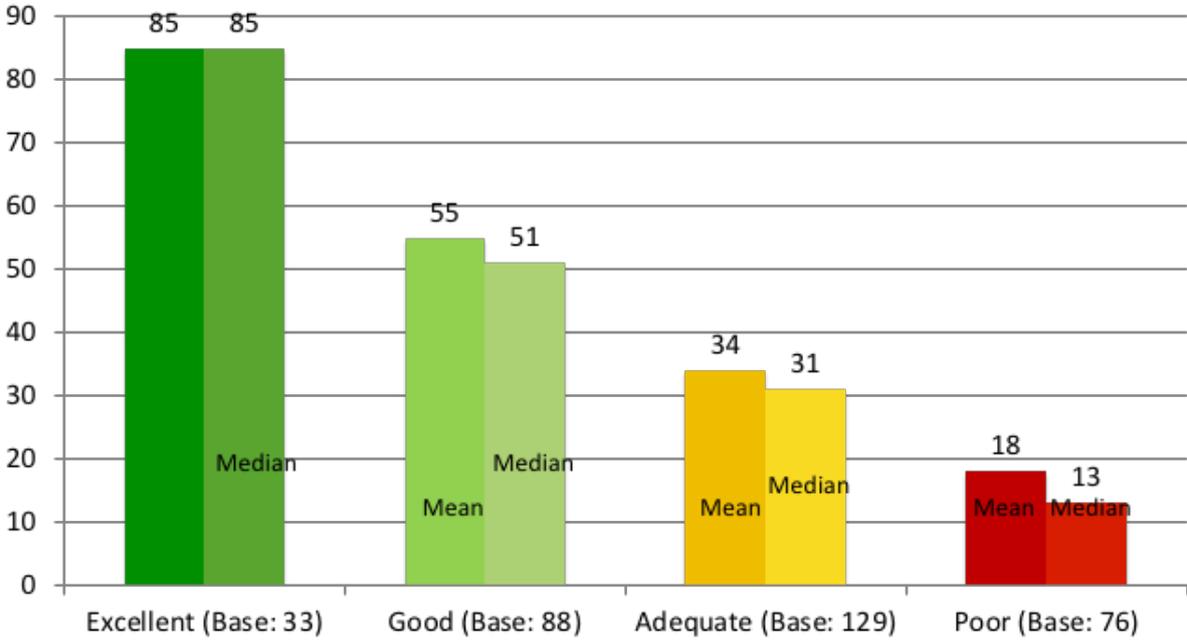
5.4 Report length

The length of reports varied widely. The shortest report was less than one side of A4, whilst the longest was 306 pages of dense text and appendices.

It is clear that the length of report is not the only criterion of effort or ability to evaluate appropriately and robustly. Some projects were smaller in scope than others, with fewer expected outputs and outcomes to evaluate and report on. However, short reports were unable to do justice to their projects. Short reports were insufficient to provide appropriate evaluative evidence of the project, considering the range of aims and objectives to be reported upon and the amount of public money for which the projects were accountable.

Figure 11 shows that the better quality excellent/good reports had a higher mean and median number of pages than the weaker adequate/poor reports.

Figure 11: Relationship between mean/median length of the evaluation report (number of pages) and overall score



Source: HLF Heritage Programme Database Analysis (Base 326 self-evaluation reports)

5.5 HLF outcomes mapped on to the 326 reports

5.5.1. Overview

The reviews of the 326 self-evaluation reports have included an assessment of which of the 14 outcomes set out in the HLF Outcomes Framework can be attributed to each project.

Mapping the HLF 14 outcomes onto the 326 self-evaluation reports has been a retrospective, qualitative exercise – as the outcomes framework was agreed following the awarding of these grants.

As we found in the previous review and mapping exercise, the degree of evidence around outcomes in the reports varied enormously. In some reports it was possible to find robust data that supported an outcome being met, for example a volunteer survey which demonstrated that people had learnt new skills. In other reports there was no objective data available, but the report included a description which could be interpreted as the outcome being met. We have been generous to the projects in our mapping, assuming that where the report included some mention of an output which might have led to an outcome being met, that we should interpret it as a ‘yes’. Where there was no mention of any activity or description of an outcome being met, we mapped that as a ‘no’.

Comparing the two review cohorts - those completed in the 2017 and 2018 reviews, has found very similar distributions of values, identical averages and very similar standard deviations. Formal tests are not therefore required to know that there is no statistically significant difference between the two cohorts - assessments have been carried out consistently.

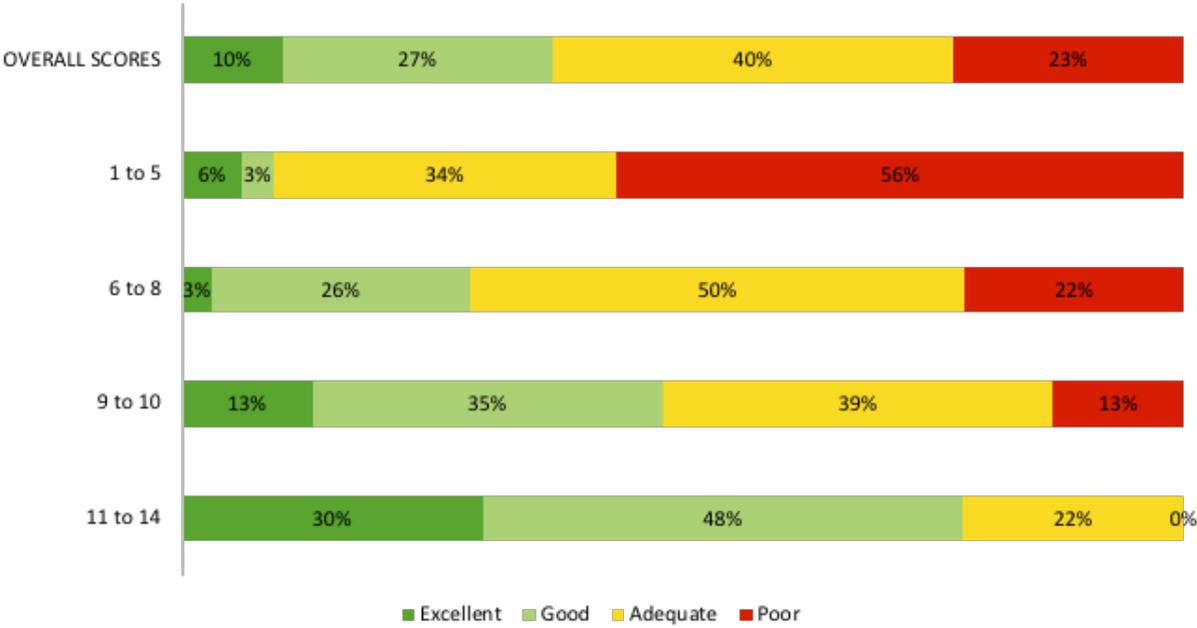
5.5.2 Correlation between number of outcomes met and quality of reports

We have tested whether the overall assessment scores, based on the 6 criteria are in some way correlated with the number of instances of achievement of an outcome. The possible link has been investigated in two ways:

- Counting the number of outcomes achieved from the 14 specified in the Outcomes Framework.
- Counting the achievement of each outcome category - Heritage, People, Communities, scored if over 50% of the outcomes in each category were attributed to a project.

Figure 12 shows the relationship between criteria scores and the sum of all outcomes achieved. Statistical tests confirm that there is a strong link between the number of outcomes achieved and the scores on the quality criteria. The likely reason for this is that the assessors only have the self-evaluation reports to go on; the better written and more evidence-based the report (the higher it scores on the quality criteria) the more readily assessors can discern if an outcome has been achieved. In effect, the outcome scores are an indicator of report quality as well as of project achievements. It would be wrong, however, to interpret the statistical link as indicating a causal connection between the number of outcomes and the self-evaluation quality criteria.

Figure 12: Criteria scores by number of outcomes met



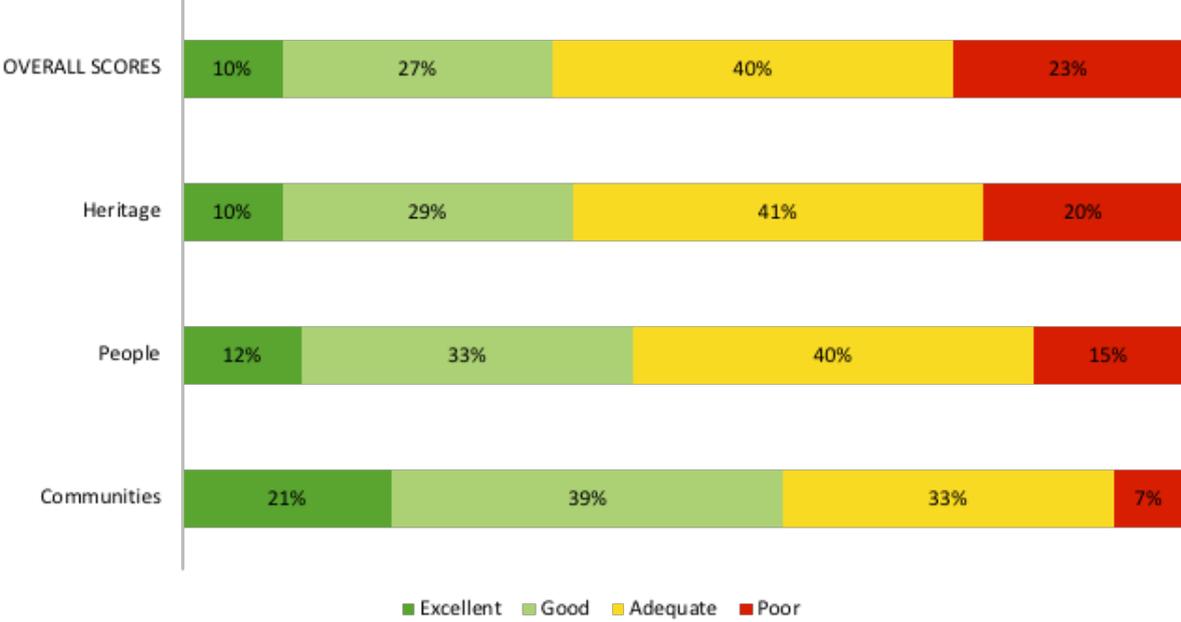
Source: HLF Heritage Programme Database Analysis (Base 326 self-evaluation reports)

5.5.3 Correlation between categories of outcomes met and quality of reports

The links between criteria scores and outcomes can also be shown for the three categories of outcome. Figure 13 shows criteria scores against the three outcome categories. We devised a system for measuring this; reports were scored as being in an outcome category if they had achieved at least half the outcomes for that category.

Reports that score on ‘people’ and, rather more so, on ‘communities’ were more likely to achieve good or excellent criteria scores. This tends to bear out the interpretation that clearly presented and well evidenced self-evaluation reports were likely to show more outcomes as the material is there to adduce their existence.

Figure 13: Criteria scores by outcome categories



Source: HLF Heritage Programme Database Analysis (Base 326 self-evaluation reports)

We repeated the regression analysis with the criteria scores expressed on a two-point scale – ‘poor or adequate’ against ‘good or excellent’ and using the three categories of outcomes (heritage, people and communities) as separate indicators. The results confirm the findings from the previous regression model, that the use of external evaluation, planned evaluation resources and achievement of two of the three categories of outcome had a substantial and statistically significant relationship with achieving a ‘good or excellent’ score. Amongst the outcome categories, people related outcomes showed the largest and most significant correlation with criteria scores, that is to say the higher quality evaluation reports were more likely to demonstrate people outcomes. The link with heritage (the most frequent outcome category) is though not statistically significant.

5.5.4 Investigating links between mapped outcomes and other project characteristics: grant size, heritage area, region

We have applied similar analytical methods to the number and types of outcome mapped by the assessors, to test if these are linked with other project characteristics.

The possibility that the number of outcomes scored was larger for bigger budget projects was tested by computing the simple statistical correlation coefficients between the number of outcomes achieved and the level of project grant. The computed correlation indicates an extremely low degree of association – the two are essentially independent.



A cross tabulation analysis and associated statistical tests also indicate a low degree of association between numbers of outcomes and heritage areas.

Similarly, there is no statistically significant pattern in the number of outcomes achieved by projects in each region.

6 Understanding the incidence of mapped outcomes

Figure 14 ranks the 14 outcomes in descending order of the extent to which the outcomes have been mapped as being met.

Figure 14: Mapping outcomes onto the 326 self-evaluation reports



Source: HLF Heritage Programme Database Analysis (Base 326 self-evaluation reports)

The following section considers each of these outcomes within their three categories:

- **Outcomes for heritage**
 - Heritage will be better managed
 - Heritage will be in better condition
 - Heritage will be better interpreted and explained
 - Heritage will be identified and recorded
- **Outcomes for people**
 - People will have developed skills
 - People will have learnt about heritage
 - People will have changed their attitudes and/or behavior
 - People will have had an enjoyable experience, people will have volunteered time
- **Outcomes for communities**
 - Environmental impacts will be reduced
 - More people and a wider range of people will have engaged with heritage
 - Your local area/community will be a better place to live, work or visit
 - Your local economy will be boosted

6.1 Outcomes for heritage

Heritage will be better interpreted and explained

- Almost four fifths (79%) of reports had some evidence that this outcome had been met.
- Some provided explicit evidence that this had occurred e.g. surveys that asked participants how highly they rated the information provided.
- However, in many cases we have assumed that the activities which were designed to better interpret and explain heritage described in the report, (for example, the production of school resource packs, the development of guide books and new display information, the creation of new interpretation boards) are an improvement on what existed before the project began.
- No reports provided baseline quantitative information which measured how well heritage had been interpreted and explained at the outset of the project (where this was relevant as it already existed, such as in a museum), and then compared it to data collected at the end of the project once the interpretation materials had been improved.

Heritage will be in better condition

- Almost four fifths (79%) of reports had some evidence that this outcome had been met.
- In many reports this was the main activity and focus of the project, e.g. heritage site restored or repaired, or an enhanced facility created to conserve heritage such as an archive centre, or a website or digitization of archive materials created, or where a new nature reserve was developed to conserve environmental heritage.

Heritage will be identified/recorded

- Just under three fifths (59%) of reports had some evidence that this outcome had been met.
- All the oral history and archive and archaeology projects met this outcome.
- Some reports contained baseline information which supported the evidence showing the increase in heritage which had been identified/recorded.
- However, few reports met the HLF criterion of demonstrating that visitors or users had told the project that this was an important part of heritage and that they valued it.

Heritage will be better managed

- This is a less frequently evident outcome for heritage, and therefore only possible to map in just under half of all reports evaluated (48%).
- This outcome overlapped, to some degree, with the community outcome '**your organisation will be more resilient**'.
- Reports containing some degree of evidence about this outcome mentioned either improvements to staffing resources and structures, or improved ways of managing resources and relationships more effectively to ensure that heritage would be better managed in future.
- Examples included: the establishment of new project management structures; improved partnership working with other historic venues; appointment of new staff; building capacity to better manage and retain volunteers.
- None of the evidence about this outcome was robust in the sense that it was measured and quantified.

6.2 Outcomes for people

People will have volunteered time

- This is the most commonly mapped outcome across all categories, with 87% of reports containing evidence of this.
- Of the 13% of reports that did not demonstrate this, 88% were adequate/poor.
- Reports varied widely in the degree of evidence for this. A few provided detailed breakdowns of volunteer hours and their financial equivalent, and some included quantitative surveys of volunteers or qualitative verbatim and case studies to support their evaluation of volunteers.

People will have learnt about heritage

- A similarly high proportion of reports, 78%, contained evidence of this.
- The interpretation of this was also fairly broad.
- In many reports there was evidence that learning had occurred from the volunteers who were involved in the project, rather than a wider group of people such as visitors or participants.
- In other reports we have assumed that this occurred as a result of the fact that the project involved educational outreach and learning events.
- Few reports contained robust evidence that this outcome had been met which is outlined by HLF as people being 'able to tell you what they have learnt about heritage and what difference this makes to them and their lives. They will also be able to tell you what they are doing with that knowledge and understanding; such as, sharing it with other people, using it in their professional or social life, or undertaking further study.'

People will have developed skills

- Over three-quarters of reports (77%) contained evidence that this outcome had been met.
- This outcome was closely correlated with 'people will have volunteered time', where the skills referred to were developed by the project staff and volunteers.
- Skills development was wide ranging, including many different types such as learning how to conduct oral history interviews, species identification, conservation, object handling and volunteer management skills.
- Evidence for skills development included staff and volunteers receiving training as well as qualitative verbatim.

People will have had an enjoyable experience

- Over three-quarters of reports (76%) contained evidence that this outcome had been met.
- This is an outcome that to some degree we have assumed has been achieved from other data gathered around activities, without a specific data source that said people had enjoyed themselves.
- In many cases the evidence for this was obtained from a range of qualitative verbatim gathered from visitor books, feedback cards, forms, emails and Trip Advisor reviews. 'Enjoyment' as a concept tends to be absent on visitor surveys, with respondents being asked to express their level of 'satisfaction' instead, although there were some examples of enjoyment being measured to varying degrees.

People will have changed their attitudes and / or behaviour

- Significantly fewer reports (33%) had evidence which could be mapped to this outcome in comparison with the other outcomes for people.
- This outcome is challenging to evidence as it requires the collection of before and after data. It is difficult to capture this quantitatively and therefore most projects did not attempt to do so.
- However, it should be possible for most projects to demonstrate some impact on attitude or behaviour even if they evidence was just gathered from volunteers about the impact on them of participating in the project.
- Many evaluation reports did not even provide evidence that this had occurred; on occasion they provided numbers of volunteers or participants/visitors, but with no commentary, verbatim or measurement to demonstrate changes in attitudes and behaviour.
- A few excellent/good reports did discuss this. Qualitative evidence supported by verbatim was the most common evidence; a few reports included quantitative measurement of how the project had made individuals think differently about heritage or their community.

6.3 Outcomes for communities

As we found before, outcomes for communities stand out as the group where there was the least amount of evidence available in the reports to support an interpretation that each outcome had been met. They are the most complex to map as they require the most consideration as well as involving complex measurement. At least in part, this is because they reach furthest outside specific heritage project activity to its wider impact. Less than half of the reports contained evidence for any of these outcomes being met.

Your organisation will be more resilient

- Just under half of reports (47%) had evidence that this outcome had been met.
- We have widely interpreted the outcome as where there was evidence (often assertions) that the organisation was stronger to some degree (i.e. it did not exist before / it now has new facilities / resources etc.) as a result of the changes made by the project.
- There was not always enough information provided to assess if an organisation was more resilient than it was before, which was why just over half of the reports were not considered to have met this outcome.

More people and a wider range of people will have engaged with heritage

- Just over two-fifths of reports (42%) had evidence that this outcome had been met.
- This is a challenging outcome to evidence because it measures two aspects: more people and a wider range of people; and ideally requires baselines to compare against.
- Many evaluation reports provided strong evidence of increased numbers but not from a wider range of people.
- Most projects did not seem to have collected baseline data about audience profile at the beginning of their activities that they could then compare with the audience profile towards the end of the project.
- For some projects this was because it did not exist before the grant, which made collecting data problematic.
- However, even a large proportion of museums did not seem confident in, or concerned

with, collecting robust data about their audience profile. Some reports acknowledged that they tried to / were expected to, reach new audiences but did not succeed in doing so.

- A few good examples of this outcome being met are from some of the projects which primarily focused on recruiting large numbers of volunteers (to record heritage, to learn about heritage). Particularly when the organisations themselves appeared to be experienced at creating research tools to collect this data and to analyse it (e.g. University of Oxford, Workers Educational Association). These grantees had clearly planned this data collection as integral to the project.
- It appears that few projects have collected evidence that proves they have met the inclusivity targets/ objectives, which were in some cases the HLF's approved aims for the grant.
- For many projects, the report's focus tends to be on what the project is achieving i.e. the build/redevelopment, along with the associated planned activities and therefore the grantees consider that success is demonstrated by having achieved this, not by collecting evidence that this will lead, or has led to, more diverse audiences.
- There is some evidence of squeamishness with collecting 'personal' data from audiences (i.e. their ethnicity, income etc), which suggests that some projects and the volunteers or staff members who are tasked with collecting data, prefer not to potentially antagonise their visitors/participants by asking for this level of detail.

Environmental impacts will be reduced

- A quarter of reports contained some evidence that this outcome had been met.
- This has been interpreted so that where reports include positive environmental impacts these have been assessed as meeting this outcome.
- We have taken a broader interpretation than the current HLF definition suggests; very few of the evaluated reports have included any information about reduction in carbon emissions.
- Most of the environmental conservation project reports evaluated were able to demonstrate some sort of positive environmental impacts ranging from enhanced biodiversity to protection and enhancement of the landscape/site.
- A few other projects also mentioned positive environmental impacts in terms of implementing environmentally sensitive measures within a capital build project.
- A few projects also mentioned that the heritage which was identified and recorded by the project (e.g. archaeological sites, heritage at risk) will have an indirect positive environmental impact by helping to protect this heritage in future from deterioration or loss.

Your local area will be a better place to live, work or visit

- Just under a quarter (24%) of reports had some evidence of this outcome being met.
- Whilst it is possible that many projects could have created this type of positive impact, evidence for this was less widespread than for many of the other outcomes.
- Examples included: evidence of community group involvement creating community cohesion and social capital; verbatim and letters or emails supporting the view that the project had helped to regenerate an area or restore an important building and/or created a vital community resource and meeting place; evaluation evidence on enhanced attractiveness of local area; establishment of new local partnerships and ongoing collaboration.

Your local economy will be boosted

- This outcome was the least commonly mapped onto the reports; we could map this outcome on to just 15% of the reports.
- Again, we have interpreted this more widely than the description provided by the HLF. A few reports described that the impact of the project had helped create additional income for local businesses.
- Most of the evidence for this outcome was about the creation of work in the area, whether this was new jobs entirely or local businesses benefitting from working on the project.
- A few reports also contained estimations and analyses of the increased visitor numbers and expenditure which would benefit the local economy.

7 Conclusions

1. The reviews confirm that the quality of the self-evaluation reports submitted varies widely.
2. The main indicators of quality are derived from how seriously the organisation invested in the evaluation process in terms of:
 - Funding – where higher amounts of expenditure were initially allocated for evaluation, the reports tended to be of better quality.
 - Expertise – external consultants/organisations tended to write better quality reports.
 - Length / time – longer reports, which contained more explanation and more data tended to be of better quality than the shorter reports.
3. There is also a link between better quality self-evaluation reports and a greater number of outcomes being mapped as met.
4. Reports that mapped more ‘people’ outcomes and, rather more so, mapped ‘communities’ outcomes were more likely to achieve good or excellent evaluation scores. This confirms our interpretation that clearly presented and well-evidenced self-evaluation reports are likely to show more outcomes as the material is there to adduce their existence.
5. It was not possible to discern any other indicators of quality connected with heritage type, type of organisation, size of grant awarded and region. Therefore, the conclusions drawn apply across the portfolio of all types of projects.
6. As we found before, the diversity of content and evaluation approaches, despite the existence of the HLF guidance, shows that many grantees were not clear about what the HLF wanted from the self-evaluation report.
7. Our reviews suggest that the self-evaluation of projects to HLF standards was not a key priority for many grantees who were more focused on meeting deliverables, rather than systematically collecting evidence of their impact throughout.
8. Some of the individuals tasked with evaluating projects were experts in research and their reports tended to be good or excellent. However, this was only the case if they were involved sufficiently early in the project to be able to collect data systematically throughout the process, or where the grantee put in place robust evaluation systems from the outset. Few evaluation reports were the product of robust systems.
9. Following the recommendations from our 2017 review, the Heritage Lottery Fund has revised its guidance for grantees. Our latest 2018 review of 326 self-evaluation reports has confirmed the relevance of these recommendations and built upon these in the following section.

8 Recommendations

8.1 Recommendations: Improving the outcomes framework

1. It seems that the following outcomes for heritage are likely to be at the heart of what most projects set out to do:
 - Heritage will be in better condition
 - Heritage will be identified and recorded
 - Heritage will be better interpreted and explained
2. However, the remaining heritage outcome 'Heritage will be better managed' is wider and perhaps more subjective than the other heritage outcomes and it also overlaps to an extent with the outcome for communities - 'Your organisation will be more resilient'. We suggest that these two outcomes should be reviewed. They could be replaced with one single outcome that focuses on the positive impact that HLF funding has on an organisation.
3. There is also an overlap between the outcome for heritage 'Heritage will be better interpreted and explained' and the outcome for people 'People will have learnt about heritage'. The first outcome may lead to the second. We recommend that these could also be reviewed and replaced with one single outcome.
4. We suggest also that the outcomes for people could also be simplified and reduced in number. 'People will have volunteered time' is the most frequent outcome recorded in our review and the easiest for many grantees to measure. Connected to this outcome is 'People will have developed skills', 'People will have had an enjoyable experience' and 'People will have changed their attitudes and/or their behaviour'. HLF could review and rationalise these by considering which is the most important outcome it wants to achieve in terms of the impact that projects have on the people who get involved with them. Indeed, projects could be encouraged to focus on elements particularly pertinent to their project rather than all of the outcomes for people.
5. The outcome for communities 'More people and a wider range of people will have engaged with heritage' is a challenging outcome as it requires grantees to measure two elements over time: who and how many people were engaging before and who and how many have engaged as a result of the project. To demonstrate they have met this outcome grantees would need to conduct baseline research which provides data on their current audience and reach and then measure this again over time and at the end of the project to compare the two. The intent appears to be that HLF funding should deliver the positive impact of reaching new audiences who would not have previously engaged with that specific heritage. HLF may wish to consider simplifying this outcome so that it focuses on either more people or a wider range of people – or separating the two out to allow for more focus when outcomes are about volume, or diversity.
6. The other outcomes for community: 'Environmental impacts will be reduced', 'Your local area will be a better place to live, work or visit' and 'Your local economy will be boosted' all overlap to an extent. They are also challenging impacts for projects to measure and prove as they require measurement over the long-term and the expertise to be able to identify project impacts in isolation from other factors which may be happening in the local community at the same time. We suggest that these three outcomes could be replaced by one positive outcome for communities.

8.2 Recommendations: Improving the quality of self-evaluation reports

1. It is essential that projects are able to demonstrate good value for money. They will only be able to do this by writing good quality evaluation reports which include evidence of positive outcomes being met. As we found before, projects would benefit from more clarity and support from HLF about how to meet and measure the revised outcomes if they are to manage this successfully.
2. There are a number of different models deployed in the field of evaluation consultancy. One of the most commonly used basic tools is the idea of a 'logic model' that lays out the project inputs and activities and considers their impact. We believe that it would benefit most projects to have to lay out the logic model of their project – and that this would help with the development and undertaking of on-going measurement activities.
3. It is clear that some projects find it difficult to go beyond a discussion of outputs, and that to track even short-term impacts is a challenge. This may be due to the inherent nature of the project / the scale of funding / activity achieved with it. It may also be due to the varying levels of expertise and understanding within each individual project's team about how to measure and evaluate outcomes effectively.
4. In summary, it is important to consider the scale of evaluation appropriate to the type and nature of the project. For example, larger projects in general terms are perhaps more likely to be engaged in driving towards longer term outcomes. Outcomes should of course be tracked specific to an individual project. It is unrealistic to expect projects with small budgets and low expertise to measure and track longer-term impacts. Perhaps projects could be encouraged to consider their outcomes at the various levels of activity and therefore understand more how they build on one another over time, but also depend on the focus of the project.
5. Since our 2017 review, the Heritage Lottery Fund has taken a number of actions to seek to improve the quality of evaluation and to support organisations to work with the outcomes framework. These have included:
 - Updating the evaluation guidance available to projects. The guidance now includes examples to help projects understand how to apply best practice approaches.
 - Providing training on evaluation to HLF staff. The central evaluation team at HLF has been delivering training across the HLF regions.
 - A more efficient and better developed feedback loop between HLF operational teams and the evaluation team to support programme learning.
 - Increasing the budget thresholds projects can attribute to evaluation as part of their grant application.
 - Producing more evaluation related content within HLF's online communities to encourage knowledge exchange between grantees and potential applicants.
 - A simplification of the outcomes framework based on this report and our prior report which will support the next Strategic Funding Framework.

Disclaimer

This report was commissioned by the Heritage Lottery Fund and written by RF Associates. Our conclusions are the result of our professional judgment, based upon the material and information provided to us by the client.

About RF Associates

RF Associates is a research and evaluation business focused on delivering high value research driven insight. We were established in May 2015 and are committed to offering great work whilst growing and developing our team. We undertake the whole breadth of research and evaluation methodologies including desk research, both qualitative and quantitative research projects, and make use of a wide range of consultancy methods. We work with clients across the UK in both the private and public sector.

APPENDIX CONTENTS

APPENDIX 1: Performance against evaluation criteria

APPENDIX 2: Analysis of heritage area, region and financial size of project

APPENDIX 3: Overall summary of self-evaluation report strengths and weaknesses

APPENDIX 4: HLF adopted approach to reviewing grantee evaluations

APPENDIX 1: Performance against evaluation criteria

1. Criterion 1: Telling the project story

The evaluation provides a logical framework setting out linkages between activities, expected outputs and outcomes for all elements of the project.

'Telling the project story' had the highest proportion of reports that scored good/excellent, with over half of all reports, 52%, achieving this score. Compared to the other criteria it also had a low proportion of reports scoring poor, 17%.

For 234 reports, we are able to analyse how the reports scored for each aspect of this criterion².

75% included an explanation of how each activity contributed to the outputs achieved

- Three quarters of these reports linked activities with outputs.
- However, a quarter did not manage to explain this sufficiently clearly or in any detail.

59% included a clear explanation of how each output led to outcomes or wider impacts

- Over two fifths did not include a clear explanation about the outcomes which occurred as a result of the project.
- Many of these reports did not include any information about outcomes or wider impacts, while others confused 'outcomes' with 'outputs'.
- The best reports included extensive discussion of the outcomes that had been achieved and evidence to support these claims.
- Reports which distinguished clearly between outputs and outcomes often had collected considerable amounts of qualitative evidence on outcome and impact from a range of perspectives such as stakeholders, participants, volunteers and the community.

27% included some sensitivity analysis or other estimation of the extent of attribution of activities to outputs or outputs to outcomes

- Sensitivity analysis is the estimation of the extent of the effect of other activities on outputs, or outputs to outcomes.
- This was frequently the sub element of the criteria that prevented the 37% of all 326 reports that scored good in this criterion, from scoring excellent.
- In some cases, the sensitivity analysis took the form of a few sentences, in others it involved a labeled section that was clearly influenced by the HLF Evaluation Guidance describing how much of a difference would have happened anyway without the project.

2. Criterion 2: Counting, involving, choosing indicators that matter

Appropriate and methodical ways of asking were used which provide robust evidence including coverage of wellbeing as well as demographic, economic, and social capital issues where appropriate.

² This analysis was possible for the 234 reports analysed between December 2016 and January 2018. This data is not available for the first 92 reports analysed in 2015.

40% of reports scored good/excellent for this criterion (+3 from review of 200 reports) and 60% scored adequate/poor (-3 from review of 200 reports). 27% overall scored poor.

The 10% of reports that scored excellent for this category provided detailed summaries of the methods used which included qualitative and quantitative research, external sources as well as internal data, included coverage of demographics and considered wellbeing.

Again, for the latest 234 reports evaluated, we were able to analyse how the reports scored for each aspect of this criterion.

64% included external sources (e.g. community feedback) as well as internal data

- Over a third of reports did not include any external sources, instead they relied purely on internal project data and the views of the project staff.

57% included a number of different ways of asking people involved in the project (volunteers, project staff, visitors, participants, focus groups)

49% included qualitative and quantitative data

- The best reports were explicit about exactly what research took place.
- Adequate and poor reports often lacked information on whether or how they had gathered data, failing to demonstrate an understanding that data needed to be gathered in a methodical, robust way.
- Qualitative data collection varied widely, from the minimum of including feedback comments (from comment cards and visitor books) to analysis of focus group discussions and depth interviews, supported by verbatim.
- Quantitative data collection also varied widely, from small scale paper based or Survey Monkey enabled questionnaires to (more rarely) professionally designed and conducted surveys where attempts were made to collect data from a representative mix of participants.
- Many of the poor reports did not refer to any research methods at all and simply described the overall story of the project. Other poor reports mentioned that there were evaluation forms or that people's views were sought but did not include any details about how many forms were completed or how the views were obtained.

28% included coverage of demographics (gender, ethnicity, socio-economic, disabled) and/or economic impacts and/or social capital issues

- Treatment of demographics was a consistently weak area.
- It varied widely from collecting basic information from visitors, participants and volunteers about gender and age, to more detailed information about ethnicity, socio-economic group and disability.
- At the simplest level, this information was collected by staff observation or counts regarding types of tickets purchased, for example: family; child; senior citizen.
- In other reports the information was collected using quantitative surveys ranging from self-completion paper questionnaires to face-to-face interviews to online surveys.
- Variance in degree to which demographic data was provided and analysed indicates that some grantees were unclear or unaware about the value of, or need for, measuring and including this data, despite the fact that the 2012 HLF Evaluation Guidance makes it clear

that projects would be asked to provide this information in the Evaluation Questionnaire, to be completed within a year of the project ending.

- Even some projects whose HLF Approved Purposes were to engage with a wider audience, did not include data to show the extent to which this objective had been met.
- We hypothesise that because this data needs to be collected during the life of the project, the degree to which it is measured is dependent on the amount of evaluation planning which took place at the beginning of a project, as well as the capacity and motivation of project staff to implement a system to capture this data.

16% adequately addressed how people’s levels of wellbeing might have changed as a result of the project

- Few reports discussed how people’s wellbeing might have changed as a result of the project, and those that did tended to be good or excellent.
- A few projects involving large numbers of volunteers collected this information in surveys and evaluation forms using open-ended qualitative questions such as *“What did you gain personally from being involved in the project?”* or *“How has your involvement with x contributed to your life in the area?”*
- Some reports included direct verbatim quotations as evidence of the impact on wellbeing. Others included detailed case studies about individuals to demonstrate the positive impact that their involvement in the project had on them.
- Some of these reports considered the economic impact for the individuals participating, providing evidence of those who had been able to find employment as a result of their involvement in the project.

3. Criterion 3: Beyond counting

Data was subject to robust analysis to provide evidence on outcomes.

Many reports were unable to demonstrate that the data (where it existed) had been analysed robustly. This criterion had the second highest proportion of reports which scored adequate/poor, with over two thirds, 68%, receiving these scores. Just over a third of all reports, 34%, scored poor.

Again, for the latest 234 reports evaluated, we were able to analyse how the reports scored for each aspect of this criterion.

44% of these reports' analysis of the qualitative data brought any explanations for any trends/patterns identified

- In the strongest reports, qualitative research was used to elaborate on the quantitative findings and was evidenced in the appendix as well as in short case study examples throughout the report.
- Some also described how their analysis of qualitative feedback gathered throughout the project had enabled the grantees to make positive improvements during it.
- In the strongest reports, the qualitative verbatim were attributed to a type of stakeholder or participant, with some reports including detailed lists of names of people who had been interviewed and their relationship to the project.
- However, the weakest reports used qualitative feedback and selected verbatim without explanation or qualification to support their positive findings.
- Some of the weaker evaluation reports only included selected descriptions or comments from surveys, evaluation sheets or comments books.
- In some cases, the survey answers were included in the appendix as evidence, but there was no attempt made in the report to explain or analyse the findings from these.

41% of the reports' analysis of the quantitative data included trends/patterns

- Whilst many reports mentioned the use of surveys to collect evidence from project participants, it was clear that the understanding of what a survey involved varied widely.
- In many of the adequate/poor reports it was used as a way of describing any form of feedback obtained from project participants or visitors, regardless of how any material was collected and how many people were included.
- In some of the weaker reports whole feedback forms were included, suggesting that there was limited understanding of how to compile and use the data in aggregate form.

40% made comparisons with a baseline

- Two fifths of the 234 reports established baselines with which later data could then be compared.
- These were often simple baselines of topline data such as numbers of visitors before and after the project work, or number of website visitors, volunteers or change in membership.
- A number of environmental projects included baseline data about species numbers before and after to demonstrate how the project work had made a positive difference to specific flora or fauna populations.

- A few good/excellent reports included more detailed baseline data to compare with, such as external surveys which enabled the project to be benchmarked against similar heritage attractions or volunteer attitudes.

31% sample sizes were considered

- Many reports were not transparent about the methodology or sample size.
- Even some of the better scoring reports did not always provide information about the sample and base sizes of the data presented.
- This indicates that there is a lack of awareness about the importance of including this information when reporting on research/evaluation findings.
- In many cases, even where sample sizes were provided, the samples were often small – without acknowledgement of the ramifications of this.
- It was common for reports which scored adequate/poor for this criterion to report findings from very small base sizes (e.g. 20 people) in percentages, which gave a misleading impression to the findings.
- Some reports included estimates calculated from overall attendance figures for events or locations at which the project featured, as evidence that the project had engaged with a wide audience, when there was no project specific data collected to support this claim. One project asserted that it had reached 111,092 people, based on a wide range of estimates. However, it did not collect any survey data from any of these individuals.

29% contained evidence that the evaluation considered the robustness and comprehensiveness of the data collected and highlighted any limitations

- Less than a third of these reports overtly considered this.
- A few reports written by external consultants included useful information about how to interpret the quantitative data such as explaining confidence intervals and the degree to which the findings were statistically significant.
- However, many of the weaker reports had compiled evidence on a self-selecting basis, for example, through the use of comment books or feedback forms that were completed voluntarily by participants.
- Some reports acknowledged this bias, but most did not, and seemed unaware of it.

4. Criterion 4: Avoiding bias

The evaluation is objective and free from bias.

Many reports struggled to demonstrate that they had sought to avoid bias. As before, this criterion had the lowest number of good/excellent scores. Three quarters of reports scored adequate/poor, with 40% scoring poor overall.

It is clear that there is an inherent tension within self-evaluation between a project telling its own story to an external audience, whilst also avoiding bias. Many reports seemed to be written primarily as project history documents, rather than documents which sought to evaluate objectively and identify openly where things did not go well and could be improved in the future.

Amongst the many adequate/poor scoring reports for this criterion, many reports appeared to be written solely from the viewpoint of the project lead. These reports recorded what these individuals had done, and it did not appear that it was in the author's interest to be genuinely objective.

Again, for the latest 234 reports evaluated, we were able to analyse how the reports scored for each aspect of this criterion.

41% included a clear template for how the evaluation would be undertaken and the conclusions reached

- The best reports were clear about how the evaluation had been undertaken and the conclusions reached.
- In just under three fifths, 59%, of recently analysed reports, no template or explanation was provided for how the evaluation took place.
- It was therefore not possible to understand what specific processes and activities had been put in place to demonstrate how the project had been evaluated.

26% achievement is referenced to external standards

- Some reports narrowly avoided achieving a poor score in this criterion because they contained some reference to achieving external standards, for example winning awards, being shortlisted for Artfund Museum of the Year, achieving a Visitor England Quality Mark, chosen by a Royal Commission to be a case study, comments by English Heritage Building Inspector.

22% had evidence of management oversight and/or quality control for the evaluation

- Some examples of this included where an external evaluator had been appointed and their involvement throughout the project was described; oversight of the evaluation by a steering group or management team; and evidence of the evaluation process leading to changes being made to improve the project.

7% had evidence of external challenge or scrutiny of the approach used and the conclusions reached

- Of the few reports who had evidence of this, none were poor reports.

5. Criterion 5: Structuring the process of understanding

The results are clearly and sufficiently presented.

45% of reports scored good/excellent and 56% scored poor/adequate for this criterion.

As before, it had the largest proportion of reports scoring adequate, with 41%, achieving this score. This was because many were able to describe clearly the project background and context, the first aspect of this criterion. However, the weakest reports failed to even include this scene setting in their report, despite the fact that it would have been created for the original grant application.

Reports varied widely in their style of presentation and the degree of effort that had gone into creating them.

Again, for the latest 234 reports evaluated, we were able to analyse how the reports scored for each aspect of this criterion.

81% clearly presented the project background and context

- Whilst just over four fifths of reports did this, almost a fifth did not even provide a summary of the background information which would have been available in their original HLF grant application documents.

37% of the reports' conclusions were clearly supported by the data presented

- Just under two thirds of these reports made conclusions for which there was no clear data evidence.

36% made appropriate use of tables, graphics etc.

- Just over one third of these reports contained graphics and charts to visually depict the data.

34% of the reports followed the structure recommended in the HLF guidance

6. Criterion 6: Improve not just prove

The conclusions and recommendations are sufficiently clear to enable stakeholders to identify and apply any lessons learned.

After **critterion 1: telling the project story**, this criterion had the highest proportion of reports scoring good/excellent (47%) and the lowest proportion of reports scoring poor/adequate (53%). The reports which scored poorly for this criterion did so because they lacked a conclusion or any considerations around efficacy and efficiency.

Again, for the latest 234 reports evaluated, we were able to analyse how the reports scored for each aspect of this criterion.

72% contained conclusions which offered insights and lessons learned

- The conclusions and recommendations were not always specifically labeled as such, so when evaluating this we gave the benefit of the doubt to reports where this material is included but embedded, rather than being within a discrete section.

67% identified areas for improvement or issues that could improve effectiveness in future. 35% included evidence that the grantees had undertaken some sort of reflective or lessons learned exercise to indicate that the evaluation was likely to lead to improvements on future or similar activities

- There was considerable variation in the amount of analysis and detail provided about the lessons learned, the extent to which they followed logically from the evidence within the report and the degree to which they could provide meaningful lessons more widely within the project, lead organisation or externally within their sectors.
- Some reports explained that changes had already been made to the project or would be implemented in the future as the result of reflecting on areas for improvement.

8% of the evaluation results had been made available to stakeholders and other interested parties

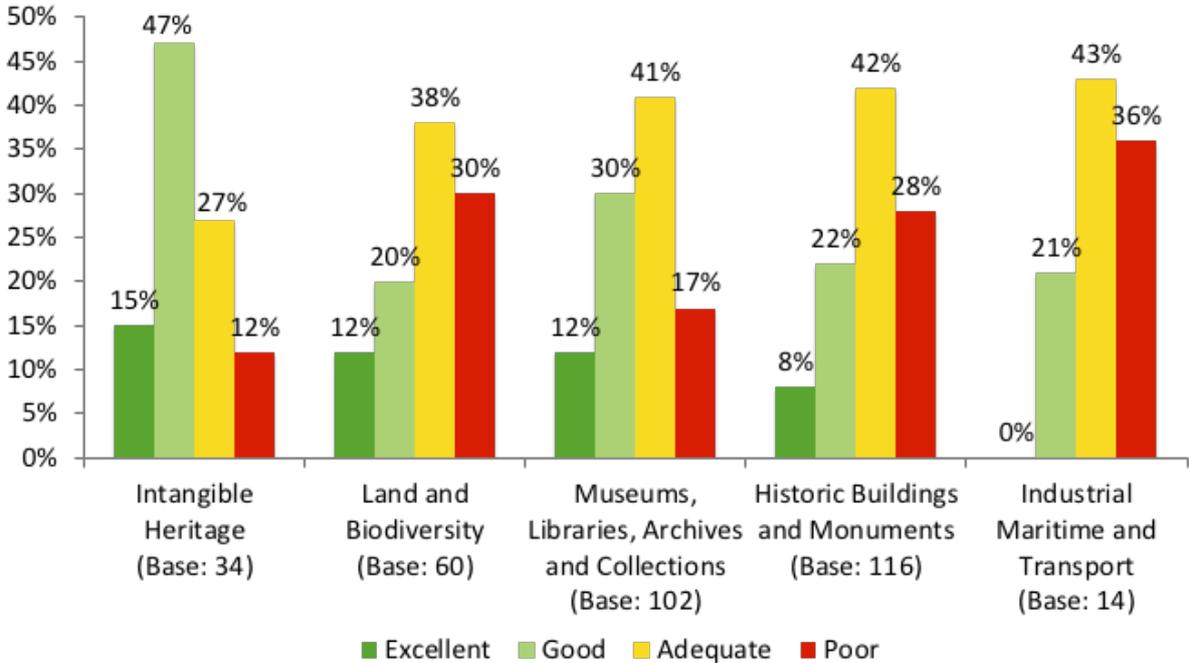
- Few reports contained evidence that the evaluation results had been made available to stakeholders and other interested parties and very few reports contained a clear list of the stakeholders to whom the report had been sent.

APPENDIX 2: Analysis of heritage area, region and financial size of project

As discussed in section 5, we could find no statistical dependence for quality of reports by heritage area, region or financial size of the project.

We have therefore included the following data as a record only.

Figure 15: Quality of reports by heritage area

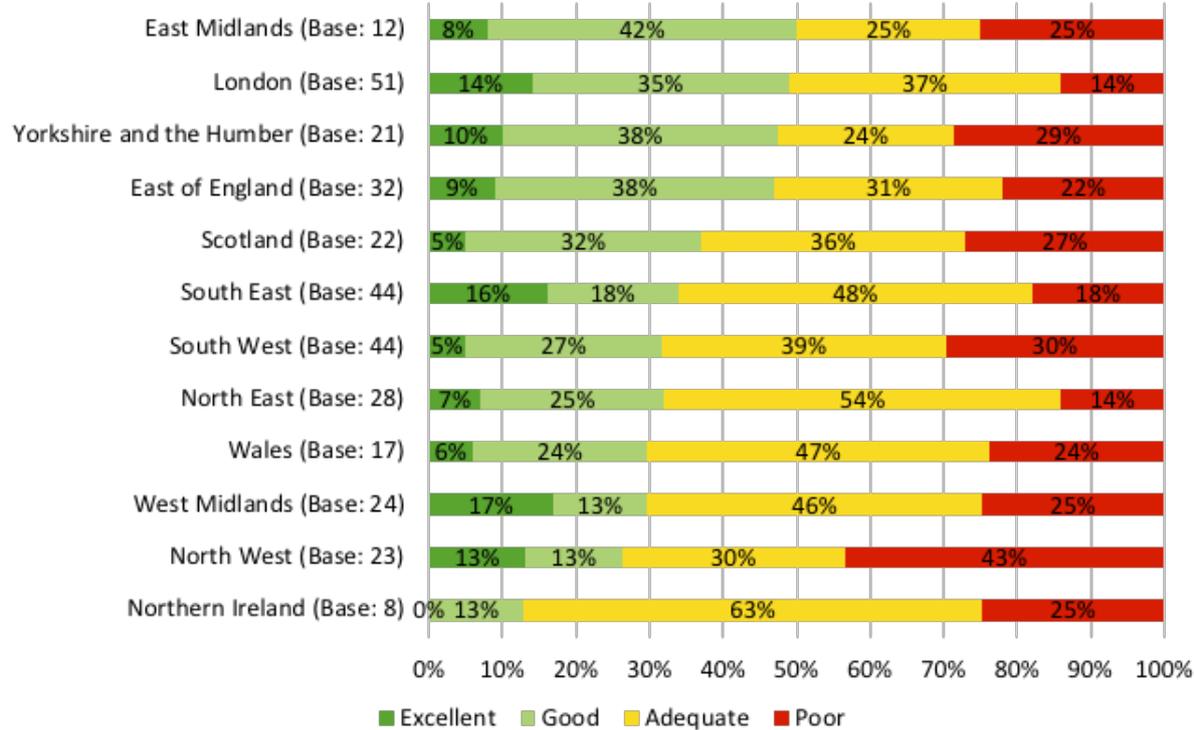


Source: HLF Heritage Programme Database Analysis (Base 326 self-evaluation reports)

We have analysed the mean values of the overall assessment score for each heritage area. Industrial, Maritime and Transport records the lowest score (1.9), but there are only 14 cases in this group, while Intangible Heritage is the highest (2.7). However, all group means are close to the overall mean of 2.2 and the differences are not statistically significant.

Figure 16 shows the distribution of overall criteria scores for each region. While these are broadly similar, the North West stands out with particularly high share of evaluations - over 40% - ranked as Poor and over 70% as Poor or Adequate, although this region also records the fourth highest share rated as Excellent.

Figure 16: Quality of reports by geographic region

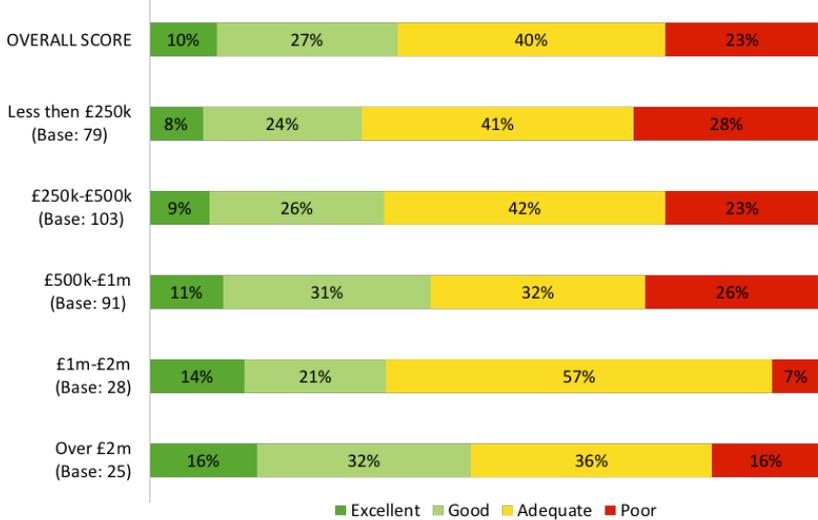


Source: HLF Heritage Programme Database Analysis (Base 326 self-evaluation reports)

We have analysed the average (mean) scores for each region with the standard deviations. The lowest average score is for Northern Ireland, but this is based on only 8 reports. There are no statistically significant differences between average scores for the regions.

Figure 17 shows the quality of reports by financial size of project. The small to medium project results account for most of the data and dominate the statistical analysis which shows very low correlation between assessment scores and project size in general - that is, scores do not improve systematically with project size. The chart indicates that the very largest projects record lower shares of poor evaluations and higher shares of reports rated excellent. But the £1 - 2 m scale projects also show the highest share that score adequate, and the larger projects (over £1m) account for less than 20% of the total.

Figure 17: Quality of reports by financial size of project



Source: HLF Heritage Programme Database Analysis (Base 326 self-evaluation reports)

APPENDIX 3: Overall summary of self-evaluation report strengths and weaknesses

In this section we have, by way of summary, highlighted particular observations from the excellent reports and the poor reports, by way of understanding how reporting might be improved.

Observations from the strengths of the excellent reports:

- All of these reports embedded robust data collection methods at the outset of the project.
- Many of them used evaluation throughout the project lifespan to continually test and refine the success of activities and to monitor their impact. Some were able to demonstrate how self-evaluation findings have already had an impact on the ongoing success of the project.
- All demonstrated expert understanding of research methodologies and of impact evaluation.
- Many provided commentary on the robustness of the data and guidance for how to interpret the findings, taking into account sample sizes and confidence intervals.
- There was often a good balance between the detail of qualitative case studies and what projects meant to people individually, combined with the broader picture provided by quantitative surveys.
- Some reports referenced evaluation metrics and project successes to external references thereby setting their findings within a wider context e.g. other surveys produced by visitor / volunteer organisations, Generic Learning Outcomes.
- In some cases, the more succinct excellent reports were supported by appendices, which included detailed supporting evidence such as a list of who was consulted, evaluation plans, activity plans and separate research reports.
- Many were explicit about how the learning from this project and the evaluation report would be used to inform best practice in future. Some explained exactly whom the report would be shared with, how they would use it and what impact this was expected to then have.
- All the reports clearly and explicitly focused on outcomes and legacy.

Observations from the weaknesses of the poor reports:

- It was clear that evaluation had not been a priority for these projects.
- Many reports were short summaries and lacked sufficient detail about their project. Some were incomplete reports despite their title being 'evaluation report': they partially reported on an aspect of the project or focused on the achievement of milestones in terms of project management and processes rather than outcomes. Some reports only contained activity summaries, some even included conservation plans and marketing plans.
- Some reports lacked a clear structure or had no introduction / aims or objectives section in which to understand the context of the project.
- In many cases there was no evidence of any evaluation data. Therefore, reports often relied on the perspective of the author/s, or used selective anecdotal data to support the findings. It was often unclear how the report's judgments had been made.
- Where some data was provided, the reports did not demonstrate that they could analyse this robustly. Some reports included information in appendices but did not refer to this in the report nor attempt to analyse it.
- Many focused on universally positive/anecdotal verbatim and commentary, asserting that objectives had all been met despite a lack of robust evidence to support this. Where some objectives or targets had clearly not been met, there was often a lack of explanation for this.

- Many lacked reflection and insight into the strengths and weaknesses of the project.
- Many did not attempt to consider objectively whether any lessons had been learned.
- Many did not consider outcomes or refer to the project's wider impact outside of the fact that it had taken place and been delivered on time and to budget.
- A few reports had clearly not been proofread and had missing sections.

APPENDIX 4: HLF adopted approach to reviewing grantee evaluations

Created to review first 92 self-evaluation reports and then used in subsequent 2017 and 2018 reviews

Please note RF Associates did not create this methodology and criteria

Methodology directions

1 Introduction

Reviews of grantee evaluations should be modeled on the Heritage Grants programme evaluation produced by Evaluation Expertise, for HLF, and published in September 2015.

Reviews should include a comparative appraisal of: -

- the quality, scope and methodology of the self-evaluated reports;
- the type, range and quality of activities and outcomes achieved by completed projects including a mapping of the impact of their work onto HLF's current framework of outcomes for heritage, people and communities.

Since 2008, organisations awarded funding by HLF have been required to submit a self-evaluation of their project work upon completion of the project. The grant application guidance³ makes it clear that as well as providing regular progress reports, the project must send a final evaluation report at the end of the project and that this is needed before the Fund can release the final 10% of the grant.

When projects apply for a grant, HLF asks applicants to tell the 'story' of their project, by explaining why they want to do their project – the need or opportunity that exists; what they will do – both capital works and activities and what difference the project will make for heritage and for people. HLF asks grantees to repeat this exercise in their evaluation, comparing what they hoped or expected their project to achieve with what actually happened. The report is also expected to include details of issues or events relating to managing the project, staffing or the programme that they consider have been successful or unsuccessful.

HLF does not prescribe how grantees should carry out their evaluation. However we have issued Evaluation Guidance (4) and expect projects to follow the principles set out in this guidance in carrying out their self- evaluations.

HLF can contribute up to 3% of the total grant amount towards the cost of evaluation.

2 Assessment of evaluation reports

It is important to have a clear and fair set of criteria against which to assess the quality of each of the evaluation reports. Drawing on best practice⁵, the characteristics of a strong evaluation include:

- Trustworthiness: the data collection and analysis should be rigorous and stand up to

³ Heritage Grants : Grants of more than £100,000 Introduction and help notes, HLF February 2012

⁴ Heritage Lottery Fund, *Evaluation Good-practice guidance*, October 2012

⁵ In particular, the original reviewers drew on the classic 1990 text by Huey Chen "Theory-driven evaluation"

scrutiny

- Objectivity: personal preferences or loyalties should not ‘contaminate’ evaluation results.
- Responsiveness: an evaluation report should be relevant and useful not only to the funders, but to other stakeholders such as programme staff, participants and local communities.
- Generalisability: the results can be used by stakeholders for decision-making in future.

HLF also specifies criteria for effective evaluation in the guidance they provided to each grantee. This guidance identifies four key factors:

- Look beyond the outputs: To evaluate change means looking at the real differences made by a project – this difference is often referred to as the outcomes of a project, or its impact.
- Tell the story: It is important to be clear about the link between activities and actions and the change that they are designed to bring about.
- Involve people: A search for a meaningful narrative requires a dialogue, rather than a mere extraction of data from people
- Indicators that matter: they must be chosen based on what people associated or benefiting from the project have identified as important to measure, and not just what is easiest to count

These sources have been synthesized to develop a set of six criteria suitable for use in assessing the evaluation of heritage projects. These are set out in Figure 1 with the relevant matching principles from the HLF’s own evaluation guidance shown in brackets.

Figure 1: Criterion used to assess and grade evaluation reports.

<p>Criterion 1: The evaluation provides a logical framework setting out linkages between activities, expected outputs and outcomes for all elements of the project (Telling the project story).</p> <p>Criterion 2: Appropriate and methodical ways of asking were used which provide robust evidence including coverage of well-being as well as demographic, economic, social capital and quality of conservation issues where appropriate (Counting, involving, choosing indicators that matter).</p> <p>Criterion 3: Data was subject to robust analysis to provide evidence on outcomes (Beyond counting).</p> <p>Criterion 4: The evaluation is objective and free from bias (Avoiding bias).</p> <p>Criterion 5: The results are clearly and sufficiently presented (Structuring the process of understanding).</p> <p>Criterion 6: The conclusions and recommendations are sufficiently clear to enable stakeholders to identify and apply any lessons learned (Improve not just prove).</p>
--

From these we have developed a set of indicative characteristics under each criterion to grade each report against. Performance against each criterion could be graded poor, fair, good or very good and we also made a similar overall assessment of each report. The grading schematic is shown in full at Annex A.

Reviewers should not follow an unduly mechanistic approach. The final grade should not be just the average of the scores against each criterion, but an overall view on the report taking into account the strengths and weaknesses demonstrated against all the criteria. The schematic should be used to guide judgements and help ensure consistency. But the report grades remain a judgement.

3 Assessment against outcomes

HLF currently uses an outcome framework which sets out 14 potential desired outcomes from its investment. The framework has 4 outcomes for heritage, 5 outcomes for people, and 5 outcomes for communities. For each potential outcome the framework describes the outcome and provides examples of how a project will know if it has achieved the outcome.

The framework was introduced in SF4 and all projects funded during this strategic plan period (April 2013 onwards) have presented the case for their projects against the outcomes framework and have been assessed against it.

Where projects were funded before SF4, it is still possible to map the evidence presented in self evaluation reports against the 14 outcomes, by interpolating from the evaluation reports.

The review should seek to identify evidence that a project has achieved an outcome under the current framework, and that this achievement is likely to have resulted: -

- because of the nature of the activity described within the report and/or
- because the quantitative and qualitative evidence presented in the report.

Where an outcome has not been mapped against the Framework, it does not mean that the project did not have that outcome but rather that the evaluation report does not provide clear evidence that can be interpolated. Once this scoring has been completed the reviewer should look for patterns across outcome and type of project in order to identify common features and lessons.

Please note that for the 2017 Review

For the purposes of evaluating the second set of 108 self-evaluation reports, in addition to following the process above RF Associates devised a numerical scoring system to ascertain the overall report grade:

- Excellent = 4, Good = 3, Adequate = 2, Poor = 1
- Reports were given 6 scores – one for each of the 6 criteria; the maximum possible total was 24 (6 x Excellent scores), and the minimum possible total was 6 (6 x Poor scores)
- Reports were then allocated an overall rating based on their total score, unless researchers considered there was a reason that their scoring should be slightly higher or lower:
 - Poor 0-9
 - Adequate 10 – 15
 - Good 16 – 19
 - Excellent 20 - 24

Annex A – Criterion Scoring

Please note that for the second evaluation review in 2017, it was agreed that the original terms of 'Very Good' and 'Fair' would be replaced respectively with 'Excellent' and 'Adequate'.

Criterion 1: The evaluation provides a logical framework setting out linkages between activities, expected outputs and outcomes for all elements of the project (Telling the project story)

Poor

1.1 There was no clear rationale explained or the activities undertaken were described but without any explanation of the outputs and outcomes.

Fair/Adequate

1.2 One of the following three issues was adequately addressed:

- There was an explanation of how each activity contributed to the outputs achieved
- There was a clear explanation of how each output had led to outcomes or wider impacts
- There was some sensitivity analysis or other estimation of the extent of attribution of activities to outputs or outputs to outcomes

Good

1.3 Two of the following three issues were adequately addressed:

- There was an explanation of how each activity contributed to the outputs achieved
- There was a clear explanation of how each output had led to outcomes or wider impacts
- There was some sensitivity analysis or other estimation of the extent of attribution of activities to outputs or outputs to outcomes

Very Good/Excellent

1.4 All of the following three issues were adequately addressed:

- There was an explanation of how each activity contributed to the outputs achieved
- There was a clear explanation of how each output had led to outcomes or wider impacts
- There was some sensitivity analysis or other estimation of the extent of attribution of activities to outputs or outputs to outcomes

Criterion 2: Appropriate and methodical ways of asking were used which provide robust evidence including coverage of well-being as well as demographic, economic, and social capital issues where appropriate(Counting, involving, choosing indicators that matter)

Poor

2.1 Ways of asking are limited or not clear

Fair/Adequate

2.2 One or two of the following four was adequately addressed:

- How people's levels of well-being might have changed as a result of the project
- There was a number of different ways of asking people involved in the project (volunteers, project staff, visitors, participants, focus groups)
- The mix included qualitative and quantitative
- The data collected included external sources (eg community feedback) as well as internal data.
- Where appropriate the analysis included coverage of demographics (gender, ethnicity, socio-economic, disabled) and/or economic impacts and/or social capital issues

Good

2.3 Three or four of the following issues was adequately addressed:

- How people's levels of well-being might have changed as a result of the project
- There was a number of different ways of asking people involved in the project (volunteers, project staff, visitors, participants, focus groups)
- The mix included qualitative and quantitative
- The data collected included external sources (eg community feedback) as well as internal data.
- Where appropriate the analysis included coverage of demographics (gender, ethnicity, socio-economic, disabled) and/or economic impacts and/or social capital issues

Very Good/Excellent

2.4 All of the following issues were adequately addressed:

- How people's levels of well-being might have changed as a result of the project
- There was a number of different ways of asking people involved in the project (volunteers, project staff, visitors, participants, focus groups)

- The mix included qualitative and quantitative
- The data collected included external sources (eg community feedback) as well as internal data.
- Where appropriate the analysis included coverage of demographics (gender, ethnicity, socio-economic, disabled) and/or economic impacts and/or social capital issues

Criterion 3: Data was subject to robust analysis to provide evidence on outcomes (Beyond counting)

Poor

3.1 Data is given but there is limited analysis

Fair/Adequate

3.2 One or two of the following four issues was adequately addressed:

- Comparisons are made with a baseline
- There was evidence that the evaluation considered the robustness and comprehensiveness of the data collected and highlighted any limitations
- The analysis of the quantitative included trends/patterns
- Sample sizes were considered
- The analysis of the qualitative data brought any explanations for the trends/patterns identified.

Good

3.3 Three or four of the following issues were adequately addressed:

- Comparisons are made with a baseline
- There was evidence that the evaluation considered the robustness and comprehensiveness of the data collected and highlighted any limitations
- The analysis of the quantitative included trends/patterns
- Sample sizes were considered
- The analysis of the qualitative data brought any explanations for the trends/patterns identified.

Very Good/Excellent

3.4 All of the following issues were adequately addressed:

- Comparisons are made with a baseline
- There was evidence that the evaluation considered the robustness and comprehensiveness of the data collected and highlighted any limitations
- The analysis of the quantitative included trends/patterns
- Sample sizes were considered

- The analysis of the qualitative data brought any explanations for the trends/patterns identified.

Criterion 4: The evaluation is objective and free from bias (Avoiding bias)

Poor

4.1 There was no information on how the evaluation was undertaken and no evidence of any controls in place to minimise the risk of bias.

Fair/Adequate

4.2 There was information on how the evaluation was undertaken but only one of the following four issues had been adequately addressed:

- A clear template for how the evaluation would be undertaken and the conclusions reached
- Evidence of management oversight and/or quality control for the evaluation
- Evidence of external challenge or scrutiny of the approach used and the conclusions reached
- Achievement is referenced to external standards

Good

4.3 There was information on how the evaluation was undertaken but only two of the following four issues had been adequately addressed:

- A clear template for how the evaluation would be undertaken and the conclusions reached
- Evidence of management oversight and/or quality control for the evaluation
- Evidence of external challenge or scrutiny of the approach used and the conclusions reached
- Achievement is referenced to external standards

Very Good/Excellent

4.4 There was information on how the evaluation was undertaken and each of the following issues had been adequately addressed:

- A clear template for how the evaluation would be undertaken and the conclusions reached
- Evidence of management oversight and/or quality control for the evaluation
- Evidence of external challenge or scrutiny of the approach used and the conclusions reached
- Achievement is referenced to external standards

Criterion 5: The results are clearly presented (Structuring the process of understanding)

Poor

5.1 None of the following issues had been adequately addressed:

- The project background and context were clearly presented
- The report followed the structure recommended in the HLF guidance
- There was appropriate use of tables, graphics etc.
- The conclusions were clearly supported by the data presented

Fair/Adequate

5.2 Only one of the following issues had been adequately addressed:

- The project background and context were clearly presented
- The report followed the structure recommended in the HLF guidance
- There was appropriate use of tables, graphics etc.
- The conclusions were clearly supported by the data presented

Good

5.3 Two or three of the following issues had been adequately addressed:

- The project background and context were clearly presented
- The report followed the structure recommended in the HLF guidance
- There was appropriate use of tables, graphics etc.
- The conclusions were clearly supported by the data presented

Very Good/Excellent

5.4 All of the following issues had been adequately addressed:

- The project background and context were clearly presented
- The report followed the structure recommended in the HLF guidance
- There was appropriate use of tables, graphics etc.
- The conclusions were clearly supported by the data presented

Criterion 6: The conclusions and recommendations are sufficiently clear to enable stakeholders to identify and apply any lessons learned (Improve not just prove)

Poor

6.1 There was no conclusion, or the conclusion was a simple assurance that all activities had been completed without any analysis of their efficiency or effectiveness

Fair/Adequate

6.2 One of the following issues was adequately addressed:

- The conclusion offers insights and lessons learned
- The conclusion identifies areas for improvement or issues that could improve effectiveness in future
- There was evidence that the grantees had undertaken some sort of reflective or lessons learned exercise to indicate that the evaluation is likely to lead to improvements on future or similar activities
- The evaluation results had been made available to stakeholders and other interested parties

Good

6.3 Two or three of the following issues were adequately addressed:

- The conclusion offers insights and lessons learned
- The conclusion identifies areas for improvement or issues that could improve effectiveness in future
- There was evidence that the grantees had undertaken some sort of reflective or lessons learned exercise to indicate that the evaluation is likely to lead to improvements on future or similar activities
- The evaluation results had been made available to stakeholders and other interested parties

Very Good/Excellent

6.4 All of the following issues were adequately addressed:

- The conclusion offers insights and lessons learned
- The conclusion identifies areas for improvement or issues that could improve effectiveness in future

- There was evidence that the grantees had undertaken some sort of reflective or lessons learned exercise to indicate that the evaluation is likely to lead to improvements on future or similar activities