



Creative Partnerships

exciting minds



Creative Partnerships National External Evaluation Audit Report

Project Director – David Wood

Project team:

John Hole

Rachel Payne

Phil Whitehead

Mandy Winters

Peter Muschamp – Adviser to the report

October 2007

Contents	Page
Executive summary	3
1 Introduction	5
2 The Audit Methods Used	6
3 How CP National Office Contributes to Evaluation	7
4 How Regional CP Offices, their Programmers and ‘Creative Agents’ contribute to evaluation	12
5 The Role of Schools in the Evaluation of CP	18
6 The Role of Teachers and Creative Practitioners in Evaluation	22
7 Conclusions	24
8 Appendix A – The aide memoire for visits	27
9 Appendix B – an explanation of terms and concepts	31

Acknowledgments

The authors wish to acknowledge the assistance and co-operation of Creative Partnership staff both at regional offices and at the National Office during the data gathering for this audit report.

Executive Summary

This is the first of three national audits of Creative Partnerships. It is designed to:

- analyse the evaluation process across the country, i.e. are reports rigorous, fit for purpose, consistent, comparable?
- validate and disseminate regional strengths and good practice in evaluation;
- synthesise and interrogate issues common to Creative Partnerships.

The audit team from Oxford Brookes University:

- reviewed 10 sample evaluation reports from each CP;
- visited 6 representative CPs to see evaluation work, and interview key staff;
- scrutinised sample evaluation material at a different 6 CPs.

The CP National Office supports evaluation by promoting a CP project evaluation system known as the 'Creative Partnerships Evaluation Toolkit,' and hosting an online database which collects evaluation data. Use of the Toolkit was patchy across CPs, although the majority of them were using the principles of the Toolkit. Most of the CP staff we interviewed were critical of the Toolkit, claiming that it had been hurriedly introduced without an external analysis or sufficient consultation with CPs. They were unanimous in suggesting more training in using the Toolkit.

We judge that the Toolkit is a valid means of evaluation, which could potentially be the vehicle for collecting large-scale reliable and legitimate information about the impact of CP. However, we agree with many of the criticisms, and recommend that a consultative process is initiated to refine the Toolkit, especially the questions which it comprises. A new programme to train the CP programmers and/or creative agents who administer the Toolkit should follow. The online database used to collect feedback from teachers and creative

practitioners is not currently well designed to produce digestible summary information for CP stakeholders. We recommend that it is overhauled.

Most regional CP offices had evaluation strategies. Elements of these could provide useful refinements to the Toolkit. In particular we saw several useful systems for eliciting feedback from pupils and young people. In CPs where there appeared to be no overarching evaluation strategy, the evaluation evidence is unconvincing. Three CPs had commissioned independent quality surveys of their evaluations. These contained useful advice about enhancing quality.

Schools and teachers usually demonstrated considerable commitment to evaluation activities. There were half a dozen examples of schools analysing their own data to throw light on the impact of CP. Occasionally school leadership teams were apathetic about, or obstructive to, CP projects. We suggest that CPs ensure that schools comply with their responsibilities to participate in CP evaluation, but also that CP staff support busy school managers by identifying CP evaluation data and preparing evidence of CP's impact in SEFs and CSEFs.

Creative practitioners tended to recognise gains in their understanding of schools and education, rather than gains in their own artistic practice.

The bulk of evaluation evidence comprises largely positive prose testimony about projects. By contrast there is currently not enough quantitative data to compare projects and CPs or make any reliable and legitimate claims about the impact of CP across the country. We believe that collecting this data is a priority if CP is to continue fulfilling its objectives and the DCMS Policy and Delivery Agreement. We accept that a revised Evaluation Toolkit would be an effective vehicle for this.

1 - Introduction

Creative Partnerships is the Government's flagship creativity programme for schools and young people, managed by Arts Council England and funded by the Department for Culture, Media and Sport (DCMS) and the Department for Children, Schools and Families¹. It aims to develop:

- the creativity of young people, raising their aspirations and achievements;
- the skills of teachers and their ability to work with creative practitioners;
- schools' approaches to culture, creativity and partnership working; and
- the skills, capacity and sustainability of the creative industries.

Creative Partnerships was initially designed and funded as a pilot programme (Phase 1) from April 2002 to 31st March 2004. This phase had a budget of £40 million. Sixteen pilot areas were selected by Ministers from a list of the most economically and socially challenged neighbourhoods in England. In the July 2002 Comprehensive Spending Round, Arts Council England was awarded funding for Creative Partnerships to continue beyond the original pilot programme. DCMS committed £70 million to continue to support the existing 16 Creative Partnerships and to develop 20 new Partnerships in 2004-2006. At the time of writing there are 36 Creative Partnerships working with over 1000 schools.

Creative Partnerships (CP)² National Office at Arts Council England commissioned Oxford Brookes University to conduct three annual audits of the project evaluation processes and practices in the 36 Creative Partnerships across England. This report summarises findings from the first annual audit, covering projects evaluated by CPs in the 2006/7 academic year. A team of five Oxford Brookes staff conducted the audit. The audit team was essentially the same as that which reported on continuing professional development in CP in 2005. Former HMI Peter Muschamp acted as 'critical friend' to the team, refining the audit process and commenting on drafts of the report.

In this first audit the Oxford Brookes audit team has trialled its audit process. This report, therefore, offers provisional observations, findings and recommendations. Through a process of dialogue and a consultation conference for key staff in all of the CPs in October

¹ Formerly the Department for Education and Skills (DfES)

² In this report we use *CP* to denote the entire Creative Partnerships initiative. We use *CPs* to denote the 36 regional offices and their processes.

2007, the audit team hope to enhance its audit procedures in 2007/8 and to assist CPs to refine their evaluations and embed their good practice.

The purpose of this external audit is to:

- analyse the evaluation process across the country, i.e. are reports rigorous, fit for purpose, consistent, comparable?
- validate and disseminate regional strengths and good practice in evaluation;
- synthesise and interrogate issues common to Creative Partnerships.

2 - The Audit Methods Used

The audit team reviewed a sample of ten completed evaluation reports from each CP, this being the size of sample recommended in the Creative Partnerships 'Evaluation Toolkit' (see section 3 below). We made visits to a sample of six CPs in the summer of 2007 and observed evaluation taking place through interviews with teachers and artists and pupils. During each visit we also interviewed key CP staff, usually the CP Director and one or two CP programmers. To ensure consistency, we developed a standard template for our visit questions. This is included at Appendix A (below). We sent our notes on the visits to each CP we visited to check their accuracy.

Finally, in a different sample of six CPs, we audited corroborating or supporting evidence of evaluation, such as Creative Self-evaluation Forms (C-SEFs) pamphlets, DVDs, school improvement plans and Self-evaluation Forms (SEFs). Again, we used a standard set of questions to guide our survey of the supporting evidence. This is also included in Appendix A. So twelve CPs were subject to more in-depth scrutiny in the first audit, and this rolling programme should ensure that the audit team cover every CP in more depth at some time over the three-year period.

This audit report is organised in sections, corresponding to the various agencies and individuals actively involved in CP evaluation. So, in each succeeding section, we outline evidence, findings and recommendations about the contributions to CP evaluation made by:

- the CP National Office;
- regional CP offices and staff;
- schools;
- teachers;

- creative practitioners.

3 - How CP National Office Contributes to Evaluation

The CP National Office supports evaluation by promoting a CP project evaluation system known as the 'Creative Partnerships Evaluation Toolkit,' and hosting an online database which collects evaluation data. In this section we describe and analyse the evidence we found about the system and database.

In discussion with the CP National Office, we worked on the assumption that the Evaluation Toolkit was the standard recommended means of collecting evaluations across CPs. It was developed in CP Kent. CP National Office ran regional training days to introduce the Toolkit and issued a booklet of guidance on using it. The booklet stated that using the Toolkit would ensure:

- common measures to help compare and contrast projects between Creative Partnerships areas;
- a shared language;
- a long-term body of evidence of accountability for DCMS and DfES;
- a planning tool to inform the ongoing development of activity.

For this reason the booklet states that:

It is vital that all area office teams engage with and administer this process (p2-3).

We judge the rationale behind the Toolkit to be appropriate. The above principles align well with the published objectives of CP to provide:

rigorous evidence, through an **agreed** programme of research and evaluation;

and with the DCMS Policy and Delivery Agreement requirements for an evaluation strategy which will provide:

*longitudinal studies to track the **impact** of Creative Partnerships,*

and evaluation which is:

*sufficiently **robust** to form the basis of future **policy development** and potential spending round bids. (7.2)*

(our emphases in bold)

In practice, however, the 36 CPs did not all use the Toolkit system in 2006/7. At the time of writing, thirteen CPs were using the Toolkit and compiling data on the XA online database

hosted by the National Office. A further nine CPs were using close variations on the Toolkit, but not using the XA database. Two CPs claimed that the CP National Office had not made it clear whether the Evaluation Toolkit was the required evaluation system. Seven CPs were not using the Toolkit at all. We discuss their alternative approaches below (see section 4).

During visits to CPs we had several lengthy discussions about the validity of the Evaluation Toolkit and explored the reasons for its uneven adoption. In three CPs the staff felt very strongly that the Toolkit information was frequently flawed, necessarily incomplete, and certainly not a reliable basis upon which to reach conclusions about CP projects and the impact on learning. These staff felt that the Toolkit's grading structure hardly reflected the depth of experience and value of project work. Staff in all six CPs we visited reported that they had received little or no training in using the Toolkit and that its introduction had been hurried and inadequate. Two CP Directors claimed the Toolkit had not been externally analysed before approval was given. At three schools, during evaluation interviews, CP staff asked the audit team how to score, interpret and otherwise administer particular parts of the Toolkit. This implies that there is a need for further training and guidance on interpreting the Toolkit questions, so that evaluations are administered consistently.

Three CP directors were critical of the Evaluation Toolkit because it failed to capture direct pupil feedback, but instead required adult respondents to offer their impressions of the value to pupils. One claimed that the questions were too difficult for busy teachers to answer. Another claimed that the data could be open to abuse and misrepresentation at national level.

Another common criticism during our visits was about the Toolkit's four-point 'Likert-type' attitudinal scoring system. In evidence submitted to us, three CPs had commissioned independent evaluations to analyse the principles behind the Evaluation Toolkit. One of these reports described teachers' objections to the numbers system:

- the subjective nature of the evaluation in general and the scoring in particular;
- the limitations of a scale with just 4 integer points on it (0, 1, 2, 3) with no half or fractional points;
- the problem of using such a simplistic scale to capture the impact of intensive work with a small number of pupils on the one hand and that of more extensive work with a larger number on the other;

- ambiguities in the definition/interpretation of the various terms used in the questionnaires, and overlaps between them;
- individual interviewees (especially coordinators) not being in a position to answer questions and give scores where they hadn't witnessed enough of the work in practice.

On the other hand, the report does not reject quantitative data outright:

It must be acknowledged that this qualitative approach is inherently subjective. The possibility of interviewee bias (conscious or unconscious) cannot be ruled out. In the medium to long term it will be possible to produce quantitative data (eg on attendance, exclusions and educational achievement) that will enable a statistical analysis of the impact of CP that can be considered alongside this and subsequent qualitative analyses.

Another independent report on the Toolkit corroborates negative views about its introduction:

Negative: How the National Evaluation Framework was adopted, reviewed and revised – what lessons have been learned (i.e. late start).

It also points out that information was incomplete and that the Toolkit was seen as unnecessarily complex:

- *Schools, despite promoting, didn't undertake the 'doing' phase.*
- *The questions still aren't right – particularly in the showing questions.*
- *Think we need to keep to a maximum of 5 key areas in each section – the practical experience in implementation is that there is so much cross over anyway in responses to earlier questions that a lot is repeated towards the end.*
- *Really don't think we need 4 phases – three is perfectly adequate and even two might suffice – it's overkill and not really telling us much more.*

However, to some extent the misgivings of CP staff about the principles of the Evaluation Toolkit are unfounded.

First, the Toolkit uses valid forms of enquiry to find out if projects meet CP objectives. So, put simply, it is designed to find out, before and after a project, whether a creative practitioner adds value to a school. Moreover, it asks teachers and creative practitioners for their views on what has changed for the school, young people, teachers and creative practitioners. All of this seems to us to be broadly well targeted at determining the impact

of CP projects. Whilst we accept the view above that questions in the Toolkit need refining and sharpening, we found that some questions devised by those CPs which had rejected the Toolkit were more ambiguous and invalid than those in it.

Secondly, in our comparison of CP evaluations across the country, it was information in Toolkit evaluations that was consistently the most robust, balanced and detailed. It contained far more of the direct 'voices' of teachers and creative practitioners. For example:

The ideas have massive value. It has totally changed the way I teach. I now have the opportunity to be outstanding. I'm now engaging a greater percentage of my class and motivating them to want more and ask for knowledge. At first the pupils may think 'what's she doing?' but when they see the results, they get it. These creative techniques build generic skills too, not just science.

Or less positively:

When well planned and organised the work had a big effect, but this was not consistent. It seemed like a lost opportunity because so much more could have been learnt.

The grades, and prose comments given in the Toolkit could usually be traced back to particular schools and respondents. In this way it was potentially possible to confirm or even challenge these views and opinions. However, the source of the information was frequently missing in CPs which didn't use the Evaluation Toolkit.

Thirdly, despite the criticisms about the lack of 'pupil voice' in the Toolkit, one CP directly transcribed pupil comments into the XA system and another CP offered written guidance on ways of capturing pupil feedback particularly at primary level. This CP suggested that evaluations could include primary pupil storyboards, outlining how pupils responded to different stages of a CP project, or even postcards to 'Barnaby Bear' explaining what they had gained from a project. We saw two other CPs which used 'smiley faces' and other images to elicit basic feedback, even from foundation level pupils.

Finally the use of a four grades in a Likert attitudinal scale (to indicate value) is a widely accepted evaluation method which can reliably discriminate individual views³. A four-point attitudinal scale gives fairly accurate views since the four phrases like 'no value,' 'some value, etc give little scope for misinterpretation. So, whilst there is a case for refining the Toolkit questions, the Toolkit method has the potential to provide large-scale reliable data, from which CPs could learn lessons, refine practice and provide national information on

³ For example, see Robson, C. (2002) *Real World Research*, Oxford: Blackwell p293ff, Cohen, Manion & Morrison (2000),

the impact of CP. Despite this, two CPs introduced 10 point attitudinal scales, and others advocated this method. Such a scale did not produce as clear feedback as the four point scale, however.

We made extensive use of the XA online database to read evaluations from the thirteen CPs which put all of them online. The system used four forms, two four-point attitudinal scales, and two designed for prose comments. The layout and composition of the online system was, however, unhelpful. For example, the 0–3 Likert-type rating scale used for interviewing was converted to an A-D scale online, with 'A,' somewhat confusingly, representing 'no value' rather than 'significant value.' Furthermore, users of the system could not 'print to fit' forms on paper; neither could they combine information from the four forms to generate a single project report. The system contains hundreds of respondents' prose comments but these cannot easily be read on an Excel-based system. These design flaws made it unsuitable either for generating large-scale aggregated data or specific evaluation data on a single project. However, if re-designed, the system could provide, for example, an end user with country-wide data on how CP projects gave new skills to teachers, or with views on the value of CP to schools.

At least two CPs we visited said it was too time consuming to enter data onto the system and around half a dozen CPs reported difficulties in entering data online. For this reason, we recommend (below) a redesigned online system so that a printable version of a project evaluation can be assembled from the four forms, thus making CP evaluation accessible to a range of stakeholders and researchers. In fact, some CPs had designed alternative Toolkit evaluation formats to produce just such reports.

The fact that the Toolkit – or close variations of it – is in widespread use, and that nine CPs have derived their evaluation systems from it, indicates that it has broad support in principle and could become the standard method of evaluation. The Toolkit uses two types of data, qualitative and quantitative, with complementary uses and relevance. At the qualitative end there are large numbers of teacher testimonies; enthusiastic assertions about CP, recording positive narratives about its effects on pupils and the school. This is rich data, classic 'thick' description, but limited to a particular project. At the other end is dryer quantitative data. This is information which does nothing to illuminate the excitement and commitment generated by CP projects. But it could potentially indicate trends: for example, the average score across 55 teachers in a CP could indicate how much they believed the CP projects added value to pupils' problem solving.

One independent consultant's conclusion chimes with our own view:

The framework could, with revision and severe editing, be adopted for use by all creative agents with schools.

So, as feedback suggests, there is room for improvement. We agree with some of the criticisms made by CP staff and independent consultants, particularly about flawed, incomplete, and unreliable information.

Recommendations We recommend that CP National Office considers:

1. editing and expanding the Toolkit booklet to include guidance on interpreting answers, definitions of key terminology and more detailed instructions on administering it;
2. running another consultation and training programme on the use of the Toolkit;
3. re-designing the online database so that raw data can be presented in accessible composite project reports.

4 - How Regional CP Offices, their Programmers and ‘Creative Agents’ contribute to evaluation

In this section we outline our findings about evaluation policy and practice in the regional CP offices. We wanted to find out whether a regional CP was maintaining and improving an effective evaluation system. To do this we looked for evidence that they were:

- **planning** an evaluation process and strategy;
- **doing** effective evaluations;
- **checking** that the system was working well and that evaluations were coming in;
- **responding** to the findings of evaluation by learning lessons and improving practice.

Planning: The majority of CPs have designed and disseminated strategies for conducting consistent evaluation processes. The audit team inferred that those CPs using the Evaluation Toolkit adopt its overall strategy. Of those CPs neither using the Toolkit nor variations on it, seven had published alternative strategies.

Case Study - One CP had created an evaluation framework, which was illustrated in a flow chart. Important features included linking project planning to impact and to the school improvement plan, a process review at the start of the project, involving interviews with teachers and practitioners, a mid-point impact on learning review which informed feedback to the project leaders and an end of project 'impact on learning' review which also involved pupils. In many ways this mirrors the CP Evaluation Toolkit, although data gathering is done by means of 'deep conversations' rather than questionnaires⁴.

Two of the seven CPs in this category had designed valid and robust alternative systems, and one other had a tightly controlled but less valid evaluation strategy. Within these alternative strategies there were some helpful ideas. For example, one project 'start' form asked very relevant questions as an alternative to those in the Toolkit:

How will your project place children and young people as equal partners in the work?

this relates CP objectives to other school priorities such as 'pupil voice,' and Every Child Matters;

How will you make sure that the impact of your work is sustained?

this is designed to throw light on the potential legacy of CP;

How will you share the work with different audiences?

this question seeks information about how lessons learnt from CP projects will be disseminated.

Case Study - One CP used a visual diagram for evaluating a particular project. A particularly interesting aspect of this project is that a graphic wheel was used. Segments of the wheel were labelled, for example, 'playfulness,' or 'inspiration,' and the wheel had ten concentric spaces from centre to circumference. Teachers, it seems, rated themselves somewhere along the ten circles on 2 occasions, with a three month gap in between. The later self evaluation was coloured red, the earlier blue. This gives an accessible graphic representation of the impact of the CP project on teacher development.

⁴ These deep conversations were often in use elsewhere as a way of developing respondents' ideas prior to filling in Toolkit forms.

Five CPs appeared to have no evaluation strategy at all. An independent researcher who reported on one CP pointed out that it:

does not have monitoring and evaluation policy or guidelines: rather, any evaluation takes place on a project by project basis (p21). She concludes, rightly, that projects could not, therefore, be compared and goes on to recommend that the CP should develop an evaluation policy.

In CPs where there appeared to be no overarching evaluation strategy, the evidence is unconvincing. For example, in one CP, no author was named for any of the 'bespoke' prose evaluations so readers and stakeholders could not challenge findings or even question the author. There was no direct testimony from teachers, pupils or creative practitioners. There was no evidence that teachers school or pupils were vouching for the accuracy of the evaluations. All reports contained only positive content and were written in a promotional, rather than an evaluative style. Some ambitious assertions, which could not be corroborated, were made:

There was a significant impact on the pupil's habits holistically and a real impact has been made in their homes. Most of the parents of the pupils were incredibly positive, saying the project had made a huge difference to their family.

The distinction between promotional and evaluative reports was most often blurred in CPs which did not use the Evaluation Toolkit. Clearly promotional reports have their place, not least to encourage more schools to adopt creative approaches to learning. But the purpose of evaluative reports is to gauge the impact of CP projects, to find out what works and to avoid practices which do not work in future.

There were frequent weaknesses in some of the 'bespoke' evaluation forms CPs designed. For example, one 'start' form asked questions about the intended objectives of the project but did not ask if they were achieved in a parallel form issued at the end. Some asked respondents to give impressions based on very limited observations, such as teachers' willingness to use creative learning techniques. Or:

Do you feel the Creative Partnerships project has had an impact on the community's perception of the school and its students?

ie the respondent's perception of another collective perception.

A commonly used question was particularly poorly constructed:

How would you describe your understanding of the importance of culture and creativity in education?

In a questionnaire this should form two questions, one about understanding of culture and the other about creativity but it also assumes that the respondent has an acceptable understanding of the two concepts. It is, therefore, complex and ambiguous and could not yield reliable information. This question is actually derived from the DCMS Policy and Delivery Agreement for Creative Partnerships (section 6.3) and was clearly not intended as a research question. Yet we read three evaluation pro-formas where it had been directly used as such. These weaknesses in evaluation questions suggest to us that CP staff would benefit from a training programme in the design of evaluation tools and research methods

Doing: CPs using the Evaluation toolkit and entering data on the XA online database had usually collected detailed information on projects. Two CPs appeared to have made strenuous efforts to collect all the feedback from every teacher and creative practitioner involved in its ten sample projects.

CPs not using the Toolkit submitted a wide range of material, from promotional DVDs to academic articles as part of the evaluation sample. For example, one evaluation was written and structured as an MA dissertation and another as an academic journal article. However, neither made mention of CP at all, and no attempt was made to relate the objectives of CP to the learning described in the study. In nearly every case, material at the margins of this range was not useful as evaluation.

A common element of reports written by CP staff such as programmers or creative agents is the use of assertions which are not backed up by evidence or explanation. For example:

The project had a significant impact on students' self-confidence and motivation.

and

The project had a significant impact on students' ability to think and act creatively.

Neither of these reports present evidence for these assertions or define what 'significant' means in this context.

On the other hand, some CPs not using the Toolkit had introduced helpful methods of obtaining feedback from pupils and young people. One CP collected feedback from 227 primary pupils who completed a questionnaire especially designed for their age group, and 196 secondary young people completed a similar one designed for them. In another CP, pupils themselves conducted evaluations:

Case Study - Using the questions given as guidance in the Toolkit, year five pupils interviewed year two pupils about a CP project. Year five pupils commented that the first question (regarding values and contribution of ideas) was too difficult for the year twos to answer and suggested simpler questions. Some of the year five pupils adapted the questions during the interviews to help the year twos answer. One year five pupil devised a sliding scale of numbers. Another pupil worked out that she was asking leading questions of the year twos in order to help them answer, 'They just repeated the things I said...'. Many of the year fives commented that they needed to know the younger pupils better to get better answers, 'We should work together more on the project...'. This pupil evaluation group were thoughtful and cleverly adapted questions and approaches in order to help the year two pupils' give meaningful feedback on the project.

Checking: Three CPs had commissioned independent organisations or individuals to survey and analyse the quality of their evaluations. All three surveys provided helpful guidance to the commissioning CPs. We also received evidence from a CP which had decided to outsource its entire project interviewing to independent consultants.

The volume of data CPs submitted on the XA database varied widely. In one notable CP 20 projects were completed, by at least two respondents in each. These individuals completed all four forms, and answered almost every question in the first two forms. One CP not using the XA database had also made every effort to ensure full completion of evaluations:

Case Study - All the project evaluations in a particular CP were compiled by a single individual who appeared to have worked hard to ensure that teachers and creative practitioners had responded to all questions both before and after each of the ten projects. The reports were helpfully colour coded to indicate whether the comment was from a teacher, pupil or creative practitioner. The reports were balanced in so far as, while most evaluations indicated the significant value of the project, one notable report contained very low grades, principally due to poor teacher-artist relationships.

However, lot of the information required in the XA database was missing. For example, respondents in most CPs omitted to provide ratings on the impact on the whole school; frequently respondents failed to grade the input to a project, and commonly teachers and

creative practitioners left out grades about the impact on themselves. Clearly the more information omitted from evaluation forms the less reliable one can claim the data to be. Despite the two instances of good practice above, we found no direct evidence that CP staff were routinely checking and making efforts to capture complete sets of data from teachers and creative practitioners. The implication is that some CPs are less robust than others in requiring completed evaluations.

At least one CP ensured that the evaluations were never conducted by any person involved in the project in question. This made it more likely that evaluations would be objective.

We agree with a conclusion in one independent CP report which recognised that:

Creative Agents need both the time and also the understanding that evaluation is as critical to their function as the project development and planning is.

Responding: As well as conducting evaluations, one CP undertook quarterly monitoring, asking schools for SEFs and seeking other evidence such as OFSTED reports from them. They go on to plan new projects using the lessons learnt from both previous projects and the quarterly monitoring. The CPs we visited usually held meetings with their creative agents and programmers at the end of an academic year to review the lessons learnt from project evaluations and to plan changes for the succeeding year. Often CPs staged review and evaluation events with representatives from each of their schools. In one CP there was no review event which brought together staff across schools.

Recommendations We recommend that CP Regional Offices consider:

1. ensuring that those CP staff designing and conducting evaluations undertake appropriate training in evaluation;
2. reviewing contract arrangements in order to commit schools to full participation in evaluation, especially at the leadership level;
3. commissioning occasional independent surveys of the quality of evaluations across their CP projects.

5 - The Role of Schools in the Evaluation of CP

Schools contribute to evaluation through their hosting and management of CP projects, through the contribution of leadership teams and other staff to CP. Schools sometimes

enhance evaluation with information and analysis drawn from their own data banks. In this section we summarise the evidence about CP evaluation at the whole-school level.

CP has introduced and distributed a school self-evaluation instrument, the Creativity Self-Evaluation Form (C-SEF), which is designed as a parallel to the Self Evaluation Forms (SEF) which OFSTED requires schools to produce. Evidence in the C-SEFs we received was often descriptive and lacking in exemplars. One, for example, referred to a, 'wealth of opportunity to be creative' through links with CP without detailed reference to evaluation processes. One school was clearly aware of this:

the school has a philosophy to adopt a creative approach to teaching but we need to implement how this is evaluated and the criteria we use.

By contrast another C-SEF adopted a very specific approach:

Case Study: - One C-SEF provided case studies as supporting evidence about the impact of CP. For example, one case study describes a student in a team of young people making a film about teenage pregnancy. The project helped him to engage at a time when he was finding it difficult to stay in school. The skills he developed in the project appeared to increase his confidence, and this was evidenced when he received a national award for citizenship. Attending a national conference to receive the award gave him a great sense of pride and achievement. His work is now used in other schools and by the health service to help promote and discuss the issues.

It was, however, surprising that only a handful of C-SEFs were submitted as part of the supporting evidence from the six CPs in this year's sample.

A significant omission in evaluation data was that only two CPs using the XA system had, across all projects, gathered ratings on the value of the project to the school. Beyond these two CPs, virtually no respondents using the Toolkit had entered any ratings on impact in the school.. There may be a range of reasons for this. These include:

- some respondents may have felt that they had too little information to make a judgement;
- CPs may have felt it wasn't their place to make a judgement about a school;
- some respondents may have felt that the project wasn't relevant to the wider school community;
- some may have felt that it wasn't valid simply to make a stab at assessing value to the school;

- some may have felt that a judgment could only be made after a longer period of time.

However, the lack of a reliable body of quantitative information about the impact on schools is an issue which we believe should be addressed as a priority in the current academic year.

Most comments indicated that school leadership teams are supportive of CP evaluations. Occasionally comments indicate that CP has changed the opinions of school leadership teams:

The teaching and learning outcomes of undertaking this process have been more profound than we initially anticipated. The most fundamental has been the realisation from the school leadership / teaching perspective of how we underestimate the potential of children. This process has raised our expectations of what we can expect children to achieve.

However, we found around a dozen accounts in which respondents, usually teachers, claimed that leadership teams were disinterested in CP projects or occasionally actively obstructive. In one example, a head teacher told an evaluator that the school improvement plan was confidential to the governors so could not be released. Another report described a head refusing to participate in evaluation at all. Sometimes prose comments imply that staff closely involved in a project miss an opportunity to disseminate the lessons learnt:

an opportunity was lost in not having the experience of the residency brought into the consciousness of the whole school and staff body.

A few CPs adopt strict systems which contract and commit schools to participation in evaluation. These sometimes require schools to make a modest financial contribution to a project, and contain explicit requirements for leadership teams to participate in evaluations.

Case Study - In order for schools to receive CP Projects and funding, XXX CP requires schools to submit a detailed application form and a 2-page application. Schools must submit their SEF, a SIP showing creativity and creative learning as an outcome, they must nominate a creative 'champion' with a teaching and learning responsibility allowance for CP, and a governor responsible for CP, head teacher participation in training and a £500 contribution to match fund the £4000 from the CP. Schools must also complete a C-SEF.

We encountered half a dozen examples of schools using their existing data to undertake a useful analysis of CP's impact. One school had looked at incidents of challenging behaviour between cohorts doing and not doing CP. This had yielded favourable information. Two others had looked at standards achieved over time comparing cohorts involved and not involved in CP. This was an inconclusive but useful exercise. Some evaluation comments encourage the schools concerned to decide, at the project planning stage, on the outcome data they will draw on:

Recommend that the measures of success are clearly defined so that, where necessary, the significant comparative data can start to be collected now

One CP had designed a questionnaire for governors, parents, support staff and heads, using the same 4 point scale as in the Evaluation Toolkit. This is particularly useful for finding out if CP projects have impact and profile beyond just pupils, creative practitioners and teachers.

Two CPs had commissioned reports on how the model of a creative school might be characterised:

Case Study - CP XXX commissioned a report from a University which outlines a 4-stage 'Progressive Maturity' model to describe schools' involvement with CP. At the initial stage the school 'learns the ropes.' The second stage involves 'projectism' where the repetition of separate externally-funded projects, which may be successful in themselves, does not move the CP agenda on. At the third 'integration' stage, the principles of CP are embedded in the school's core curriculum as a self-supported initiative. The fourth stage is described as 'dissolution,' when CP activities become so integrated into the school's learning and teaching as to lose the CP branding.

Another CP had devised a self-analysis tool for schools to assess their creativity. This CP defines a creative school as a place where certain factors facilitate a rich and varied experience of creative learning. School staff self-assess elements of institutional creativity such the school's ethos, environment, curriculum, teaching, pupil involvement, and staff development against descriptors, in three developmental categories of school: *beginning*, *progressing* and *exemplary*. The planning tool commentary suggests that staff in the school should spend about an hour each year on the tool. Clearly the results of this process could provide CPs with direct evidence of the impact of CP on schools.

We sampled supporting evidence in six CPs and some of these submitted school documents such as SEFs. The audit team read around a dozen of these, which provided useful evidence of the profile and impact of CP on a particular school. This evidence highlighted a small group of projects which aligned their objectives to other common school priorities such as Every Child Matters, emotional well being, and extended schools. Whilst dovetailing CP objectives with other school priorities might be practical, in these SEFs the profile of CP objectives was suppressed by the school's own priorities. For example one school CP-funded project purported to investigate 'happiness.' In another CP a research project aimed to 'increase family involvement'. Such projects are unlikely to yield data which is accurate and reliable and the link with CP objectives is unclear. In a third CP an independent evaluator pointed out that a high profile project had omitted to acknowledge CP funding or objectives at all. So, in principle, it makes sense if CP projects match and align with other school imperatives such as ECM and extended schools, but in practice the distinctive objectives of CP and the extent to which the objectives are met through CP can be compromised and diluted by these other priorities.

In contrast, another SEF claimed that its CP-funded project was investigating the links between creativity and improved language skills / increased vocabulary / using language for thought. Here, the relationship with CP objectives is much more explicit. This was, however, the exception. In the small sample of around a dozen SEFs we saw, there is very little other comment about CP, other than passing references to its beneficial influence. These SEFs tended to claim that CP provided enriched educational opportunities but presented no direct evidence of evaluating impact on standards.

This implies that CPs could intervene more often to sharpen the influence of CP on SEFs. Schools clearly have multiple priorities and might understandably resist the imposition of extra burdens. So the facilitating role of the CP staff is vital in assisting schools both to articulate the aims and outcomes of creative learning, and to conduct systematic self-evaluation. For example, perhaps a school can show whether pupils in years 7-9, involved in CP, present with fewer than average incidents of challenging behaviour than the weekly average across the lower school. Perhaps a school could compare absenteeism rates between the non- CP and the CP cohorts in upper primaries.

Recommendations We recommend that School managers could consider, with the assistance of CP staff:

- interrogating their existing data for evidence of the impact of CP;
- devoting explicit sections of their SEFs to creativity;

- making more use of C-SEfs.

6 - The Role of Teachers and Creative Practitioners in Evaluation

Teachers and creative practitioners run CP projects and provide the direct feedback which comprises project evaluations. In this section we survey the contribution they made to evaluation in 2006/7.

Nearly all teachers participate actively and positively in evaluation. During our visits, teachers gave a generous amount of time to evaluation and demonstrated a commitment to the work of Creative Partnerships. Nevertheless, both creative practitioners and teachers using the XA system almost never submitted a full set of data. Most commonly they failed to rate the value to schools on the system (though they offered prose comments). Some teachers left out the value to themselves, and there were frequent examples of teachers omitting to rate the input stage of a project. Clearly the less complete the data entry, the less reliable the evaluation is.

Creative practitioners rarely reported a change in their practice as artists, and often appeared to discount development of their practice as a by-product of working in CP. Correspondingly, their scores on the XA database rarely indicated a positive development in their risk taking. On the other hand creative practitioners frequently reported that they had learnt a lot about how to work in schools as a result of CP, and recognised themselves more as learners:

Yes, in my ability to do teaching side. Our practice, happy with it but being able to get things across, more confident to try new things. Boost to confidence taking largish groups of young children around school. Probably not affected personal confidence.
YP: Yes, a lot. Some quite quiet and came out of themselves, particularly 1st session when meeting us - by end all putting hands up.

Teachers commenting on their creative partners tended to corroborate this:

Artists felt that they have not learned new skills as a result of the project as the skills required were already in situ. They learned more about schools, not to have preconceptions and young people; they learned to expect to be surprised. The process of the project was a key element and they had to be strict not to rush it and skip bits out.
The practitioner stated that it was useful for her to engage with members of staff she hadn't worked with before and that every time she works with this school it increases

her capacity to work with the education sector, as she gets real insight into the politics etc within a school and the pressures they are under.

Well over three-quarters of prose comments in evaluations record the positive benefits of CP projects:

Staff here have really opened up their teaching practice and committed themselves 100% into the process of working in partnership and in a way that is uncomfortable to them. They have embraced the Creative Partnerships approach because they have wanted the change. The ball is rolling now and with more time the creative planning model can be tested and put into place. More staff are becoming aware and so more time is needed now to spread out into the school.

The ideas have massive value. It has totally changed the way I teach. I now have the opportunity to be outstanding. I'm now engaging a greater percentage of my class and motivating them to want more and ask for knowledge. At first the pupils may think 'what's she doing?' but when they see the results, they get it. These creative techniques build generic skills too, not just science

Creative Practitioners demonstrated their commitment to CP in most of their comments, and reported positive partnerships with teachers. Their commonest criticism was that projects did not permeate the wider school community, usually because, they claimed, school leadership teams were disinterested. This evidence was supported by our visits, during which CP staff claimed that committed heads and deputies was a key factor in successful projects. Creative Practitioners were usually reluctant to predict or judge impact on schools, presumably for the reasons we cite in section 5 above.

In isolated cases complete prose evaluations reports were written by the creative practitioners who ran the projects in schools. In practice these reports tended to be less robust and self-critical.

Recommendations: We recommend that teachers and creative practitioners:

1. attempt to provide a full picture of their opinions by answering all evaluation questions;
2. creative practitioners should consider and plan ways in which CP projects could develop their artistic practice.

7 - Conclusions

CPs have accumulated a wealth of predominantly positive testimony and feedback about the impact of CP projects on pupils and young people, schools, teachers and creative practitioners. At present there is no more viable and valid system for evaluating CP than the CP Evaluation Toolkit. There was evidence that CPs had used ideas which could refine the Toolkit, but none of them had a system which could potentially replace the Toolkit with a more rigorous, fit for purpose alternative.

In evidence submitted to us, it is clear that CP across the country has not accumulated sufficient aggregated and comparable information to be reliable. This is because the majority of Toolkit respondents have left out questions, and that CP staff have interpreted the questions and administered evaluations in widely differing ways. In the absence of large-scale data collected by consistent methods, no reliable and legitimate comparisons across CPs, schools, regions could be undertaken and nothing generalisable could be claimed about the impact of CP. Nor could policy decisions be drawn from secure evidence. This applies as much to projects across individual CPs as to the national initiative.

The Evaluation Toolkit, however, has the explicit as well as implicit support of the majority of CPs and, if refined through an open process of consultation, has the potential to secure widespread support. It could provide a valid and effective method of gathering reliable data. Using a mutually agreed and well supported method of evaluation would avoid duplication of effort across the CPs, which currently devote significant time to 'bespoke' systems. A programme of general training in conducting evaluations and research methods would help the cadre of CP programmers and creative agents conducting evaluations, many of whom requested such training. This training may also help to secure a consensus about the Toolkit so that more consistent and better understood processes can be used.

The XA online database is potentially an efficient means of gathering respondent feedback but it currently lacks an accessible and clear 'front end;' a composite report format which could convey meaningful information in a digestible form for schools and other stakeholders. This may necessitate a re-design and/or another software platform. We saw three or four examples of project evaluations conveyed by clear graphical representations.

We agreed with the CP staff who felt the Toolkit did not give sufficient profile to the 'pupil voice.' It will be useful to disseminate and experiment further with the methods some CPs used to gather pupil feedback. In particular the use of pupil evaluators as a means of adding an extra dimension to feedback seems to us to be worthy of wider trials.

CPs could usefully review and refine the extent to which they ensure CP funding is contingent on schools participating fully in evaluation and respondents filling in answers to all the Toolkit questions. There are likely to be real benefits to a system in which CP staff routinely support schools in fulfilling all the responsibilities they committed to CP. C-SEFs and SEFs could usefully contain specific evidence of the impact of CP projects, and schools should more regularly be encouraged to analyse its existing data for the same reason. Because of the manifold pressures on schools, CP programmers or creative agents will need to support and facilitate this work by liaising with school staff and leadership teams.

In summary, we judge that there is a wealth of information on the positive effects of CP, and sufficient feedback to help staff refine CP at a local level year on year. With the introduction of more consistent processes, we believe that it is perfectly possible, as we move through the 2007/8 academic year, for CP to take action locally and at the National Office, to capture this information more reliably.

Appendix A The aide memoire for visits

CP AUDIT: interview of CP managers

Note: purpose of audit:

- To evaluate the self-evaluation process: are reports rigorous, fit for purpose, consistent and comparable?
- Validate and disseminate regional strengths and good practice
- Synthesise and interrogate common CP issues across the country
- Challenge and support CPs in their work

CP:

Interviewees:

Brief description of CP e.g. management structure, number of employees, schools involved (core and other)

EVALUATION PROCESS

Brief description of process:

(ie Is there a means of contracting schools to deliver evaluation as part of the project? Is there a means of selecting one project as having more impact than another, or refining projects so they have optimum impact?)

If not using - or using variant of – CP Evaluation Toolkit model, why?

How are schools and CPs prepared for evaluation?

Feedback from schools on user friendliness, time taken, value?

Usefulness of CP data base?

(ie the accessibility and user interface with XA system and *Athens* a parallel system at the Arts Co.)

Effectiveness of evaluation process thus far? (strengths and weaknesses)

In light of experience, any changes likely?

EVALUATION OUTCOMES

Impact on school improvement? Evidence?

Most critical factors in successful projects?

CP SELF EVALUATION FORM

CP and school:

Date/time:

Attendees' roles (e.g. CP co-ordinator, CP agent, head, class-teacher):

Project focus and objectives:

OUTCOMES (in relation to objectives) for:

Pupils

Teachers

Artists

School

Community

Other, unexpected, outcomes:

Evidence of outcomes:

ORGANISATION & MANAGEMENT of partnership:

Training for teachers and CPs:

Issues:

CP AUDIT: Aide memoire for scrutiny of supporting evidence.

Note: purpose of audit:

- To evaluate the self-evaluation process: are reports rigorous, fit for purpose, consistent and comparable?
- Validate and disseminate regional strengths and good practice
- Synthesise and interrogate common CP issues across the country
- Challenge and support CPs in their work

Are there 10 evaluations in document form? Yes / no
Do they follow the evaluation Toolkit (XA) format? Yes / no

Is there any evidence that reports are quality controlled?
(is there any form of document critiquing the evaluations, pointing out good ones and suggesting improvements)

Are there *hybrid* tools, ie variations on the CP Evaluation Toolkit? Yes / no

(And if so are they clear, unambiguous, are terms defined, are the meanings of questions interpreted?)

Have the reports been signed off to confirm their accuracy?
(ie has a teacher or creative practitioner made a direct intervention – signature or otherwise - to confirm that they can vouch for the accuracy of the report?)

Are there other evaluation tools?
(such as C-SEFs, SEFs – are these rigorous?⁵)

Are there other planning tools?
(Such as school improvement plans, local authority improvement plans which demonstrate that CP objectives are permeating school improvement and therefore impact in some ways)

Are there examples of external evaluation and are these rigorous and robust⁶?
(eg evaluation commissioned from a university, college, consultancy, freelancer external to the CP)

⁵ By *rigorous* we mean balanced, containing negative as well as positive points.

⁶ By *robust* we mean report which can stand up – in some degree – to challenge. So if an evaluation claims ‘significant gains in confidence,’ it can define what ‘significant’ counts as.

Appendix B - an explanation of terms and concepts

Our principal task was to find out the extent to which CPs evaluation is rigorous, fit for purpose, consistent, comparable (see section 1 above). In this section we outline how we interpret these terms in the context of CP. We believe this explanation is essential so that CP staff across the country reach a common understanding of the criteria by which we make the judgements and recommendations in this report.

a) By rigorous we mean to what extent do Creative Partnerships strive to collect all of the data it asks for? How common is it for CPs to make efforts to assemble complete sets of evaluation data from schools and creative practitioners? Clearly the more complete the data, the more comparisons can be made across schools, projects and CPs. Rigour also means the extent to which the evaluation reports are balanced, by both negative and positive points. Clearly it is unlikely that every project across the country, or even across a CP, is going to be wholly successful or unsuccessful.

The term rigour also denotes for us whether the evaluative evidence is robust, ie does the evidence stand up to challenge. For example, if an evaluation claims:

CP has made a significant improvement in pupil problem solving.

we can expect to get a clear answer to the question

What do you mean by significant?

b) By fit for purpose we mean are the evaluations and the methods they use valid? That is, are the evaluations designed to tell all of us, with vested interests in CP, what it is important to know? Is it using the most logical and appropriate means to find out what we want to find out. Taking our cue from the LC Associates Report and the Evaluation Toolkit booklet we made the assumption that evaluations and the strategies behind them should be designed broadly to find out whether projects were addressing the objectives of CP, ie:

1. the creativity of young people, raising their aspirations and achievements;

2. the skills of teachers and their ability to work with creative practitioners;
3. schools' approaches to culture, creativity and partnership working; and
4. the skills, capacity and sustainability of the creative industries.

The overarching dimension to this is that CP is accountable to the public, to the education system, to parents to pupils. A fit for purpose system of evaluation ought to be accessible to such stakeholders so people can make their own mind up about how successful CP is. In other words a fit for purpose system can provide evaluation data to end users and the interested public.

We understand consistency to refer to the extent to which evaluations are conducted in the same way. The more CPs administer the same sort of evaluations in the same way the more one can potentially derive large scale information from the aggregated results. The Evaluation Toolkit is explicitly designed to achieve this consistency. In this audit, we analysed the extent to which the Toolkit is in use.

d) By reliability we mean the more the same sort of data is aggregated, the more reliable any findings derived from the data can be. If results from evaluation are to be in any sense convincing, the questions should be asked and responses treated – as far as possible - in a consistent way across the country. Whilst the CP Evaluation Toolkit's set of attitudinal questions are not like a litmus test or a blood count, the reliability of the results is dependent on the extent to which the test is administered consistently by trained people.

In summary, therefore, this audit attempts to gauge the extent to which large scale data is being collected by consistent methods, and yielding convincing, unambiguous, balanced information which tells stakeholders whether CP is meeting its objectives. Clearly CP evaluation can never be an exact science any more than any form of analysis of people and organizations. But the more we have a system the better we will know whether CP is making a difference.