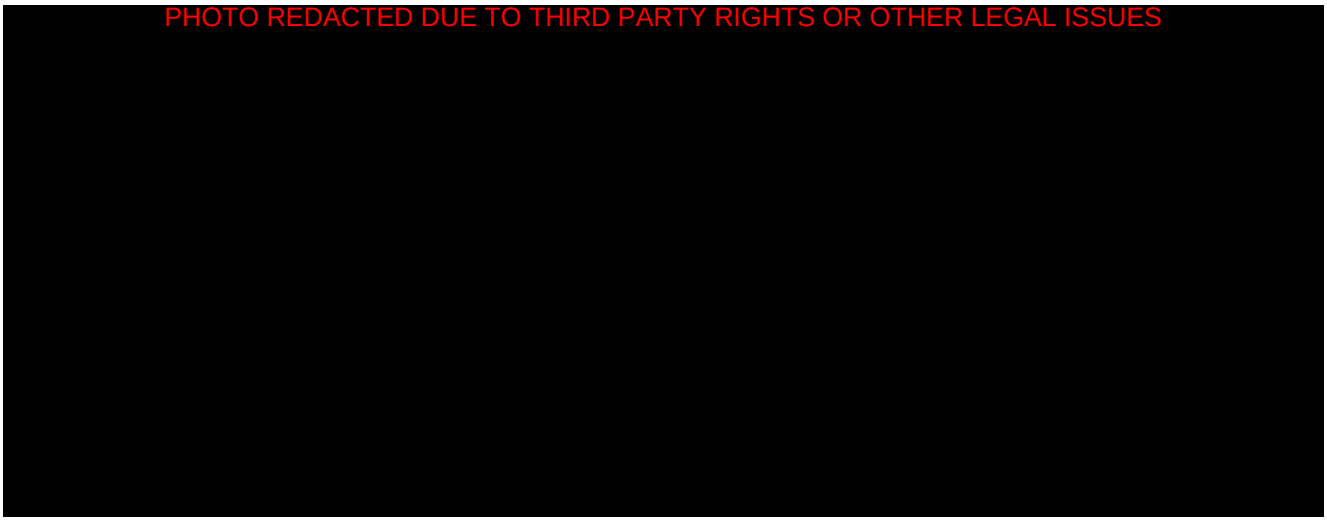

Test Bed project evaluation additional maturity model analyses
Technical Report 26/07

Jean Underwood and Gayle Dillon Quantitative Evaluation Team
Psychology Division, Nottingham Trent University

Contents

Points for policy makers

The analyses below show:

1. maturing systems and school performance go hand in hand
2. the six maturity models form a complex web of relationships and level of disparity between institutions on any one model have fluctuated over the lifetime of the project
3. findings from the second year of the project support the concept of the technology dip reported throughout the evaluation. Embedding systems is not cost free – the technology dip apparent in year two of the project is exemplified by a severing of the link between maturity and performance. A link that is re-established in year three, when the learning experience, not the technology is once again the prime focus.
4. Models from the final year of the project show a plateauing of progress, which is reflected in low variation between the schools.

Readers note

The analyses reported here represent data from each of the four years of the Test Bed project (2003-2006).

Preparation of the data sets:

Maturity model data:

These analyses use a merged data set collated from each institution's self-assessment and the evaluation team assessment based on data collected during each year of the project. For each year, a final data set was created per institution by taking an average score of the institutions' and evaluation teams' assessment on each of the dimensions.

School data:

The school performance data was the same as that used in the process of benchmarking. The average point score for each institution calculated by the DfES from the national test results was the starting point for the schools' data. In order to be able to run the analyses with all institutions entered

simultaneously, it was necessary to standardise the performance data to be used as the dependent variable. A single score for each institution was calculated using the mean of the available test data for each institution. Thus, for those primary schools for which Key Stage 1 and Key Stage 2 data were available, the mean score of the two average point scores for both tests was calculated and used in the analyses. In the case of the secondary schools, unlike in previous years, only the GCSE score was used as a result of a delay in publishing Key Stage 3 data and changes made to the A-level performance score calculations. The resulting single score for each institution was then converted to a z score in order to standardise the data and make the data directly comparable for each sector.

Cluster analysis

Identification of school profiles for each of the four years of the project was examined using cluster analyses. Cluster analysis, or numerical taxonomy, does not provide an indication of causation; analyses to identify causal relationships are presented in the regression analyses below. Rather, it enables the researcher to structure samples in terms of homogenous subgroups. To identify the different profiles a hierarchical cluster analysis was performed on the mean scores achieved by each school on the six models and their performance on the relevant national tests (entered as a z score to enable a standardised score to be entered). Analyses were run with both primary and secondary sectors entered; given the small sample sizes it is not possible to run the cluster analyses with these sectors individually. Table 1 displays the cluster solutions and mean scores for each project year.

Cluster analysis findings with year one data (2003)

A two cluster solution was identified for the year one maturity model mean scores and national performance data. The two clusters were characterised as those schools (n=14) belonging to a ‘lower maturity/lower performance’ cluster and those schools (n=12) belonging to a ‘higher maturity/higher performance’ cluster.

Cluster analysis findings with year two data (2004)

A three cluster solution was identified with the year two data, characterised as 'low' (n=2), 'medium' (n=9) and 'high maturity' (n=15) clusters. It is interesting to note the fluctuations in performance data for these clusters; with those schools who fit the 'medium maturity' profile demonstrating the highest performance scores and schools in the 'low' and 'high' maturity profiles performing equally poorer. An explanation of these findings is offered in a later section.

Cluster analysis findings with year three data (2005)

A two cluster solution was identified for the year three maturity model mean scores and national performance data. The two clusters were characterised in the same way as for the year one data, with schools demonstrating either a 'lower maturity' (n=15) profile or a 'higher maturity' (n=11) profile.

Cluster analysis findings with year four data (2006)

With the year four data, a two cluster solution was selected given the presence of isolates in other cluster solutions. The two clusters are characterised by 'lower maturity' (n=15) and 'higher maturity' (n=11).

Table 1: Means (and standard deviations) for each model and standardised z score of performance data for each year

	Year one - 2003		Year two - 2004			Year three - 2005		Year four - 2006	
	Cluster 1 Lower maturity/lower performance N=14	**Cluster 2** Higher maturity/higher performance N=12	**Cluster 1** Low maturity N=2	**Cluster 2** Medium maturity N=9	**Cluster 3** High maturity N=15	**Cluster 1** Lower maturity/lower performance N=15	**Cluster 2** Higher maturity/higher performance N=11	**Cluster 1** Low maturity N=15	**Cluster 2** High maturity N=11
Maturity model 1 mean Technological	2.49 (3.44)	3.06 (0.31)	2.75 (0.45)	3.75 (0.15)	3.85 (0.29)	4.19 (0.31)	4.37 (0.20)	4.24 (0.27)	4.75 (0.18)
Maturity model 2 mean Curriculum	1.94 (0.17)	2.49 (0.37)	2.59 (0.23)	2.87 (0.26)	3.11 (0.23)	3.63 (0.30)	3.91 (0.41)	3.56 (0.32)	4.29 (0.29)
Maturity model 3 mean Leadership and management	2.35 (0.17)	2.95 (0.28)	2.75 (0.43)	3.28 (0.25)	3.64 (0.36)	4.01 (4.73)	4.28 (0.57)	3.96 (0.44)	4.65 (0.23)
Maturity model 4 mean Work force	2.37 (0.26)	2.93 (0.24)	3.03 (0.60)	3.69 (0.32)	3.89 (0.26)	4.16 (0.30)	4.43 (0.24)	4.14 (0.29	4.63 (0.26)

Maturity model 5 mean Linkage 1 - Intra	1.83 (0.39)	2.49 (0.37)	2.58 (0.12)	2.92 (0.15)	3.27 (0.37)	3.70 (0.37)	4.16 (0.37)	3.80 (0.43)	4.27 (0.34)
Maturity Model 6 Mean Linkage 2 – external	1.54 (0.29)	2.08 (0.48)	1.58 (0.04)	2.30 (0.60)	2.85 (0.37)	3.13 (0.47)	3.18 (0.67)	3.14 (0.59)	3.72 (0.33)
Performance z score	-0.07 (0.58)	0.36 (0.94)	-0.38 (0.60)	1.06 (0.28)	-0.42 (0.78)	-0.63 (0.51)	1.01 (0.29)	0.34 (1.22)	-0.47 (0.03)

Table 1 shows that over the four years of the project, schools had demonstrated an increasing maturity across all six models. It also demonstrates that the clusters are ordered and distinct. In year one schools in the low maturity/low performance cluster consistently scored lower overall on all maturity models and national tests than schools in the high maturity/high performance cluster. In year two, as previously mentioned, schools whose profiles demonstrated a mid level attainment on the maturity models were also those schools demonstrating the highest performance scores. The two clusters of schools with the lowest and highest levels of maturity were also those with the lowest performance data. This finding was followed up by discriminant function analyses which suggested that the driving force behind the cluster profiles in year two was each school's performance on the tests. The importance of performance data in the second year in differentiating between the clusters of schools is attributed to the technology dip previously reported. The lack of consistency in performance between the high, medium and low e-mature clusters suggests that in the second year of the project, when schools were in the initial embedding phase, is further evidence of the technology dip. As schools efforts focused on establishing working systems there was a closing down of innovative teaching. By the third year of the project, however, the schools demonstrated a recovery, again settling in distinct two distinct clusters of low maturity/low performance and high maturity/high performance. In your four, this pattern was repeated, although performance was found to have stabilised between years three and four, which discriminant function analyses revealed to now drive little differentiation between the two clusters. Put simply, following the technology dip in year two and the performance driven difference between the clusters, all schools recovered and stabilised beyond their initial starting points and the driving force of the differences between schools became attributable to levels of maturity by the end of the project. Whereas performance outcomes drove the cluster differences initially, by the end of the project when achievement had become consistently high across the Test Bed schools, levels of e-maturity became the important differentiator. This was particularly the case for performance on model two (curriculum maturity) followed by model one

(technological maturity), model three (leadership/management) and model four (workforce maturity).

Regression analyses

As in previous years, a series of regression analyses were conducted for each institution which enabled us to establish which, if any, of the dimensions within each of the models predicts performance outcomes. This was to identify causal relationships, rather than group structures that were discussed in relationship to the cluster analyses. These analyses also enable us to compare the models globally and ask which of the six models is able to best predict performance outcomes from the third year of the project.

Three levels of analysis were conducted on these data. The three levels represent the level at which the data were broken down and can be defined as follows:

1. Macro level analyses: The macro level analyses were performed as global analyses in order to establish the predictive power of scores on each of the six maturity models overall on performance outcome. In order to conduct these analyses, the mean score for each institution was calculated for each model, resulting in the generation of six data points for each institution.
2. Meso level analyses: The meso level analyses were performed on each independent model in turn, resulting in one analysis per model. For each model the number of independent variables was equal to the total number of dimensions contained within that particular model. For example, maturity model one: Technological maturity, has seven dimensions resulting in seven independent variables being entered into this analysis.
3. Micro level analyses: The micro level analyses were performed on those models that contained subsections, namely models two, three and four, each of which is made up of three subsections. The micro analyses were conducted such that each subsection constituted its

own analysis, resulting in either three or four analyses being run per overall model. For example, model two: Curriculum maturity is made up of the subsections 'institutional', 'teacher' and 'pupil/learner', resulting in one analysis being conducted for each of the 'institutional', 'teacher' and 'pupil/leaner' levels.

Findings from year four:

Macro level analyses

At the macro level (mean scores from the six maturity models) no significant model was found. Unlike in 2005 where models three (leadership and management) and five (internal linkage) were found to predict performance, the plateauing of maturity by year four meant that there was less collective variation on scores between the models and thus, reduced predictive power.

Meso level analyses

At the meso level, model three (leadership and management maturity) was found to be a significant predictor of performance in year four ($F_{(6,25)} = 4.06$, $p<0.05$), explaining 70 per cent of the variance. Dimensions two (planning; $t = -2.65$, $p<0.01$), three (policy; $t = -2.42$, $p<0.05$), and seven (analysis of attainment and progress; $t = 2.33$, $p<0.05$) were found to be the significant dimensions.

Micro level analyses

At the micro level, three models were found to be significant. The first was the institutional level of curriculum maturity model (model two) ($F_{(6,25)} = 4.06$, $p<0.05$), explaining 30 per cent of the variance. Within this model, dimension one (curriculum ICT policy) was found to be a significant predictor of success ($t = -2.36$, $p<0.05$). The second significant model was the leadership level of the leadership and management maturity model (model three) ($F_{(6,25)} = 4.40$, $p<0.05$), explaining 47 per cent of the variance. Within this model, dimension two (planning; $t = -2.28$, $p<0.05$) and dimension four (implantation; $t = -2.41$, $p<0.05$) were significant predictors).

The final significant model was that of the workforce maturity model (model four) at the technical support level ($F_{(6,25)} = 5.10$, $p<0.05$) explaining 31 per cent of the variance. There were no significant predictor variables for this model.