The likely impacts of target setting and performance rewards in local transport

Dr Greg Marsden*

Charlotte Kelly

John Nellthorp

Institute for Transport Studies

University of Leeds

Leeds

LS2 9JT

United Kingdom

*Corresponding Author

E-mail: G.R.Marsden@its.leeds.ac.uk

Tel: +44 (0)113 3435358

Fax: +44 (0)113 3435334

Abstract

All local transport authorities in England have, since 2000, been obliged to submit five year plans for local transport. The plans set out the overall strategy, key policies that will be implemented and how the strategy will be resourced. The central government now adjusts the funding allocations up or down by up to 25% based on the quality of the plans and, on an on-going basis, achievement against the targets proposed in these plans. This paper presents a theoretical and practical assessment of the impacts of these incentives on local authority performance.

The research has employed a mixed methods approach with interviews, questionnaires, the development of a game theoretic representation of the process and a laboratory experiment. The findings have been discussed with practitioners. The research suggests that the presence of performance rewards, in a scheme where authorities believe they have a reasonable chance of being rewarded, leads to authorities setting more ambitious targets. Whilst it is not certain that these targets will be met it appears that the absolute outcomes achieved are likely to be better than they otherwise would have been. Generic conclusions are drawn about the conditions under which target-based performance reward schemes will work best.

1. Introduction

Targets set out the level of performance that an organization aims to achieve for a particular activity within a given timeframe. This might be for example a commitment to reduce fatalities on urban roads by 10% over the next five years. Managing public services through the use of targets is not new (Hood, 2006). It is however, becoming increasingly widespread globally (FHWA, 2004; Moynihan, 2006; Hodgson et al., 2007 and Tuominen and Himanen, 2008) and in the field of transport. The US federal government, for example, is currently considering linking funding for new infrastructure to state level performance measures. A key difficulty with such a proposal is "developing effective performance measurements without negative unintended consequences" (NTPP, 2008, p6). A recent review of the use of targets in transport found that "there is little published evidence on the effect of targets on the performance of the transport sector." (Marsden and Bonsall, 2006, p191). This paper adds to the knowledge base and, whilst based around a case study in England, many aspects of the findings have broader applicability.

Since 2000, the Department for Transport (DfT) has required local authorities in England to prepare 5 year 'Local Transport Plans' (LTPs) in which local authorities set out the policies and expenditure required to make the integrated transport vision a reality (Kelly et al., 2006). The LTPs were required to contain targets for progress against which the plans would be assessed. This process has evolved into one in which authorities are ranked and rewarded according to their

performance. The most recent round of LTP (LTP2) forms the case study for this paper.

For LTP2 each authority is given an initial funding allocation for integrated transport measures for the period 2006/07 to 2010/11 using a formula allocation mechanism. The formula was developed based on four agreed priorities (congestion, accessibility, air quality and safety) and adapted on a needs basis (e.g. 20% of the total funding allocated is based on the road casualty figures for 1994 – 1998 (DfT, 2005)). The funding allocation was to be varied up or down by up to 25% based on the quality of the LTP2 plan, past evidence of delivery, the ambition of the targets set and, over time, achievement against these targets (see DfT, 2004 for full details).

The goal of central government in developing this system appears to be a desire to reward those authorities that appear to deliver results most cost effectively. It also has aspirations to raise the quality of planning and, through some form of competition, increase the net outcomes of its spending. Critics suggest that such processes are limiting in several ways. They may for example focus behaviours only around those aspects that can be measured, can encourage short-term decision making and create perverse incentives (Smith, 1995 and Hood, 2006).

A mixed-methods approach was adopted for the research to identify the likely outcomes of a performance-led rewards scheme in local transport planning. To

assess the underlying theoretical expectations of the funding competition a game-theoretical framework was established. Game theory can provide insights into the key drivers of outcomes in a situation where players interact. To enable a meaningful game to be constructed an initial investigation of the details of the performance regime and how local authorities were approaching the target setting task was therefore undertaken using interviews and a postal survey and this is described in Section 2. Section 3 provides an outline of the game theoretic model and its key findings. Finally, a laboratory experiment was designed to test the hypotheses generated by the game theory and evidential work and this is described and presented in Section 4. Implications of the combined findings and some conclusions are presented in Section 5.

2. The Local Transport Plan Game in Practice

Two key features of the LTP2 funding process were that:

1. The DfT stated that in total no extra funds would be available for performance (a zero sum game). The implication of this was that there would be both winners and losers and this could affect what was delivered in any authority.
2. Clear guidance was given on what had to be measured and how it would be assessed. 15 mandatory indicators were identified[1] (Table 1) and a maximum of 40 indicators was recommended (DfT, 2004). Thresholds which

[1] Some authorities were exempt from the air quality and congestion indicators

would denote satisfactory and stretching[2] performance were to be published for the 15 mandatory indicators[3] with assessment of the ambition of local indicators to be made on a case by case basis for which the authorities were to provide supporting evidence.

Insert Table 1 about here

The DfT had therefore established a system whereby the local authorities bid against each other for performance reward based on the targets they set and the apparent (and then actual) ability to deliver them. The key guidance given to local authorities relating to how they would be assessed was that targets should be based on outcomes (i.e. authorities were competing on the number of fatalities they reduce not the amount spent on casualty reduction schemes) and that targets should be "challenging but realistic" (DfT, 2004, p6).

To obtain an in-depth understanding of how local authorities were approaching target setting, interviews were conducted with a range of people from six authorities involved in LTP2 submission. This varied from those responsible for individual targets, heads of LTP, regional government officers, and consultants to those with political control (e.g. councillors). These interviews were used to help provide a perspective on the process of LTP2 target setting and to develop the

[2] For example, a satisfactory target level for slight casualties was defined as no change over recent levels whilst a stretching target was defined as a 10% reduction over recent levels.
[3] Uniform national thresholds for stretching and satisfactory performance only proved possible to set for 10 indicators due to the large variation in local circumstances

questionnaire that was sent, in March 2006, to all 82 LTP submitting authorities investigating how they chose their targets for LTP2. 31 of the 82 authorities responded (a 38% response rate) and analysis suggests that no particular type of authority was under-represented in the sample. The responses from the questionnaire were subsequently compared with how the submitted targets had actually been assessed by the DfT (DfT, 2006).

96% of authorities agreed with the statement that they were more likely to increase/maintain their future funding levels if they achieved the targets they had set. In addition 60% agreed that if they did not receive their full allocation of funding (100% formula) that they would not be able to achieve their targets in 2010/11. Authorities were therefore aware that they needed to set targets that they could meet, as this would impact on their funding levels.

The interpretation of challenging was more varied across authorities. Authorities were asked to rate the likelihood of meeting the targets they had set. The average values for all authorities are shown in Table 2.

Insert Table 2 about here

Across all authorities the average expectation of meeting the targets is 70 to 80% suggesting that authorities are indeed challenging themselves and taking some risks. Interestingly, with the exception of Killed and Serious Injury accidents (where

the differences are small), authorities are more certain of achieving targets which are assessed as stretching. This is indicative of authorities playing strategies where they set stretching targets in areas they know they will perform better in.

Authorities were asked to make an assessment of how they thought the DfT would assess their targets and how they would assess their own targets for the mandatory indicators. The authorities were marked with a 2 if they set (or were assessed as having) stretching targets and 1 if they were not stretching. A comparison of the forecast and actual assessments (DfT, 2006a; DfT, 2006b) is shown in Figure 1.

Insert Figure 1 about here

It is clear that the local authorities took a pessimistic view of how they would be assessed by the DfT. This suggests that during the LTP2 process, local authorities felt some pressure to be ambitious with their targets, an important goal of the performance regime. The actual assessments are more closely in line with the authorities' own perceptions of the degree of stretch that the targets represented, given their local circumstances. This implies that the DfT has had some success in establishing a level playing field where it assesses each target on its merit and not just according to national rules. Whilst these findings hold true at an aggregate level statistical tests of differences in assessments between the DfT and authorities

were conducted for three main groupings of authorities (metropolitan areas[4], smaller urban authorities and rural authorities). Significant differences exist in assessments on bus patronage for urban and rural authorities. Urban authorities differ from the DfT on the numbers of killed and seriously injured accidents whilst rural authorities also differed from DfT for footway condition, unclassified road condition and slight accidents (Kelly et al., 2008). This suggests that it might be easier to establish a level playing field for assessment for some indicators than others.

Only 16% of the authorities responding to the questionnaire stated that they were clear about how the DfT assessment would work. The interview evidence showed that whilst the scope of the assessment was clear[5] and the need to set and deliver "challenging but realistic" targets was understood, the exact process by which their LTP submission would be turned into a performance reward was not fully revealed to the local authorities in advance. From DfT's point of view, this decision not to reveal the full details of the assessment process could be rationalised in at least three ways:

- to avoid creating perverse incentives for authorities to focus on narrow, pre-specified indicators;
- to allow DfT the flexibility to judge local targets in relation to local conditions, including an element of subjective judgement;

[4] There are six areas classified as metropolitan in the UK (outside London) and this is too small a sample for the tests to be reliable so the contrasts in this section are reported for smaller urban and rural authorities.
[5] i.e. the set of assessment criteria, although not the details of how those criteria would be scored and weighted.

- furthermore, if the overall budget was indeed a fixed sum then it is possible to state in advance the relationship between performance and *assessment*, but not the relationship between performance and reward.

A small number of examples were found of authorities failing to set "challenging but realistic" targets. For example, some authorities chose to set targets for 0% growth in bus patronage even where the evidence suggested that a decline was the best achievable. This may have been due to the local political implications of 'aiming for decline' or due to fear of the potential funding implications. E.g. *"targets were set to at least meet the minimum criteria of the LTP guidance even if this seemed ambitious"* (LTP officer). Cases such as this were the exception rather than the rule.

We conclude that, overall, the authorities were actively engaged in setting challenging but realistic targets. They understood the relative difficulty of meeting targets in different policy areas and carefully chose the amount of stretch they put into their plans accordingly. The funding reward was important to them for delivery, but they could not be confident in advance what reward would follow from a given set of targets.

The integrated transport settlements for 2006 to 2007 and their adjustments for the best and worst performing authorities are shown in Table 3. The table shows that the system allows small authorities as well as larger authorities to perform well and

that predominantly urban, rural and authorities with a mix of both characteristics can also perform well. This implies that there is a level playing field for authorities to compete in the process. Whilst the four worst performing authorities are all comparatively small, a regression of funding adjustment against initial funding allocation finds no relationship between the two ($R^2 = 0.039$) which again suggests that the process does not reward bigger authorities (those receiving more money) nor penalize small authorities (those receiving less).

Insert Table 3 about here

3. The Theoretical Local Transport Plan Game

To inform our understanding of the likely impacts of the performance reward scheme for LTP2 a theoretical model was developed. Game theory was originally devised to study strategic interactions between players whose actions impinge on each other (Dutta, 2000) and is a "core approach to the analysis of institutional relationships" (De Palma and Lindsey, 2003, p1). Here, game theory was used to define a model of the relationships between DfT and the local authority players, and to generate hypotheses for testing using the laboratory experiment. The key characteristics of the game are presented in Sections 3.1 to 3.6 and the model results in 3.7. The characteristics are derived from the work described in Section 2. A more extensive description can be found in Nellthorp and Marsden (2009).

3.1 Set of players

The key players of the game are:

- DfT
- 82 Local Transport Plan authorities.

This simplifies the real game by excluding third parties such as bus companies and the regional tier of government, which provide important parts of the context for target setting, but do not themselves either set the targets (local authorities' role) or determine the assessment criteria and funding (DfT's role).

3.2 Motivations of the players

The motivations of the DfT were highlighted in Section 2 as getting better use of resources and encouraging better planning. The interviews indicated that the local transport authorities' main motivation in the LTP process is to maximise funding. Whilst their transport strategies as a whole serve wider goals, the LTP process offers the prospect of greater resources to support their strategies. The key trade off, therefore, that local authorities have to make is how much extra effort to put in compared with the expected reward they would receive. Under the current rules, if no additional effort is put in and a very poor LTP is produced, authorities would still receive 75% of their indicative funding allocation.

For the theoretical model, therefore, the local authority players are assumed to maximise

$$Max\left[E(V_i) - c(z_i)\right] \qquad (1)$$

where:

$E(V_i)$ is the expected financial reward to authority i, in the form of performance funding;

$c(z_i)$ is the cost of playing the game (at effort level z_i).

3.3 Set of actions the players can take

Two simplified measures of the extent to which authorities are competing are:

- planned stretch – the level of ambition of the targets (at the time of target setting);
- achievement – recent progress towards the targets that are in place.

Correspondingly, each LTP authority has two sets of actions:

- prepare the Local Transport Plan – including setting targets;
- implement and report on the Local Transport Plan.

The LTP authorities therefore have to make decisions about how ambitious to be for each indicator, which will be reflected in the strategies they choose to adopt. There was no evidence of opportunities for collusion and, whilst other funding opportunities exist, the LTP game was taken seriously and was important enough in its own right to be modeled as an independent system.

DfT's most important action within the game is to assess the plans and progress on delivery and allocate a fixed block of funding between the 82 LTP authorities, annually.

3.4 Payoffs

The payoffs to the LTP authorities will be their additional funding net of the cost of effort exerted (equation (1) above). Initially DfT indicated the range of additional funding would be +25% to -25% compared with the nominal, formulaic funding guideline (see Section 1). In fact nothing lower than -12.5% was awarded. The payoff to DfT, aside from the financial loss and effort in adminstering the process, is the gain in useful outcomes on the part of the local authorities – discussed above.

3.5 Sequence

The time sequence of the game is unusual, compared with others in the literature, in being in two distinct stages: each player plans their action first – and has that assessed by DfT – before going on to implement their action and be assessed on that as well. Consideration was given to a two-stage theoretical game, although it was concluded that this would complicate the analysis for no gain in predictive power. The interviews confirmed that the authorities also view each five year LTP as something closer to a one-shot game, subject to some adjustments during the implementation period.

3.6 Information

The key characteristic here is the one already discussed – that the authorities were given incomplete information about the way they would be assessed. This is reflected in basic mathematical structure of the game (below) where the reward is probabilistic, rather than a certainty based on certain performance.

Players were able to see each others' targets from the first LTP period and some informal benchmarking networks exist for performance comparison. However, target setting was conducted independently and the authorities did not have the opportunity to see the others' final targets before deciding their own.

3.7 Theoretical Model

A wide range of game theoretic models were considered, before settling on the rent seeking contest (Tullock, 1980). In a simple rent seeking contest with two players, *i* and *j*, and two prizes – one high (e.g. +25% in the LTP game) and one low (e.g. -25% in the LTP game), V_H and V_L, the game can be set up as follows. Player *i*'s probability of winning over player *j* is given by the ratio of *i*'s effort to total effort

$$p_i = \frac{z_i}{z_i + z_j} \qquad (2)$$

…so if *i*'s and *j*'s efforts are equal, then *i*'s probability to win the higher prize, V_H, is equal to $\frac{\alpha}{\alpha+\alpha}=\frac{1}{2}$. When i doubles her effort but the probability to win increases to only $\frac{2\alpha}{2\alpha+\alpha}=\frac{2}{3}$.

The model has a random element (*probability* to win), but can be made more deterministic by applying a higher power to the effort terms in (2), e.g. z_i^2, z_j^2.

The expected payoff for each player is their expected reward net of effort exerted (as in (1)). It is then necessary to find a Nash equilibrium solution – one that is optimal for each player given the behaviour of others. Adding additional players (up to 82 in total) and additional prizes increases the complexity of the game. Mathematical solutions are given in Nellthorp and Marsden (2009).

The key results from the theoretical analysis are:

1. The symmetric Nash equilibrium has a desirable property of maximising aggregate effort across the players. Achieving a symmetric Nash equilibrium requires a 'level playing field' between the 82 local authorities, hence DfT's success (or otherwise) in taking account of any inherent advantages or disadvantages of particular authorities plays an important role in the game.
2. Furthermore, DfT's use of assessment methods which are partly *subjective* gives them the flexibility to take these differences into account.
3. The prize structure plays a critical role in determining the strength of the incentives. Reducing the number of prizes (with a fixed prize fund) tends to

increase the competitive force between the players, provided the principal can maintain the level playing field. This is manifested as greater total effort (see Figure 2), and hence – presumably – better transport outcomes.

Figure 2 shows the sensitivity of effort exerted to a range of prize structures, varying both the number of prizes and the prize gradient (the ratio of the first prize, v_1, to the k^{th} prize, v_k). A prize fund of V = 100 is assumed. The dot indicates the prize structure in the local transport plan game as executed.

Figure 2: Effort responses in the 82 player game

Figure 2 suggests that the greatest effort will be exerted when a single prize is offered. However, if the designer choses to offer a large number of prizes, the prize gradient begins to play a major role. For example, "when there are in excess of 60 prizes, moving from a prize gradient of 2 to a prize gradient of 10 has the effect of roughly doubling the amount of effort exerted, for the same prize fund" (*Ibid*., p13).

4. Additionally, if the number of prizes exceeds 0.63*N (the number of players) = 52 in this case, then a symmetric Nash equilibrium may not be achievable.
5. However, in a game with players of mixed ability (heterogeneous players), there may be compelling reasons for having a greater number of prizes. In particular, a weaker player needs to know that they have a realistic chance of winning at

least the lowest prize, if they are to have the incentive to compete (see Blavatsky, 2004; Symanski and Valletti, 2004).

6. Optimal incentive design therefore relies on balancing these factors in a particular case. If we believed that DfT could successfully achieve a completely level playing field, it might be appropriate to recommend a 'winner-takes-all' or Challenge Fund type arrangement with a single prize. Conversely, if it is judged that the 'level playing field' is not sustainable, we might consider adding more prizes for 'weak' performance, or we might consider breaking the contest down into several separate pools of more evenly-matched players. Nellthorp and Marsden (2009) found that the loss of incentive from splitting the contest appears to be small (~1%), and such action might be attractive if the costs of administering the current assessment system are high.

Finally, the game theoretic model allows us to establish some testable hypotheses, which were taken forward to the experimental phase of the work (section 4):

H1: The effort exerted by authorities when faced with performance incentives as in the LTP game will be greater than that without performance incentives;

H2: Where the authorities are broadly homogeneous in their ability to compete for the prizes on offer, fewer prizes will encourage greater effort – for a given prize fund – than many prizes;

H3: Conversely to H2, if authorities are heterogeneous in ability, more prizes are likely to stimulate greater effort than fewer.

4. The Experimental Assessment of the Local Transport Plan

To help bridge the gap between the simplified approach of the theoretical model and the realities of decision-making in a real-world environment, an experimental approach to understanding the behaviour of local transport planners under incentives has been adopted. In particular, the laboratory experiment provides an opportunity to understand further the behavioural dynamics between competing players over the course of the game.

4.1 Experimental Description

A simulation of the Local Transport Plan system was developed. The experiment involved groups of players using the PLUTO land-use and transport software model (Bonsall, 1994). The model places the user in the role of the local authority with decision making powers on interventions commonly available to local authorities such as infrastructure investment and public transport subsidy. Targets are set for five years and policy and budgetary decisions conducted on an annual basis. The city is a stand-alone city with no transport interaction with adjacent cities. The software was modified to allow the research team to play the role of the DfT whilst players simultaneously play the role of the local authorities.

The experiment placed eight groups of five players in the Local Transport Plan game environment each with an identical city. Within each group the five players were to compete against each other – each independently developing and submitting a plan with targets. Rewards and penalties were applied to the capital

funding settlement according to the experimental condition they faced and their ranking in the competition in terms of ambition of targets and achievement against targets. The eight groups were exposed to each of three different experimental scenarios in a random order:

- E_0 – no incentives ($10m per annum irrespective of performance)
- E_1 – one prize ($20m for best performer, default $7.5m funding for others)
- E_2 – multiple prizes ($12.5m, $11.25m, $10m, $8.75m, plus the $7.5m default).

The hypotheses set out at the end of Section 3 are therefore tested as follows:

H1: True if effort in E_1 and $E_2 > E_0$

H2: Effort in $E_1 > E_2$ if players are homogenous;

H3: Effort in $E_2 > E_1$ if players are heterogeneous.

One of the main limitations of the experimental approach is of replicating the motivations of the real game players (ecological validity). The use of LTP practitioners would clearly be preferable to reduce the impact of mismatched motivations of players. This was impractical however so undergraduate students from the faculties of engineering, mathematics, sciences, computing and environment with no previous experience of transport planning were used. A further challenge is the establishment of the relative capabilities of the players (homogeneity). The effect and implications of this are discussed in Section 4.2.

Pre-experimental familiarization with the software and the task was provided (Funke (2001)). For each experiment participants visited the computing laboratory for a maximum of three hours to set their targets (of which seven were required) for the experimental scenario they were presented with. The funding rules were signaled clearly to the players. A questionnaire on the effort exerted setting targets was completed using the NASA Task Load Index scales (Hart and Staveland, 1998). They returned on a subsequent day to receive their initial settlement (based on target ambition) and then to play the five year strategy through the software simultaneously with the competing players. One hour was allowed for each simulated year of play which proved generous. The ranking of players was not announced in experimental scenario E_0 until after year 5, was only indicated through the settlement to the best player in E_1 and was presented to all players in E_2 as this is implicit in the five settlements. Participants were asked to complete a questionnaire on effort exerted during the task similar to the target setting exit questionnaire. They were also asked to complete a more detailed questionnaire on the whole experiment after the last session.

Participants were paid a fee based on attendance, not performance. The decision not to offer a performance-based fee in the experiment mirrors the lack of a direct financial reward to LTP staff in the real game – even if their authorities gain financially. Although contentious, there is a literature which supports this approach (Wickham, 2007; Beattie and Loomes, 1997).

The key metric from the theoretical model is ‘effort’. Funke (2001) suggests that both results and process oriented measures should be taken. Table 4 shows the measures adopted.

Insert Table 4 about here

4.2 Experimental Validity

We begin by assessing the external validity of the game through responses to the end of experiment questionnaire shown in Table 5.

Insert Table 5 about here

The results suggest that the experiment reproduced the key features of the LTP system. For example, participants understood the relationship between stretching targets, risk and performance rewards. The findings are less conclusive on performance penalties where it appears that the penalties did not have as large an impact on participants’ ability to achieve their plans. This may be, as in the real game, due to the ability to offset losses through other measures (such as slightly increased parking fees). Encouragingly participants, on average, reported competing with each other, understood the software tool and did not appear to just put in effort at the start of the game. There was evidence that participants understood which indicators it was easiest to achieve stretching performance for

and could focus their efforts on those (as the interviews and questionnaire suggested happens in the real game). The motivations of the participants in respect of the key features of the game therefore map reasonably well to practitioners suggesting that the ecological validity is not substantially compromised.

Despite the provision of a training day, there was very strong evidence of learning effects across the three experimental rounds which underlined the importance of randomized play. Figure 3 shows the results of planning time invested in the first, second and third rounds of play. Whilst the time spent setting targets for all 40 players followed the order $E_1 > E_2 > E_0$ the differences were not significant statistically and likely to be somewhat influenced by learning.

Insert Figure 3 here

Despite the selection of students in an attempt to attain homogenous players it appears that capabilities differed between players within groups and also between different groups. Figure 4 shows the target setting and performance over five years for two of the groups playing the same experimental scenario.

Insert Figure 4 here

It can be seen that there is substantial variation between the target ambition set and achieved by some players in group X and little in group Y. This is found throughout the data set so there is heterogeneity to varying degrees in all groups. When asked in the exit questionnaire 30% of respondents agreed that their groups were homogenous and 30% disagreed. This is sufficiently large to limit the ability of the experimental findings to differentiate between hypotheses H2 and H3.

4.3 Self-reported effort results

Table 6 shows the effort committed by players in setting targets and playing the five year game across the three scenarios. There is no significant difference between the self-reported effort invested in target setting across the three experimental scenarios but self reported effort in playing the game in both E_1 and E_2 are statistically significantly greater than for E_0. The t – test results found that for $E_1 > E_0$, t = -2.411, p= 0.021 and $E_2 > E_0$, t = -2.208, P = 0.034. Page's L test was run on the eight groups (each group was treated as single data point) and found that there was a significant level of agreement between the predicted ranking for homogenous players (E1>E2>E0) and the experimental ranking (L = 104 α = 0.05, one tailed test) for self reported effort in playing the three games.

Insert Table 6 here

4.4 Transport outcome results

Table 7 shows the summary results for target ambition (set at the start of the experiment), target attainment (were the targets met in year five) and absolute outcomes (how much was achieved ignoring whether or not the target was met).

Insert Table 7 here

The first row of Table 7 suggests that for 3 of the 7 indicators the level of target ambition followed the expected assumptions for homogenous players that $E_1 > E_2 > E_0$ and that this increases to five out of seven indicators if the conditions are relaxed to E_1 and E_2 both being greater than E_0 but in any order. This suggests that the targets set are more ambitious with some form of incentives in place than without. On target attainment it appears that setting more ambitious targets does not mean that targets will be achieved with two of the five indicators for which more ambitious targets were set in both E_1 and E_2 not being met (5/7 reducing to 3/7). Politically this may be important if ambitious targets cannot be delivered. It is however necessary to look at the absolute level of achievement to see if the system has delivered its goals of incentivising better performance. In this regard, 3 of the seven indicators were higher strictly in the order $E_1 > E_2 > E_0$ and this increases to six of the seven indicators if the conditions are relaxed to E_1 and E_2 both being greater than E_0 but in any order. This implies that although the targets might not all be met, the absolute outcomes are better when there is some form of incentive than when there is not. It has not been possible to establish statistical significance within such a small sample.

4.5 Summary of experimental results

Taking the experiment as a whole, we can conclude that there is statistical support for H_1 from the t-tests on self-reported effort and the evidence on outcomes (although this is lacking in statistical significance). On H_2 and H_3, there is a more mixed picture, including some statistical support for H_2 but other evidence pointing to either H_2 or H_3. Although less conclusive, this is at least consistent with the theory which suggests mixed homogenous/heterogenous groups would lead to a mix of rankings of E_1 and E_2.

5. Conclusions

This research set out to determine the likely impacts of a performance reward funding regime for local transport planning which is linked to achievement against targets. None of the methods adopted alone would have provided a picture complete enough for us to draw practical conclusions. Taken together however, there are some key themes and conclusions arising.

First, linking performance rewards to target setting will lead to competition between authorities and, designed correctly, this will lead to greater levels of achievement against the key metrics in the system than if no performance rewards are available. The cost-benefit case for performance rewards remains unproven, however. Data on the cost of effort to local authorities, the benefits of greater achievement and the additional management costs would all be required.

Secondly, incentive design needs to take account of ability differences between the local authority players. If the players are inherently similar (homogeneous) in ability, or if the game can somehow be designed to offset ability differences, then the strongest incentives will be provided by offering a small number of prizes or even a single prize. By contrast, if there are ability differences that make some authorities inherently weaker than others, and if there is nothing that can be done about this, then total effort is more likely to be maximised by offering multiple prizes.

Thirdly, in the real LTP game (circa 2006-7) we found that the DfT was using subjective assessment to compensate for ability differences with apparent success. However, there is no guarantee that that approach will be transferable to other contexts, it lacks transparency and has a substantial management cost. Splitting the contest into several smaller pools of more evenly-matched authorities would have only a small negative impact on the incentive to effort although potentially inflates the risk of collusion.

Fourthly, our findings suggest that the effort spent in establishing a level playing field was, to some extent, wasted by offering a large number of prizes. This may be motivated by a political sensitivity where it is more attractive to spread the prize fund widely to avoid being seen to penalise a large number of localities. There are lessons to be drawn on running competitions for clearly defined funding pots (e.g.

to tackle urban congestion). Here, only those self-selecting authorities that believe they have a good chance of winning will compete. Smaller numbers of prizes are likely to be more appropriate and politically acceptable in such situations.

Finally, the new approach to target setting appears to have had the desired effect of focusing local authority transport planning departments on the quality of their planning. Whilst this is welcome and acknowledged, there is still a huge tension between their support for the five year planning process and the feeling that too much is required of local authorities by central government.

This study has provided some empirical evidence on the potential impacts of targets and performance rewards on transport outcomes and on the behaviour of participants under any such regime. Targets and performance rewards are becoming more widespread globally throughout public services. Whether they will prove to be a good thing for transport remains to be seen. Whilst the system appears to offer benefits of enhanced performance on defined indicators this may lead to imbalances in priorities between those things for which performance is rewarded and those for which it is not.

Acknowledgements

The research was funded by EPSRC under grant number EP/C002350/1 ‘Optimal Incentive Structures for Integrated Transport Strategies’. The authors are grateful to the Department for Transport and the authorities that participated in the research

and to Dr Astrid Gühnemann for her contribution to the PLUTO work and Dr Henrik Orzen and Dr Claus Abbink for their advice on the game theoretic framework. We are also grateful to the reviewers for their comments. Responsibility for the findings rests solely with the authors.

References

Beattie, J and G. Loomes (1997) “The impact of incentives upon risky choice experiments”, *Journal of Risk and Uncertainty*, **14**:155-168.

Blavatskyy, P. (2004) *Why the Olympics have three prizes and not just one*. Working Paper No. 200. Institute for Empirical Research in Economics, University of Zurich.

Bonsall, P. (1994) PLUTO: computer assisted training for transport planners. *Proc. Conference on computer aided learning in engineering*, University of Sheffield.

Clark, D.J. and Riis, C. (1998) Influence and the discretionary allocation of several prizes, *European Journal of Political Economy*, **14**, 605-625.

De Palma, A. and Lindsey, R. (2003) *Approaches to Institutional Issues*, Deliverable 2.2 to the Transport Institutions in the Policy Process project, EU 5th Framework Project Contract No: GMA2/2001/52043 S07.15419 TIPP, DG TREN.

Department for Transport (2004) Full Guidance on local transport plans: second edition www.dft.gov.uk/, last accessed 04/03/09

Department for Transport (2005) Financial Planning guidelines for local transport plans. Consultation paper, July 2005 www.dft.gov.uk/, last accessed 04/03/09

Department for Transport (2006) *Transport Innovation Fund*. DfT, London. www.dft.gov.uk/

Dutta, P.K. (2000) *Strategies and Games: Theory and practice*, MIT Press, ISBN 0-262-04169-3

FHWA (2004) Transportation Performance Measures in Australia, Canada, Japan, and New Zealand, International Scanning Study Team Report, US Federal Highway Administration, http://international.fhwa.dot.gov

Funke, J. (2001) 'Dynamic systems as tools for analyzing human judgement', *Thinking and Reasoning*, **7** (1), pp69-90

Hart, S.G. & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In *P.A. Hancock and N. Meshkati (Eds.), Human Mental Workload* pp.139-183. Amsterdam: North Holland.

Hodgson, L., Farrell, C.M. and Connolly, M. (2007) Improving UK Public Services: A Review of the Evidence, *Public Administration*, **85**(2), 355-382

Hood, C. (2006) Gaming in Targetworld: The Targets Approach to Managing British Public Services, *Public Administration Review*, **66**(4), 515-521

Kelly, C., Nellthorp, J. and Marsden, G. (2006) *How do local authorities set their LTP targets?* Proc. Transport Practitioners Annual Meeting, Manchester, 25-26 July

Kelly, C., Marsden, G. and Nellthorp, J. and (2008) An assessment of target setting in local transport plans in England. *Project Working Paper 1*, www.its.leeds.ac.uk

Marsden, G. and Bonsall, P. (2006) Understanding the role of performance targets in transport policy, *Transport Policy* **13** (3), 191-203

Moynihan, D.P. (2006) Managing for results in state government: Evaluating a decade of reform. *Public Administration Review* **66**, 77-89

Nellthorp, J. and Marsden, G. (2009) Performance incentives in multi-level government: an application of rent seeking contests in the transport sector, *submitted to the Journal of Public Administration Research and Theory* (March 2009)

NTTP (2008) Commentary on the July 2008 U.S. DOT Transportation Policy Proposal, National Transportation Policy Project, Bipartisan Policy Center, September 10, http://www.bipartisanpolicy.org/, last accessed 04/03/09

Smith, P. (1995). On the Unintended consequences of publishing data in the public sector, *International Journal of Public Administration* **18**(2/3): 277-310.

Tullock, G. (1980) Efficient rent-seeking, in *Toward a theory of the rent seeking society* (Eds.) J.M. Buchanan, R.D. Tollison and G. Tullock, Texas A&M University Press.

Tuominen, A. and Himanen, V. (2008) Assessing the interaction between transport policy targets and implementation: A Finnish case study, *Transport Policy* **14**(5): 388-398

Wickham, P. (2007) *Financial Incentives in Economic Experimentation*, ITS Research Discussion note, Institute for Transport Studies, Leeds, February 2007.

Table 1: Mandatory and Best Value Performance Indicators

Mandatory LTP Indicators	Mandatory Best Value Performance Indicators
LTP1: Accessibility target	BVPI 223: Principal road condition*
LTP2: Change in area wide road traffic mileage	BVPI 224: Unclassified road condition*
LTP3: Cycling trips	BVPI 99x: Total killed and seriously injured casualties (KSI)*
LTP5: Bus punctuality indicator*	BVPI 99y: Child killed and seriously injured casualties *
LTP6: Changes in peak period traffic flows to urban centres^*	BVPI 99z: Total slight casualties*
LTP7: Congestion^	BVPI 102: Public transport patronage*
LTP8: An air quality target^	BVPI 104: Bus satisfaction*
	BVPI 187: Footway condition*

Key ^only a requirement for certain authorities
*Indicators where stretching/ satisfactory thresholds were set.

Source: DfT(2004)

Table 2: Risk Assessment by target classification (satisfactory, stretching)

Indicator	Mean % chance of meeting target		
	All Targets	Satisfactory Targets	Stretching Targets
Killed and Seriously Injured	79.2	80.4	78.2
Child KSI	75.8	76.2	74.6
Slight accident rate	76.3	73.9	78.4
Bus Satisfaction	72.6	71.9	77.5
Bus Patronage	72.2	63.0	76.8
Bus Punctuality	70.2	68.3	71.7
Unclassified Road Condition	74.1	72.3	79.3
Footway condition	74.1	69.3	80.0

Table 3: Settlement adjustments for best and worst performing authorities 2006/07

Authority	Population	Nature	Adjustment	Initial Allocation (£M)	Cash Adjustment (£M)
Cornwall	501267	Rural	+25	8.53	+2.13
Reading	143124	Urban	+25	4.86	+1.21
Halton	119500	Urban	+25	1.69	+0.42
Derby	233700	Urban	+25	0.52	+0.13
Nottingham & Nottinghamshire	1055500	Mix	+25	10.48	+2.62
Cambridgeshire	597400	Mix	+17.5	7.37	+1.29
Norfolk	840700	Rural	+17.5	9.42	+1.65
Leicester	279000	Urban	+17.5	5.91	+1.04
Shropshire	289300	Rural	+17.5	3.81	+0.67
East Riding of York	314113	Rural	+17.5	2.81	+0.49
North East Lincolnshire	157983	Rural	-12.5	2.50	-0.31
Windsor and Maidenhead	141000	Mix	-12.5	1.55	-0.19
Milton Keynes	184506	Urban	-12.5	1.80	-0.22
Luton	184000	Urban	-12.5	2.87	-0.36

Table 4: Effort Measures

Process oriented effort measure		Outcome oriented effort measure	
Measure	Description	Measure	Description
Level of stretch set	The ambition of the targets set	Indicator levels vs. stretch set	The extent to which targets are met
Planning time invested	The amount of time invested in deciding what to set	Indicator levels vs. no incentive scenario	The change over a system with targets but no incentive
Number of scenarios run	The number of times participants tried different strategies each year		
Self reported input effort	The reported effort on the task		

Table 5: End of experiment questionnaire results

Statement	1 = Disagree to 5 = Agree	
	Mean	**σ**
Setting ambitious targets led to greater rewards	**3.7**	**1.1**
Setting ambitious targets was more risky	**3.6**	**1.0**
I understood how the assessments were made each year	**4.0**	**1.1**
Performance rewards helped me to achieve more	**3.5**	**0.9**
Performance penalties stopped me achieving my goals	**2.8**	**1.1**
I was penalised if I did not achieve the targets I set	**2.5**	**1.0**
I tried hard harder at the start of each five years than the end	**2.5**	**1.3**
I changed my strategy a lot for each experiment	**2.8**	**1.2**
The strategies I developed would be acceptable to the public	**2.8**	**1.2**
I competed against the other players	**3.6**	**1.2**
I found the PLUTO model difficult to use	**1.6**	**0.7**
The players in my group were of an equal ability	**3.0**	**1.1**

Table 6: Average Self Reported Effort in Setting Targets and Playing the Game[6] (effort score range = 0..100)

	Experimental Scenario		
	E0	E1	E2
Target Setting Effort	37.52	39.33	37.48
Experiment Effort	34.00	40.12	39.34

[6] Self reported effort is calculated using the NASA TLX scales

Table 7: Targets and Outcome Results

	Number of Indicators E1 > E2 > E0	Number of Indicators (E1 AND E2) > E0
Target Ambition	3/7	5/7
Target Attainment	1/7	3/7
Actual Attainment	3/7	6/7

Figure 1: Average Scores for each target for the three assessments

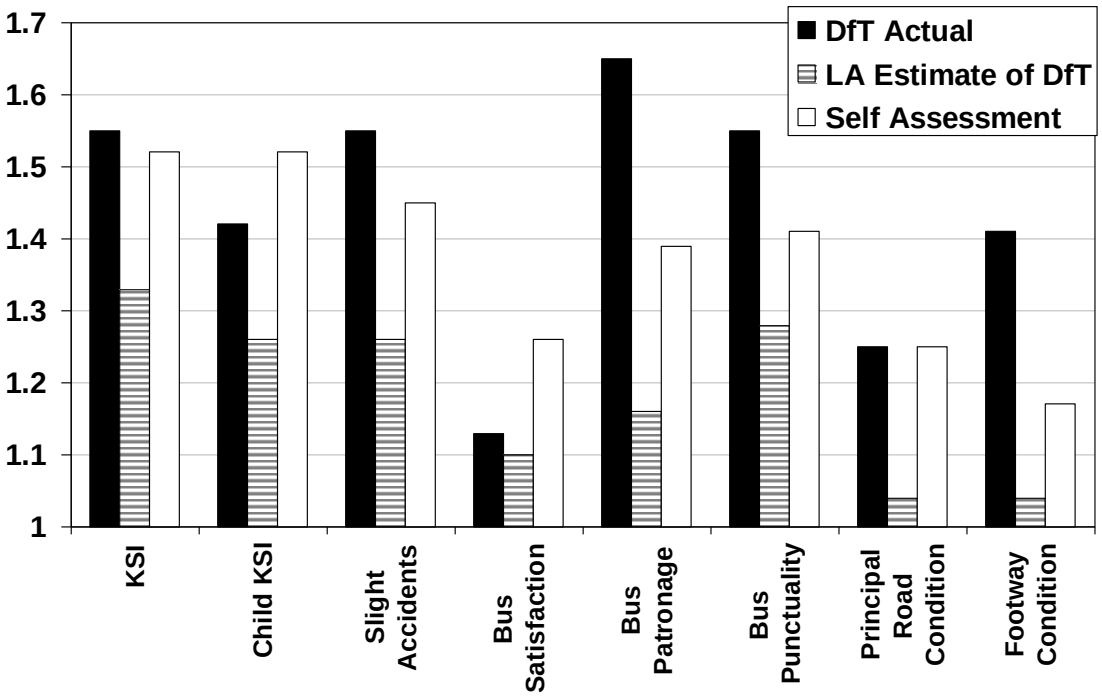


Figure 2: Effort responses in the 82 player game

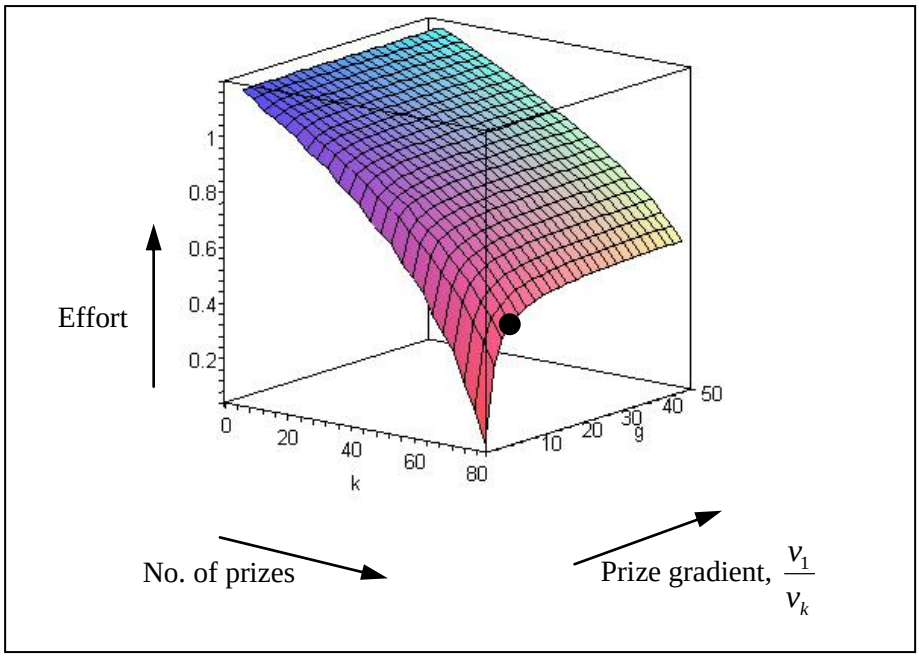


Figure 3: Time spent setting targets for the three experiments

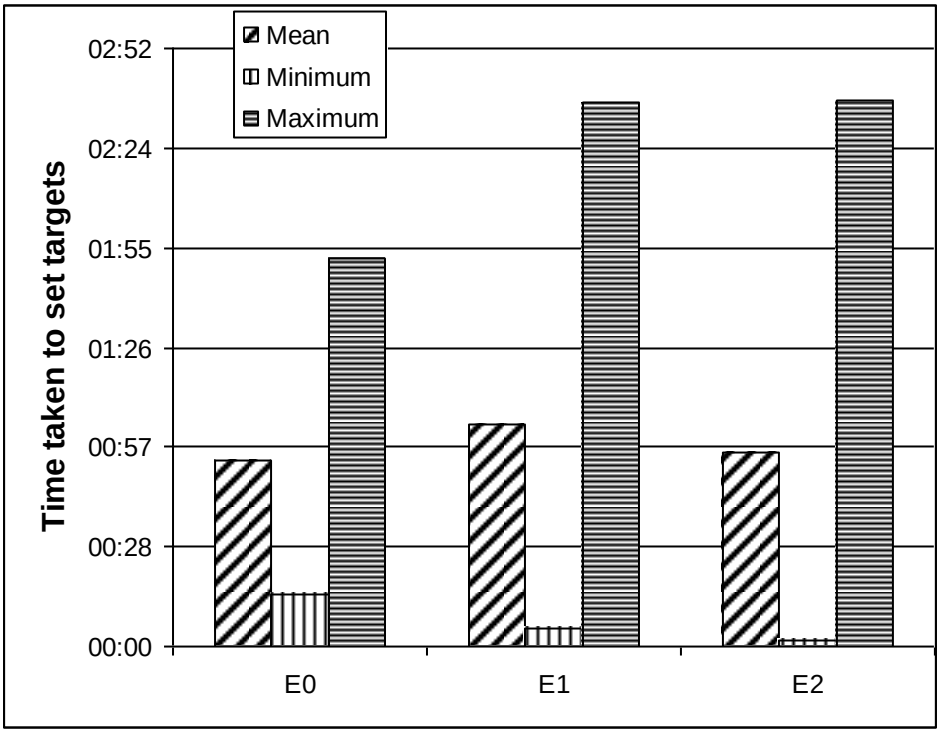
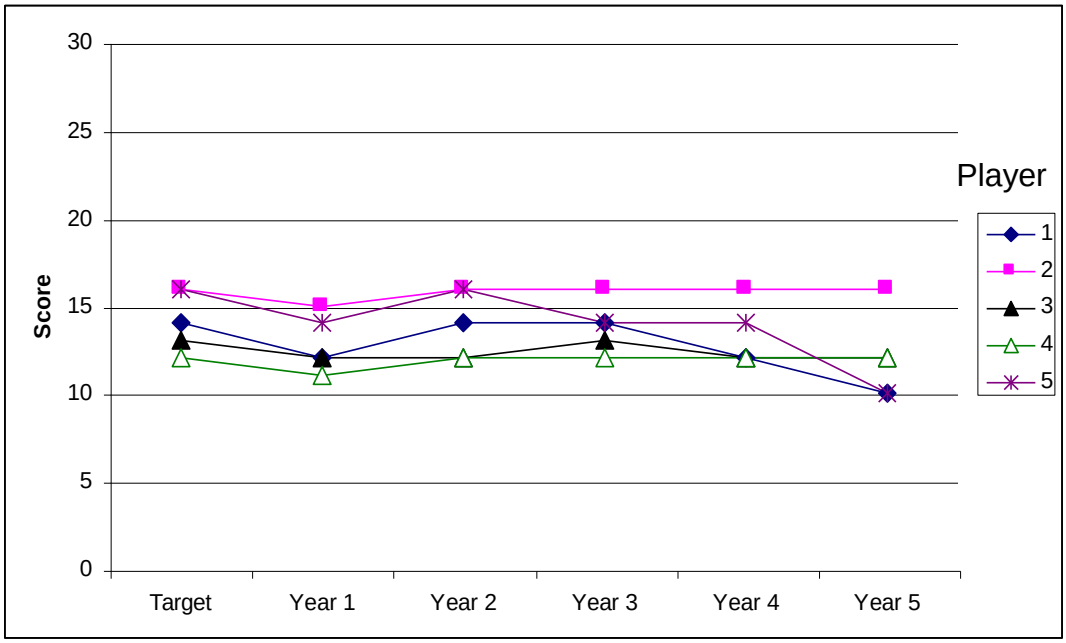
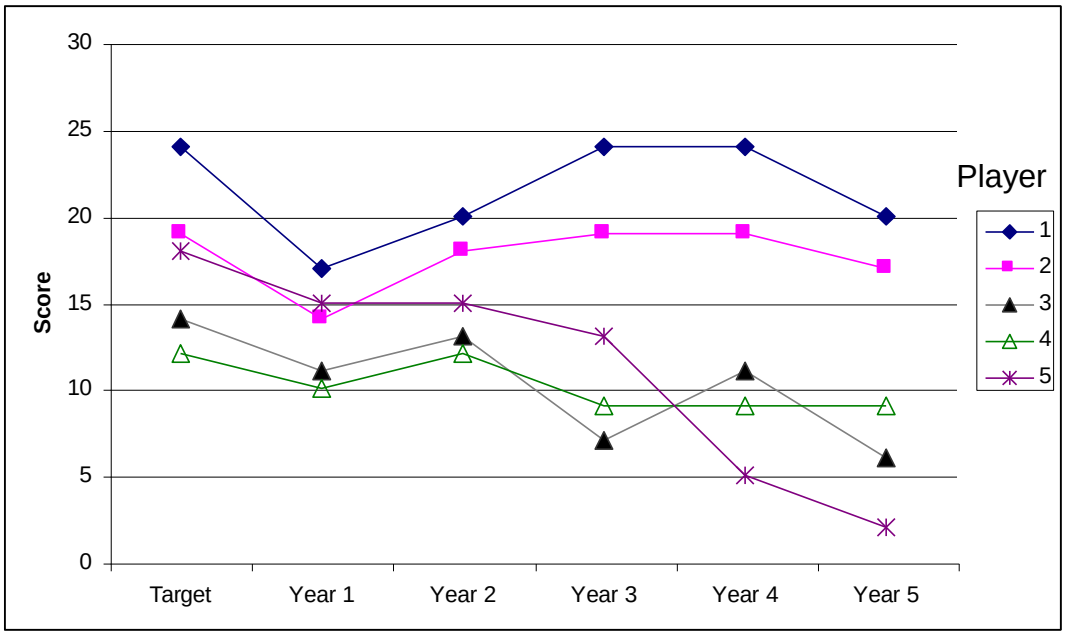


Figure 4: Target setting and performance scores over the 5 years for the five individuals in groups X and Y for experimental condition E2



Group X



Group Y