Sir Malcolm Bates
[Title and Address]

Dear Malcolm,

In preparation for the upcoming round of contract consultation, we have undertaken a detailed analysis of the contemplated PPP performance regime of the PPP contracts. We have focused on the performance regime because it has not been subject to critical scrutiny yet it represents the lynchpin of the PPP. I am sure every PPP advocate would agree that if the performance regime proves unworkable in practice then the PPP will fail. This letter sets out the results of our analysis, expressing concerns and raising questions about some of the areas where the documents are unclear.

The performance regime taken as a whole is staggering in its complexity. And deploying it – placing the very future of the Underground and indeed London's lifeblood and economy at its mercy – is a very risky endeavour.

Because the PPP contract does not provide for payment to the Infracos based on concrete deliverables -- a new fleet of cars, or carrying out specific and routine track maintenance, for example -- LUL is forced to resort to paying the Infracos based on how the Underground is performing and whether the Infracos are responsible for improvement or deterioration of performance.

It sounds simple and maybe even promising <u>in</u> <u>theory</u>. But how can one really tell how a system as complex as the Underground is performing for purposes of administering a contract, and who is responsible for that performance? Simply put:

- A very complex mathematical model of the entire Underground system had to be developed and incorporated into the thirty-year contracts.
- A system for attributing responsibility for "incidents" on the system had to be developed and incorporated into the contract.

The idea of incentivising Underground improvements almost purely through operation of a mathematical model that captures thousands and thousands of minute elements of Underground performance is fascinating to think about – in the abstract. However, it is reckless to experiment with implementing such an untested idea – and to bind yourself to the experiment for thirty years. (And while I recognize that the model may have been vetted in shadow running by the still LUL controlled infracos, truly that is no test at all when there is no real contractual adversity between the parties.)

Consider historic models, and how other aging metro systems (including Boston and New York) have been turned around. Putting it very simply, problem areas were identified by management, various capital works and maintenance solutions were identified and prioritized by management, and ultimately implemented, often by private contractors. The contractors would be paid as they delivered desired products, and would be responsible if the products didn't work as promised (warranty and reliability requirements).

The PPP performance model turns tried and true turnaround methods on their head, resulting in systems analysis run amok. LUL is to pay the Infracos based on how the system is doing, and whether improvement or deterioration of the system is contractually attributable to the Infracos. But how do you figure out how the system is doing and who gets credit? Well, if you are going to have a chance of succeeding, you have to be able to measure the performance of the Underground in excruciating detail. You need to measure things like how quickly the customers can get from place to place on each line (capability measure, with an enormous number of physics based variables and inputs), how quickly they are getting from place to place (a cross of the capability and availability measures, including crush factor and dozens of other relevant inputs), what the system looks like to customers (Is there chewing gum residue on the walls of any station? Are there four or more pieces of litter larger than 5 cm in one square meter of one car?), and how comfortable the ride is on each line. You need to measure all the minutiae that make up the Underground network as a whole, and put them in a formula that assigns proper weights to each element, and allocates pounds to each category of performance. Then you need to gather the data. And you need to figure out who gets credit, or blame, for each incident that arises. Finally, you must reach common agreement or pursue dispute resolution for each attribution of responsibility.

"Very complex" is an understatement. The mathematical model LUL intends to employ is scattered throughout the 200 plus page Performance Measurement Code, the massive Service Outputs Schedules, Line Specific Performance Measurements Schedules, and Nominally Accumulated Customer Hours (NACH) Tables, the Capability Model (which has not been provided to us), and finally the Performance Payment Mechanism Schedules. Key elements also require cross-reference to other lengthy and complex documents. The model is replete with dense formulae such as

"$\mathrm{ASP}^{\mathrm{AS}} = \sum_{\mathrm{i}} \left\{ \mathrm{ASP}_{\mathrm{i}}^{\mathrm{AS}} \text{ x } \left(1 + \mathrm{UT}_{\mathrm{i}}\right) \right\} - \mathrm{ASP}_{\mathrm{THD}}^{\mathrm{AS}}$ **,if**

$$ASP_{THD}^{AS} < \sum_i \left\{ ASP_i^{AS} \; x\,(1 + UT_i) \right\}$$

$$ASP^{AS} = 0, if \;\; ASP_{THD}^{AS} \geq \sum_i \left\{ ASP_i^{AS} \; x\,(1 + UT_i) \right\}$$

where:

ASP^{AS} **is the total number of Abatement Service Points allocated for Defects and Failures attributable to Infraco and Closed-out in the Payment Period;**

ASP_{THD}^{AS} **is the Facilities Threshold Service Points performance level set out in Appendix 6 of Schedule 2.2 of the applicable PPP Contract;**

ASP_i^{AS} **is the number of Abatement Service Points allocated for Defect or Failure *i* attributable to Infraco and Closed-out in the Payment Period, as determined in accordance with Schedule 8 *(Service Points)*; and**

UT_i **is the time for which a Facility is subject to Defect or Failure *i*, being the difference between the Time of Notification and the End of Duration as defined in Part B of Schedule 5 *(Data Requirements)*, expressed as a number of twenty *(*20*)* hour service days calculated to two *(*2*)* decimal places."**

And:

"The Excess Run Time required in order to calculate the value of H_i^{SR} according to the applicable Train Service NACHs Tables set out in Schedule 2.3 of the applicable PPP Contract *(*NACHs Tables*)* shall be calculated in accordance with the following formula:

$$ERT = d \; x \; 3.78 \; x \left(\frac{1}{v_{SR}} - \frac{1}{v_L} \right)$$

where:

ERT **is the Excess Run Time deemed to result from the Speed Restriction caused by Incident *(*i*)*, calculated to the nearest second;**

d **is the length of track to which the Speed Restriction applies, measured to the nearest metre;**

v_{SR} **is the maximum permitted speed of a train under the Speed Restriction, measured to the nearest kilometre per hour; and**

v_L **is the Nominal Line Speed Parameter for the length of track to which the Speed Restriction applies, as set out in Appendix 7 of Schedule 2.2 of the applicable PPP contract *(Nominal Line Speed Parameters)*."**

And:

"*Y = (vis x wis) + (vmn x wmn) + (vdb x wdb) + (vtm x wtm) + (vcd x wcd) + (vms x wms) + (vfyr x wfyr) + (vfyt x wfyt) + (vfa x wfa) + (vfb x wfb) + (vfta x wfta) + (vftb x wftb)".**

If PPP is concluded and TfL inherits the Tube, I will be explaining to delayed Tube passengers that their experience was or was not attributable to the recently privatized Infracos based on the application of these formulae! (Maybe it would help to clear things up if I replaced your "Publicly Run – Privately Built" ad posters with these particular formulae.)

More seriously, these formulae – and there appear to be hundreds of them in the documents– and the massive amounts of data that must be gathered, categorized and assessed to deploy the formulae actually form LUL's PPP safety net! They form the very heart of LUL's ability to require the Infracos to deliver Underground improvements for the billions of pounds they will be paid.

I believe it is your fiduciary obligation, and that of your colleagues on the LT Board, to think very hard, and ask staff many questions, before putting LUL's eggs in this performance measurement regime basket – a basket which becomes increasingly more fragile with every revision of the contract documents. I question whether you and senior LUL staff have even been made aware of some of the ways the performance regime has been eroded in the latest round of contract documents issued after preferred bidder designation, examples of which I have included below.

I also wonder whether anyone at the controls of PPP implementation has even the most elementary understanding of the implications of this regime. I am sure that DTLR wasn't being intentionally misleading when it recently published "Your Tube – Publicly Run – Privately Built, " which unequivocally asserts that a train breakdown at Victoria for 10 minutes in the morning rush hour would cost the Infraco £21,000, and that an escalator at Oxford Circus station being out of service for a month would cost the private sector £45,000. Perhaps DTLR just did not appreciate the implications of the following language:

"Actual Availability is expressed in terms of the sum of:

(i) the rolling average value of Availability in any consecutive three (3) Payment Periods of Lost Customer Hours deemed to result from Disruptions, other than Speed Restrictions, caused by Incidents which are attributable to Infraco in respect of each Line; and

(ii) the rolling average value of Availability in any consecutive six (6) Payment Periods of Lost Customer Hours deemed to result from Speed Restrictions caused by Incidents which are attributable to Infraco in respect of each Line."

This language means that the Infracos are not charged for each train delay or escalator breakdown incident that occurs– overall performance is instead averaged over three months (or for Speed Restrictions, six months), and then that average is compared to a benchmark allowance of "Lost Customer Hours" – tens of thousands, and on some lines hundreds of thousands, of permitted Lost Customer Hours per month. Only if the three-month average exceeds the benchmark level of permitted customer delays will the Infraco begin to pay anything for this type of problem in performance.

Indeed, if they are above "benchmark" (which we understand will be 5% below current performance), the Infracos will get a bonus. Imagine that. Service reliability deteriorates below what it is today – and a bonus is paid?

And incidents such as the stalled train in the tunnel reduce pay only if LUL successfully manages to attribute responsibility for the incident in question to the Infracos. The contract provides a disagreeable Infraco with ample opportunity to challenge and/or delay being held accountable for any Incident that gives rise to a potential for abatement – a total of ten procedural hurdles along the way to final attribution of responsibility, each presenting opportunity for delay and distraction, including initial attribution, review, second review, Level A meeting, Level B meeting, Contract Managers meeting, PPP Board meeting, Senior Representative review, Adjudicator review and finally – if anyone can remember what happened or still cares - a court of law. And, as we read the contract, the Incident doesn't even go into the hopper for payment purposes until any dispute over responsibility for the incident is settled. Very realistically, this could be years.

I'm not arguing that an Infraco should be held financially accountable for every incident of customer delay it causes. I'm asking you a more fundamental question: Is this any way to run a major metropolitan transit system? How effective can a system like the performance measurement regime be, when the process for imposing penalties for shoddy Infraco performance is so convoluted and so attenuated from day to day work? And how did DTLR reach the conclusion that one train stopped for 10 minutes in rush hour will cost the Infracos £ 21,000?

Complexity and attenuation from reality is far from the only issue. The performance regime is also rife with opportunities for the Infracos to play the system, some of which have been created recently, after the BAFO and selection of preferred bidders. Some examples:

- To measure how quickly customers will get from station to station, data such as "Full Speed Run-out Run-in Time" and "At Rest Run-out Run-in Time" and multitudinous other physics-based variables that affect travel time need to be run through complex computer programs. The data needs to be updated from time to time, as service improves or deteriorates. This is to be done in part through actual test runs of trains. Test results will of course depend on which trains are tested, so LUL originally provided that it would have the right to choose the trains that would undergo testing. The Preferred Bidders must have objected to this,

probably concerned that LUL might choose to test the worst trains in the fleet. A reasonable compromise would have been for the test trains to be selected randomly. Instead the most recent draft of the documents provided that the Infraco identifies a pool of trains from which test trains may be selected. Fully 40% of the trains in a fleet can be kept out of the pool – allowing the Infracos to have only its better trains count for this critical measure driving their own compensation.

Who would agree to such a trap?

- Infracos have far too much control over the test process. They design the test programme, and they can retreat to the lengthy dispute resolution procedure if they are unhappy with any LUL objections to that test programme. And it seems that, if a test programme is implemented only after dispute resolution, the results are retroactive for financial purposes only if the Infraco prevails in the dispute resolution process. We found no comparable retroactivity if a court rules in favour of LUL on the test programme. (The idea of a court ruling on the test programme is itself extraordinary.) Given lopsided retroactivity and the potential for wearing LUL down in the dispute resolution process, why would a rational, profit-seeking Infraco not do everything in its power to delay a test that may prove deterioration of service?

- The Infracos will control much of the information that is necessary to evaluate the cause of incidents that are to be input into the various "availability" formulae. Controlling the information gives them a very decided advantage.

I have asked you before, and I ask again: what kind of contract administration resources will LUL have to throw at this Hydra? I understand that there will be thousands and thousands of incidents to evaluate every month, just to execute one facet of the multi-faceted performance regime. Who is going to do it, and what is the budget for the resources necessary to give LUL half a chance against Infracos who might perceive that they stand to gain more from obstructing and playing the system than from fixing the track?

Words simply cannot begin to accurately convey the complexity of this mathematical modeling of the Underground, and the inherent risks it presents. I am deeply disturbed by the prospect that the riding public in London, indeed the very economy of London, will be held hostage to this model and the inevitably continuous battles that must be fought as the private sector attempts to eke advantage from it.

I am also disturbed by what I believe to be fundamental erosion of performance standards that seem to be introduced in the latest set of contract documents. For example:

- The Performance Measurement Code sets standards for how quickly various faults have to be rectified. Putting aside drafting problems in the schedule and the interplay of the schedule and the formula, the draft contract documents previously

provided that for each day beyond an established deadline that a particular fault was not rectified, the Infracos incurred a specified number of Service Points, each of which was to cost the Infraco about £ 50. The total number of service points for this particular component of performance was to be added up and multiplied against £ 50. The amount derived was to be docked from the monthly payment (assuming the dispute resolution process wasn't invoked, of course).

In the most recent draft of the documents this appears to have been watered down dramatically. As we dug into the documents, we found a change to the formula for adding up service points for fault rectification delays that appears to give each Infraco a sizeable allowance of Service Points for such delays– over 40,000 per four week period in the case of one Infraco, and approximately 30,000 per four week period for each of the other Infracos. Unless we have been stumped by the complexities of the formula, this allowance appears to add up to an LUL waiver of well over £ 2 million per month for one Infraco alone, and over £ 1.5 million per month for each of the other Infracos. This creates an enormous potential for service below the standards the Infracos originally bid to meet.

Remarkably, it appears that the relief afforded to the Infracos from performance standards in this category doesn't even stop with this monthly allowance. In the BAFO documents, bidders were told that only missed deadlines would be taken into account when the aggregate abatement was calculated. Now it seems that the preferred bidders are being offered credit against delay-related abatements if they manage to rectify other faults ahead of contractually specified deadlines.

- Material relief also appears to have been granted to the Infracos in connection with their obligation to improve the condition of "Track", "Earth Structures", "Bridges and Structures", and "Track Drainage" – at least for the first 7 ½ year contract period, which is the only period for which there is fixed pricing.

  These asset categories apparently fall outside of the mathematical model for measuring performance. Thus, Infraco performance with respect to these assets is measured by the actual condition of the assets – what percentage of the assets in question are permitted to be in what state of deterioration at the end of each 7 ½ year contract period.

  A few examples of the relief provided to the Infracos on this front in the latest draft:

  At the time of BAFO, 50% of open section rail on one line was permitted to be in "D or worse condition" at the 7 ½ year mark. (There are five asset condition categories, "A" being the best and "E" being the worst.) Now, fully 100% of open section rail on that same line is permitted to be in "D or worse condition" at the 7 ½ year mark.

BAFO requirement: 46% of "sleepers and base concrete" in tube sections on one line could be in D or worse condition at 7 ½ years. Preferred Bidder revised requirement: 92.5%.

BAFO requirement: 0% of viaducts on one line could be in E or worse condition at 7 ½ years. Preferred Bidder revised requirement: 7%.

What persuaded LUL to adopt these and other reductions in Infraco obligations so late in the game? Will the ongoing negotiations result in additional degradation?

* * *

In addition to the general concerns set out above, I have a number of particular questions about the Performance Measurement Code to which I would appreciate answers from your staff before the next round of contract consultations:

1. The Availability measure was changed in the last draft so that speed restriction-related availability events are now averaged over 6 months, rather than three. (Performance Measurement Code, Section 4.1.1) Another section of the documents was revised to provide that all speed restrictions are deemed closed out after three months. (Paragraph 3.4 of Schedule 3 to the Performance Measurement Code.) What was the philosophy underlying the special treatment afforded to Speed Restrictions at this point in the bid process? And why the deemed close-out of Speed Restrictions?

2. In several places, the documents provide that, notwithstanding the thousands of words and formula and input parameters identified in the documents as controlling the Capability adjustment, the "Capability Model" takes precedence. The current model, which is on a diskette, was not provided to us, and the documents indicate that it is still being validated. Who controls the model? What changes are being made to it? How will they be vetted, tested and incorporated in the contract documents, so that all parties, including TfL, can understand the impact of changes late in the game? The model on this diskette is a key ingredient of the performance regime. Is it going to be escrowed, as key contractual testing software usually is these days?

3. What was the philosophy underlying, and the origins of, the addition of the following language to the Performance Measurement Code (Section 4.3A.7):

"If the value of ERT, calculated in accordance with paragraph 4.3.6 is greater than 120 seconds, the value of H^{SR} shall be determined according to the following formula:

$$H_i^{SR} = \frac{H_{i(120)}^{SR}}{120} x\, ERT$$

where:

$H^{SR}_{i(120)}$ **is the value of** H^{SR}_{i} **when the value of ERT is assumed to be 120 seconds, determined according to the applicable Train Service NACHs Tables set out in Schedule 2.3 of the applicable PPP Contract** ***(NACHs Tables)*****."**

Does this limit Excess Run Time to a deemed 120 seconds? Why?

4. To weight the various ambience factors, the Performance Measurement Code (Section 5.1.2) provides that:

(b) **For each Infraco, LUL shall calculate the total Ambience score for each Quarterly Period *(A)* using the following formula:**

$$A = \frac{\sum_{js} QASAS_{js} \; x W^{S}_{js} + \sum_{jL} QATAS_{jL} \; x W^{T}_{jL} + \sum_{kL} QATAS_{kL} \; x W^{T}_{kL}}{\sum_{js} W^{S}_{js} + \sum_{jL} W^{T}_{jL} + \sum_{kL} W^{T}_{kL}}$$

Similarly, Train Ambience is weighted as follows:

$$TMS = \sum_{j} \left(\frac{QATAS_{j} \times W^{T}_{j}}{\sum_{j} W^{T}_{j} + \sum_{k} W^{T}_{k}} \right) + \sum_{k} \left(\frac{QATAS_{k} \times W^{T}_{k}}{\sum_{j} W^{T}_{j} + \sum_{k} W^{T}_{k}} \right),$$

Is there a document that explains in plain English the relative weightings of ambience factors? For example, how will "very clean train seats with only very minor areas of dirt or 'dull' appearance" affect payments to the Infracos relative to "some areas of dust, dirt or staining, e.g., dirt build up in corners and crevices" in the Station Ticket Hall? What is the relative weight of graffiti in the Ticket Hall to non-scratched graffiti in the trains? Or to scratched graffiti in the trains?

5. Schedule 5 to the Performance Measurement Code states that Train delays are to be logged against one Train only, even though other Train(s) may be affected. Similarly, Signal Failures are logged against one Train only, even though other Trains may be affected. Is it safe to assume that cascading delays are caught for purposes of the abatement formula by the NACHs tables? Can you explain how this is accomplished in which formula and provide arithmetic examples? Different types of train delays can have varying degrees of cascading effect – how is this captured and how has it been validated?

6. Can you identify particularly sensitive parameters of the Capability measurement?

7. Last spring, we were advised by LUL staff and consultants that caps on performance related bonuses provide a safeguard against potential errors in the formulae and to eliminate the potential for windfalls. We were thus surprised to note in the latest

draft contract documents that these caps have been eliminated in the "availability" category. On what basis did LUL conclude that the caps were no longer required to protect against errors and windfalls?

8. Has Freshfields opined as to the enforceability of the Performance Measurement Code, and, in particular, whether the Service Points regime represents enforceable liquidated damages rather than penalties? We would like to review their legal analysis in this regard.

In the April consultation, we were first exposed to the extraordinary dynamic simulation model that LUL's consultant developed in order to mimic Underground operations. (This is the model that led the consultant to conclude "if the [PPP] bidders are subjected to risks thought likely to occur, performance in some [PPP] cases appears likely to decline and service to passengers will suffer.") Frankly, the flow charts offered by the modeler were almost comical in their complexity. Certainly they were difficult to challenge – anything so inordinately complex is tough to challenge. Yet we did find arrows between boxes going in the wrong direction. Of course we did. The odds of errors are great when one attempts to reduce the workings of a complex dynamic enterprise to a model and a formula. And the resources to find all the errors and understand their implications are probably far greater than those demanded to create the model in the first place.

And yet it is just this type of model that London will be banking on for the next thirty years.

What a dangerous game.

Sincerely,

Robert R. Kiley