Findings of a review of the National Passenger Survey

February 2006

Passengerfocus
putting rail passengers first

Findings of a review of the National Passenger Survey

February 2006

Contents   **Page**

Executive summary

Consultation

The NPS Review included a consultation stage during which 39 NPS users, representing 17 TOCs and 13 other NPS user organisations, expressed their views. The results are described in Section 2. Briefly, they indicate that:

- All the people we consulted regard the NPS as an important or essential source of information. The NPS is seen as the only source of passenger satisfaction data and passenger evaluations of rail performance which covers the whole of GB, measures trends over time and allows comparisons to be made between TOCs.

- Although most TOCs commission considerable amounts of research in their own regions to allow more detailed analysis (e.g. on individual stations or routes) or to investigate certain questions in more detail (e.g. ticket sales or pricing), all the TOCs we spoke to regard the NPS as an essential tool and use it heavily.

- Other NPS users varied in the extent to which they use the detailed NPS data but they all regard it as an important source of valuable information.

- The NPS questionnaire is regarded as comprehensive. Most of those we spoke to use all the satisfaction and rating data "a lot" and look at other parts of the data at least occasionally. There are no parts of the questionnaire which can be eliminated without impairing the survey's utility for at least some users.

- The NPS survey is regarded as reliable (movements in NPS ratings for individual TOCs were said to be consistent with the TOCs' own research).

- Most users want the NPS to report twice a year - or more often.

- The NPS is used by Passenger Focus and the DfT as one of the key measures of the performance of individual TOCs. The feedback we received suggests that the mere existence of the NPS ratings puts pressure on TOCs to improve performance. Also, some TOCs use the NPS to benchmark "best practice" performance levels or to argue the case internally for the resources necessary drive up their own performance.

- A few NPS users regret that the survey does not cover more goals - e.g. explain why scores are at the level they are, or indicate the best way they can be raised, or go into more detail on particular topics or places (which may be why TOCs do so much additional research of their own). Rather more were concerned that the questionnaire is already quite long enough.

- Some feel that the methodological documentation of the NPS should be made more detailed and more accessible.

- Detailed knowledge of the NPS is very patchy: e.g. most TOCs seemed very familiar with it and commission extra analysis (or analyse their own copy of the NPS database), but some of the other users were unaware that such options were available.

Conclusions and recommendations

This NPS Review also included a study of relevant documentation, special analysis of NPS and other data, an evaluation of various survey elements, a technical review of the sampling and weighting design and procedures of the NPS (described in Section 3). After evaluating all of the above, our opinions are as follows:

- The current self-completion survey method is very appropriate for the purpose of achieving the key NPS objectives.

- The questionnaire requires no major changes or additions (and major additions should be resisted as it is already a relatively long questionnaire).

- The question scaling methods used on the NPS (5-point semantic) should not be changed, for they are effective, well understood, and to change them would sacrifice all the historical comparisons.

- The NPS response rate is remarkably high for a long self-completion survey with no respondent incentives: 38% average over the last 4 waves. However, the refusal rate is unknown, and because of differential response rates, the survey under-represents passengers under 35. This should receive attention.

- The NPS sample design is robust in terms of selecting stations, numbers of fieldwork shifts and interviews per shift for each TOC.

- The NPS sampling procedure for shift allocation by time of day is deficient. In particular it over-represents morning peak travel and under-represents evening peak travel. This must be addressed.

- NPS weighting technology (rim-weighting) is satisfactory, but the actual weighting matrix has some weaknesses - principally that journey-purpose weighting is based on ticket type, which is not an accurate predictor of journey purpose. This should receive attention.

The national railway system is an extremely complex entity: designing a survey to measure attitudes to it presents a challenge for the researcher, and any successful solution is likely to be a complex one. In many respects the NPS design is effective. In particular it is very effective in achieving its key goal of providing a consistent measurement of satisfaction scores and detailed evaluative ratings over time for each TOC, and for GB as a whole.

The description of the methodology included in most NPS Reports is inadequate for a survey of such complexity. The technical documentation should be made more complete and more accessible.

In Section 4 of this report we have recommended various steps that should be taken to improve those aspects of sample and weighting design which we judge need to be improved. Care must be taken to avoid disturbing the continuity of the results, but we believe these changes could improve the sample profile to the extent that less weighting will be required (which will in turn improve the accuracy of the data). On the most optimistic assumptions, these changes may even allow sample size to be reduced slightly without any loss of accuracy or reliability.

We have found no other options that we think are likely to reduce cost, but we believe there are many ways in which greater benefit could be derived from the NPS - principally by improving communication to users and potential users, and encouraging more and wider use of the NPS data, and we have made several recommendations to this effect. These include producing:

- a "user's guide to the NPS"

- regular, detailed reports monitoring the contribution of individual service elements to overall customer satisfaction, at various levels of performance (positive & negative drivers)

- a series of user-friendly reports on other specific topics.

All these could be produced using existing NPS data. If additional *ad hoc* research can be funded as well, the scope for producing information of value to the rail industry will be that much greater.

To the extent that such reports draw greater attention to the levels of performance which are being achieved, or help to understand and explain how these performance standards can be raised, they will be of benefit to rail passengers generally.

In Section 4 we have made a number of other recommendations for improvements to the survey, the most radical of which would be to extend it to cover 48 weeks of the year, instead of the present 20. We believe this could be achieved at little or no extra cost.

This Review has provided strong evidence of the ways in which the NPS is already contributing to improving rail performance and thus passenger satisfaction. We believe that the recommendations we have made will help to make that contribution even more significant.

Introduction

Background and objectives

The National Passenger Survey (NPS) was introduced by the Strategic Rail Authority (SRA) in 1999. The survey was designed to provide reliable data on trends in satisfaction with each Train Operating Company (TOC) that provided franchised services to the SRA. NPS satisfaction ratings were used alongside Public Performance Measure (PPM) data (technical measures of punctuality and reliability).

Fieldwork for the survey takes place in two waves each year: spring and autumn. In each 10-week wave, about 80,000 self-completion questionnaires are distributed and approximately 30,000 passengers return completed questionnaires (each relating to a named trip made by an individual passenger with a specific TOC). The survey was originally carried out by The Oxford Research Agency (TORA), and since 2003 has been conducted by Continental Research (CR).

Passenger Focus is the statutory watchdog for rail passengers. Its mission is to champion the interests of rail passengers and to secure measurable improvements in the quality of rail services in Great Britain. In 2005, Passenger Focus (at that time named the Rail Passenger Council) took over responsibility for the NPS from SRA. This is subject to a memorandum of understanding with DfT Rail which commits Passenger Focus to a number of actions such as consultation on any changes to the survey and maintaining the continuity of the dataset as far as possible.

Having taken over the NPS, Passenger Focus wished to conduct a thorough review of the survey. The broad objectives were to identify if improvements can be made to its methodology and to determine if the results can be used more effectively to drive improvements to train services for the benefit of passengers, while bearing in mind the benefits of maintaining the continuity of the data set so that long term trends remain meaningful.

Passenger Focus asked Peter Bartram, a Fellow and former Chairman of the Market Research Society (MRS), to submit proposals for this project. Peter assembled a qualified project team for this purpose. The team includes: Richard Roberts-Miller FMRS, Mary Bartram FMRS and Gary Bennett MMRS, as well as Peter himself.

In brief, the NPS Review project consisted of a review of a range of documentation relating to the NPS, consultations with a range of NPS users, and a technical evaluation of the survey, followed by the preparation of a series of conclusions and recommendations. Details of the method are included in the Appendix B.

This document reports on the results of the NPS Review.

We would like to thank Passenger Focus, Continental Research and all of the consultees for their co-operation and assistance in completing this Review.

Commentary

1. Study of documentation

The brief suggested that the NPS Review project should consider a range of written documentation, data and information relating to the NPS. This process continued throughout the whole NPS Review project. At the beginning of the project we were provided by Passenger Focus (PF) with a wide range of documentation relating to the NPS. Further documentation was provided by Continental Research (CR). This material was examined by members of the Project Team. This in turn led us at various stages to request further information from Passenger Focus, CR and others, and to carry out extra analysis of the NPS and other data sources.

There is no separate report on this aspect of the project, but the work done contributed to the design of the Consultation Exercise (see Section 2), to the work done in the Technical Review (see Section 3) and to the development of our Conclusions & Recommendations (see Section 4).

For a list of documents reviewed see Appendix E.

2. The consultation exercise

2.1 Our approach

In December 2005, 23 consultation sessions were conducted with a sample of NPS users: 14 were face-to-face sessions (six of those were group sessions) and 9 were conducted by telephone. Two other NPS users volunteered written comments. In all, feedback was received from 39 individuals: 14 from TOCs and 25 from other organisations. For full details of the method see Appendix B. For a detailed list of consultees see Appendix C.

TOCs

Arriva Trains Wales
c2c
Central Trains
Chiltern Railways
First Great Western
First Great Western Link
First ScotRail
Gatwick Express
GNER
Midland Mainline
one
Silverlink
South West Trains
Virgin Trains (CrossCountry)
Virgin Trains (West Coast)
WAGN
Wessex Trains

Other stakeholders

Association of Train Operating Companies
British Transport Police
Continental Research
Department for Transport
Greater Manchester PTE
London Travelwatch (London Transport Users Committee)
Merseytravel
Network Rail Infrastructure Ltd
Office of Rail Regulation
Passenger Focus (Rail Passenger Council)
South Yorkshire PTE
Strathclyde PTE
Welsh Assembly Government

The following section reports on the Consultation Exercise. Selections from verbatim feedback received during the consultations have been included in the following pages as anonymous illustrative quotations (identified as from a "TOC" or "Other" stakeholder).

2.2   How the NPS is used

NPS users evidently vary in the ways they use the NPS. For example: national and regional bodies have differing perspectives; specialist users will have a greater interest in certain questions than in others; some TOCs have contractual NPS satisfaction level targets, others do not. However, the NPS is clearly used widely by those consulted as an important source of benchmark data on passenger satisfaction with rail services. It is used to evaluate trends in performance over time: nationally, regionally and for individual TOCs and it is used to benchmark individual TOCs against national or sector averages.

"The NPS is a common measure irrespective of what's actually in the contract - we've actually got a common mode of measurement of this important element of service delivery across all the TOCs - and with a time series behind it." (Other)

"We see the NPS as a key customer satisfaction survey." (Other)

"Vital . . . it gives us a national picture on the more generic issues, but more importantly for me . . . it gives me a comparison between the different TOCs, and trends." (Other)

The point was also made that the NPS provides information that cannot be provided by punctuality data or similar "PPMs", or by mystery customer surveys.

"It's very important because it covers a lot of what we call soft data attitudinal perception issues - which objective measures of performance can't pick up. Objective measures can tell you something about how punctual the trains have been … but it can't tell you anything about staff helpfulness … people's judgement about whether the railway is providing value for money, whether trains are as clean as people believe they should be - so for these aspects of the travel experience which are all important to end users the NPS is the primary source of information." (Other)

As the NPS is so widely used as a principal measure of perceived performance, it is perhaps not surprising that all TOCs appear to pay great attention to the results (whether or not they have NPS targets written into their franchise agreements).

"If your results are poor you've got to be able to . . . say what you're doing about it." (TOC)

"In terms of customer satisfaction we take it very, very seriously." (TOC)

"TOC MDs are going to hate being seen to be doing less well." (TOC)

TOCs use it as a key management tool - to see how their performance has changed over time, to see how they rate against targets, or national averages, or other TOCs.

"What's very useful is how it benchmarks us against other people." (TOC)

"Without the common survey you wouldn't get very far." (TOC)

"It's increasingly used as a key guideline to how we perform." (TOC)

"The true value of all these things is to help benchmark us against other TOCs, because there are very few industry-wide measures. It's very important to us." (TOC)

It is also used to determine what might be termed "best practice" performance levels, and to drive up performance.

"We look at time series, but then we look at other TOCs . . . to identify areas of excellence really." (TOC)

"There have been occasions in the past when we have been below the sector average . . . we have very much used it *(NPS)* **as an internal tool to drive up the allocation of resources to where the problem was." (TOC)**

Consultees were also asked which other passenger surveys they look at, apart from the NPS. Among other stakeholders the picture varied. Some PTEs conduct their own local research, some national bodies conduct surveys from time to time, but no clear picture emerged. However, all but one of the TOCs we spoke to commission their own customer satisfaction surveys (broadly similar to the NPS) to supplement the data from the NPS. The usual reasons were: (i) to cover certain topics in more detail than does the NPS and/or extra topics, or (ii) to generate a very much larger base size for their TOC than does the NPS (allowing for more detailed analysis, e.g. by station), or (iii) to report more frequently than the NPS (usually 4 times a year rather than twice a year).

"We use *(our own)* **passenger research to cover a greater number of people** *(i.e. larger base than the NPS)* **. . . if we wanted to do it by sector - cross-tabbings and things - we would use our own research." (TOC)**

"We do four of those each year, they give us greater frequency, so instead of just looking at spring and autumn . . . we get seasonality, trends, they give us information faster, early warning. We use them very much as key performance indicators so that we can work (on) delivering greater customer satisfaction. On demographics they're slightly more detailed *(than NPS).***" (TOC)**

"Alongside the NPS we run our own customer satisfaction monitors - at the same time as the NPS fieldwork dates - our survey goes into more detail … for instance if they are dissatisfied . . . we ask follow on questions as to why." (TOC)

"We go into a bit more detail . . . about which ticket window did you you use, why did you travel with us rather than by car . . . and more detailed profiling of our customers." (TOC)

One TOC we spoke to does not supplement the NPS with a parallel survey of their own. In that case they have chosen instead to conduct a very large "mystery customer" programme to provide more detailed feedback on their customers' experiences.

"We . . . have a mystery customer research programme which gives a tremendous amount of specific detail about customer experiences." (TOC)

In fact <u>most</u> of the TOCs we consulted use mystery customer research to supplement the NPS data (as well as their own customer satisfaction surveys) with more detailed information on customer experiences, e.g. to monitor ticket office performance.

"It gives us the full picture - how our actual performance is related to perception." (TOC)

Their generally heavy use of customer satisfaction surveys and mystery customer surveys indicates rather conclusively that TOCs take customer satisfaction very seriously.

However, this level of independent research activity by TOCs does not seem to reduce their interest in the NPS, as all of their own research covers their own TOC only. The NPS is their only source of comparative data on their own perceived performance over time versus other TOCs. It is regarded as invaluable by all the TOCs we consulted.

"Very much so - it is a very, very important research tool for us." (TOC)

"We desperately want the NPS . . . it's amazing how much is in it." (TOC)

This view was shared by the other stakeholders we consulted, almost all of whom stated the NPS to be important, if not essential.

"Essential, absolutely essential. It's what passengers are telling us; it's not our perceptions, it's genuine feedback from passengers on the big issues. And that means when we are tackling issues on behalf of passengers, we can say that we are representing what they tell us are the big issues." (Other)

"It's extremely helpful for us. It's extremely important for us - it's our only *(survey)* source of general rail passengers." (Other)

"This is the only national industry-wide data we have on satisfaction." (Other)

"There is nothing that is of comparable utility to NPS." (Other)

2.3   Which parts of the NPS are used most

The consultation investigated which sections of the NPS are used most. As part of this exercise we used a checklist of sections of the NPS questionnaire to ask which NPS question groups they "use a lot", which "sometimes" and which "never". Detailed responses were received from only 22 consultees, so these results cannot be treated as statistically accurate. However, they are likely to be indicative. Question numbers refer to the Autumn 2005 version of the NPS questionnaire (Wave 13).

The most frequently used sections are those dealing with satisfaction with the journey, the station, the train and the route: more than two-thirds said they use them "a lot":

SECTIONS USED "A LOT"	All	TOCs	Other
Base:	22	10	12
Overall Opinion of Journey Qs 27-28	82%	90%	75%
Opinion of Station Qs 15-18	77%	100%	58%
Opinion of Train Qs 20-21	77%	90%	67%
Opinion of Route Q 19	68%	80%	58%
Security on Railway Qs 32-36	64%	50%	75%
About You Qs 46-51	45%	50%	42%
Ticket Purchase Qs 10-14	41%	50%	33%
Delays Qs 22-26	41%	50%	33%
Train Details Qs 1-3	32%	40%	25%
Your Journey Today Qs 4-7	32%	30%	33%
Disability Qs 8-9	27%	10%	42%
Frequent Users of Route Qs 29-31	23%	10%	33%
Incident Witness or Victim Qs 37-40	23%	10%	33%
General Information Qs 41-45	14%	20%	8%

Most consultees said they looked at all sections at least sometimes.

"Everything is relevant at some point in time, but not all the time." (Other)

"I would probably say I have used all of them at some time or other." (Other)

Among the less used sections, the General Information (Qs 41-45 demographics etc.), the Train Details (Qs 1-3 departure time etc.) and Journey Details (Qs 4-7 journey purpose, party composition etc.) are necessary largely for sample checking and/or for survey analysis rather than as information in their own right.

The only section which most respondents (12 out of 22) said they "rarely or never" looked at was the section on being the witness or victim of an incident (Qs 37-40 on Wave 13). However, for those who did use it, it was considered important, and the more general section on security (Qs 32-36) was used "a lot" by most respondents (64%). Low frequency of use does not necessarily correlate with unimportance.

"Disability is not central to our business . . . but it's important - it's a key interest for those who are interested." (TOC)

2.4   Comments on the NPS Report format and distribution

Probably the widest distribution of the NPS results in the past has been through the National Rail Trends report (which covers a range of metrics including: passenger kilometres, revenue and journeys; punctuality; complaints, etc.). However, the National Rail Trends report contains only a few selected NPS satisfaction results for each TOC and very limited details of NPS methodology and is unlikely to satisfy those wishing to make in-depth use of the NPS data. The main topics of our discussions were NPS TOC Reports and NPS Consultees Reports, rather than the National Rail Trends report.

Each TOC receives an "NPS TOC Report" with results in a standard chart and table format (but with data selected to cover their own TOC's latest results versus national and category averages). These reports also include a brief section on NPS methodology (1 page), a section on issues affecting fieldwork (e.g. 2 pages), plus a section containing a list of stations sampled and weighted and unweighted sample profiles (e.g. 5 pages).

TOCs seemed happy with the NPS TOC Report format.

"It gives you top level performance ... there's not reams and reams. It's quite easy to read … It gives you what you need." (TOC)

"The way it's presented is generally pretty satisfactory. It's easy to go through and pick out the key scores." (TOC)

"It's detailed. It gives an idea of where we are, whether we are either succeeding or in relation to our competitors." (TOC)

A small number of PTEs receive an "NPS PTE Report" with similar format to that of the TOC Report (but shorter and with data selected to cover their own area[1]).

"Pretty good really. No gripes. There's no perfect way of showing information but it seems easy to read." (Other)

Most of the organizations consulted receive the "NPS Consultees Report" which is similar in format to the NPS TOC reports, but covers national, regional and sector analyses (84 pages of charts and 16 pages of tables), plus brief sections on NPS methodology (1 page), fieldwork issues (2 pages), sample composition and weighting (2 pages). Most of those who see the Consultees Report also seemed happy with its format.

"Yes absolutely. It's simple to use, it's in clear English. Very good." (Other)

"I find the report fairly straightforward to use. I can generally find what I need. I think it's quite well presented." (Other)

"Yes, it is user-friendly . . . it's succinct, it's very clear." (Other)

---

[1] NPS PTE Reports are unweighted (as the NPS weighting design relates to the TOC sample design).

The only material criticisms of the format were that some wanted the report to contain more cross-analysis (see section 2.6), and one preferred tables to charts.

"A lot of it is taken up with charts - the numbers could be presented in a much more compact way in a tabular form which would make the whole thing less bulky." (Other)

The normal distribution of the above NPS Reports is, we understand, one or two bound copies per user organisation. Only two users specifically stated that they would like extra copies. Most users were content with one or two hard copies. Some were content because they also receive a copy of the dataset in digital format (a facility that seems not to be widely known). In the case of the TOCs they may have been content because a number of them - as soon as they receive the NPS Report - load the data into spreadsheets and produce a more TOC-specific report of their own for wider internal circulation (for some this a major exercise which requires the relevant staff to drop everything for a few days - such is the urgency with which the latest NPS results are handled).

One TOC user was against the circulation of extra copies of the NPS TOC Report - preferring to circulate their own internal version of the results which presumably includes guidance on the relative importance of various changes in the scores.

"I would have sleepless nights if there were too many floating around … we had people reacting to blips … and how important are these?" (TOC)

We also invited comments on the speed of publication of NPS results. Most were reasonably happy with the turnround time.[2]

"I think now it's pretty timely." (Other)

"It would be better if it was quicker but I'm not sure if that's practicable." (Other)

"I can contain my excitement for that length of time." (Other)

"Quite good. I would prefer it if they could turn it round quicker. (TOC)

"OK. Faster would be good. Knowing when it's going to come is important." (TOC)

"Business always wants the information faster. The NPS - I think it finished on the 7th of November - we will get it *(advance top-line figures)* on the 19th of December - that's not bad for the size of survey." (TOC)

2.5   Comments on the NPS coverage and content

Most of the consultees expressed the view that the existing NPS questionnaire covers all the necessary topics.

---

[2] Fieldwork for NPS Wave 13 ended on 18th November 2005. The TOC Reports were circulated on December 19th and the Consultees Reports were circulated on January 20th.

"The list of topics is fine." (TOC)

"Nothing obvious is missing." (Other)

A few suggested additions to the regular questions, but there was no consensus.[3]

Several suggested that the questionnaire was long enough already.

"Nothing is missing . . . Sometimes we feel there is too much." (TOC)

"I would be worried if questionnaire was any longer." (TOC)

"There are questions here that - if it was up to me - I just wouldn't ask." (TOC)

"Keep it as a customer satisfaction survey. Don't be tempted to add in marketing questions." (TOC)

"That's been stretched over the years and I'm sure there is a good reason for all of them but I think that no harm is done by sitting down periodically and considering whether all the questions are still earning their keep - one that occurs to me immediately is that there used to be one about train cleanliness. Because it wasn't obvious whether that was about interior or exterior cleanliness, two new questions were put in to make that distinction but the old question got left in . . . that's an example of one that can be retired if the others stay there . . . there may be others." (Other)

Two consultees suggested that the NPS does not meet all their information needs, but a third volunteered that there were limits to what the NPS could do in this regard.

"E.g. '70% of people are satisfied' . . . the NPS is seen as example of a survey that produces that type of very high level information without contributing to the in-depth understanding." (Other)

"The narrative is entirely limited to remarks about the weather and strikes and things that might have affected the survey - what they don't do is . . . offer any narrative commentary about what's going on here, what might explain this, what is interesting, what conclusions we should draw from all this. What the NPS does is to throw up a lot questions - but it doesn't answer them." (Other)

"There are other information needs that the NPS can't really capture." (Other)

Broadly, all the TOCs and the majority of other consultees endorsed the present NPS as providing a measure of all the key elements of rail customer satisfaction.

---

[3] The only specific suggestions (none mentioned by more than one individual) were: quality of ride, how people got to the station, where they came from (e.g. postcode), connections with public transport, their expectations, more on ticket sales, more on provision of train timetable information, and a broader question about how delays are handled.

2.6   Comments on the level of detail of cross-analysis reported in the NPS

Some consultees said that the reports do not contain enough information.

"We are . . . not sent a basic set of tables breaking down the results _(if I were)_ I would be able to make far more use of the data." (TOC)

Some were satisfied with the level of detail as it is.

"It seems fine to me." (Other)

"I suppose there always could be more . . . you can't meet everybody's needs in the one document because it would be huge." (Other)

"Just provides us with an overview - as a tracking study it's good." (TOC)

Others were equally satisfied with the published level of detail, but this was in the context of receiving extra analysis or datasets which gave them a lot more NPS data.

"I will be receiving the dataset. If you ask for it, it is there." (Other)

"We don't get the raw dataset but I understand it is available. We have had additional analysis in the past." (TOC)

"No problem getting other information - special analysis. We get it electronically as well as in hard form." (TOC)

Most TOCs had used the facility, but at least one TOC consultee and several of the "other" consultees appeared unaware that these options existed.

"We are sent hard copies and it really needs to be electronic." (TOC)

"If you could ask for particular cross analyses you want that would be of use." (Other)

"I would like to have an electronic version of results." (Other)

"Electronic output? I've never seen it. If it exists that fact is not headlined." (Other)

One suggestion re: access to detailed data was that access to open-ended responses would be useful.[4]

"We would like to see the comments from the open-ended questions." (TOC)

The NPS report produces averages for three categories of TOC: "High Speed Long Distance", "Regional Services" and "London & SE" (as well as producing overall national averages). We asked

---

[4] TOCs currently receive a CD-ROM of verbatim comments from Q45 only.

the TOC consultees which of the three categories represented the best yardstick for comparison with their own TOC. Most said that their network included more than one of the three types of service.

"Regional. But we have a lot of short-distance commuting too." (TOC)

"Interesting question. How do you differentiate between the two - because they are all one company." (TOC)

"Too simple. Each . . . varies quite a lot, all have some of each characteristic." (TOC)

As there has been a tendency for TOCs to become grouped into larger and fewer entities, "mixed TOCs" have increasingly become the norm. While this means that few if any TOCs precisely match one of the three yardsticks, this does not seem so far to have become a controversial issue, as the three categories are well-understood and seen as useful analysis tools in any event. Some TOCs carry out extra analysis to produce more relevant comparisons - or at least have considered the possibility.

"Our TOC is mixed." *(Is that a problem?)* **"No because we segment the data out. We do the analysis and compare our mainline to mainline and so on." (TOC)**

"One figure for our company is great, but to manage our business we need to break it down. I can get it from Continental Research - it's available." (TOC)

"We would like to compare the results for our passengers on a route with *(a rival's)* **passengers on the same route." (TOC)**

"We would like to be able to segment our data - what we want from the NPS is to be able to to look at our company data - by market segment and line of route." (TOC)

In this context we believe it may be useful to consider adding one or more extra analysis breakdown columns to the standard NPS report tables, to refine these comparisons.

2.7 Comments on the frequency of the NPS

About half the consultees were happy with the current frequency i.e. two fieldwork waves and two reports per annum.

"I doubt whether the additional insights that would be provided by upping the frequency to quarterly would be commensurate with the cost - the numbers don't change dramatically six months on six months, or year on year. Unless there is a powerful reason for changing it I would leave it as it is. " (Other)

" I think twice a year is probably just about right. A six-month period gives a TOC enough time to adjust if a particular issue is perceived as bad." (Other)

"I think that's about right." (TOC)

Most of the rest (including most TOCs), would prefer frequency to increase.

"If there could be a third phase in the annual cycle it would be good." (Other)

"It needs to be spread more across the year. It should have 3 or 4 fieldwork periods in the year rather than just two." (Other)

"We would prefer 4 waves a year rather than two - or rolling data." (Other)

"Having data all the year round may be useful in terms of ruling out any seasonal patterns." (Other)

"I would prefer quarterly - a rolling sort of programme - with quarterly reporting." (TOC)

"It would be great if they did it four times a year - we've got autumn and we've got spring but what about summer and the winter - are there differences there?" (TOC)

"It would be quite useful if it came out quarterly. That would definitely be useful." (TOC)

"Four per annum would be good." (TOC)

"More frequent polling, with fieldwork to cover winter and summer as well as autumn and spring - this would gain fuller understanding of the changes in customer satisfaction issues over the course of the year." (TOC)

Few would be pleased if the survey became annual.

"No, I'd prefer to stay with two." (Other)

"Once a year is not enough for us." (TOC)

"Two per annum is a bare minimum - because of seasonality." (TOC)

Some said they might be happy to see some money spent on ad hoc or diagnostic research going into various subjects in greater depth (but in most cases not if this was at the price of a reduction in NPS frequency). However, the only possible topic for occasional ad hoc work which was mentioned more than once was the relative importance of the satisfaction parameters measured.

"The importance of satisfaction values needs to be checked from time to time." (TOC)

2.8 Comments on NPS survey design and method

Most consultees had no spontaneous criticisms of the survey design or method.

"No, none whatsoever." (Other)

Most were happy with the self-completion method employed and the response rates achieved (approx. 40%).

"That works quite well, because you can put down exactly what you feel. 40% isn't bad." (Other)

"I'm not sure I know a better way. 40% is quite good." (Other)

"I think that's quite a good way of getting the results - it gives the customer the choice of filling it in at a later period - especially with shorter journeys." (TOC)

"Distribution at the station is probably the most logical way to do it. It's probably the best way to deliver it - then you can be much more clear about where people started their journey. 40% is a phenomenally good response rate for something that is self-completion, it is quite a lengthy survey, and there's no incentive." (TOC)

One consultee pointed out that the response rate itself is not the only relevant factor here, and wanted to see an analysis of the non-response bias and refusal rate.[5]

"It's not necessarily response rate - it's the non-response bias. And I don't know what the bias is . . . there's a refusal rate as well." (Other)

Most were happy with the questionnaire design (although a few repeated that they were worried about its length). On the topic of sampling and weighting the most frequent comment was that they would like larger sample sizes.

"I would prefer a greater sample in this area because if I had that I could do more with it." (Other)

"Bigger sample size would be nice." (TOC)

"The only limitation is the sample size, I wouldn't say it's a criticism - but it's a limitation on how much we use it." (TOC)

Most were happy with the NPS sample design at national level and TOC level.

"That seems perfectly sensible to me." (Other)

"We are happy. It covers all the big stations and a good selection of the rest." (TOC)

---

[5] The response rate for the NPS is calculated by dividing the number of questionnaires received by the number distributed. Any passengers refusing to accept the questionnaire will not have been included in that calculation.

"I think it is adequate to represent our business." (TOC)

Some mentioned that the NPS was less useful at the local authority level, or station or route level (this is due partly to the sample sizes per TOC and partly to the survey design - as it was designed to produce a representative sample of journeys on each individual TOC <u>as a whole</u>).

"We don't find *(the results)* **particularly useful in terms of the sample size when you start getting down to our area." (Other)**

"NPS works for macro (i.e. whole TOC) but not micro analysis." (TOC)

Most consultees had no comments at all on the NPS weighting.

"Don't really know the ins and outs of that." (Other)

Some said they had no problems with the weighting.

"We have certainly never had issues with sampling and weighting." (Other)

"I've had no queries . . . I'm not sure exactly how it is done. It's not an issue for me." (Other)

"No problems." (TOC)

One said that the weighting method was reasonable, but might need to be updated.

"I've a feeling it should be updated . . . It's reasonable - from the station work I've been doing - its not bad." (TOC)

One suggested that the description of the weighting should be made clearer.

"There's quite a lot of weightings in the report and there's references to the weightings - I think it needs to be made much clearer how these weightings are done and what they are for . . . you need to know a bit more than is currently made overt about that." (Other)

Three other consultees suggested that the published information on the NPS methodology <u>as a whole</u> was too limited.

"The methodology section is just a page - I think it could be expanded on - to be clearer - I would appreciate that." (Other)

"Information collected at public expense should be freely available to the public - full information on the methodology . . . should be more openly accessible." (Other)

"The NPS is perceived as a bit secretive . . . all of the details aren't released." (Other)

Another suggested that this situation is better than it used to be.

"In Rail Trends - there's absolutely no information about the *(NPS)* method used - it used to be impossible to find out the response rate, how the stations were sampled. That's changed - the dissemination *(then)* was through National Rail Trends, but that's not the case any more." (Other)

2.9   Comments on NPS accuracy, reliability and credibility

We asked for feedback on the accuracy, reliability and credibility of the results produced by the NPS. Two-thirds of the comments we received on this topic were positive.

"I've got no reason to doubt them whatsoever. They seem to be sensible." (Other)

"It has never been questioned in the time that I have been here. We rely on it to be accurate." (Other)

"I think they're okay. I've not come across any inconsistencies." (Other)

"We've had no reason to question the accuracy and integrity of the research. Had there been any issues . . . they would have come to our attention." (Other)

The TOC representatives typically said that where they were able to compare NPS data with data from their own research, the results matched.

"The trend lines follow each other (ours and NPS) . . . it correlates with our own data. NPS is reliable - I don't recall unusual results which couldn't be explained . . . it has not produced answers which cannot be explained overall. Looking at individual stations you may have an issue. It may be a bit less reliable at the station level. But NPS is a good survey - correlates well." (TOC)

"*(NPS results are)* reasonably close to the results we get from our own surveys. I don't think in terms of the sampling there's any great issue. The stations are stratified. The range of stations they're using is pretty reasonable." (TOC)

"We get very, very similar results to NPS. As a tracking study it's absolutely adequate for what it was designed to do - we just look at it on an overall level." (TOC)

"There is definitely a correlation between the NPS and our own results." (TOC)

"We look at the data - it tends to be consistent with our data. When figures haven't gone up we know why. When it goes up we believe we know why." (TOC)

There were only two comments to the contrary, and they related to question wording. The second of the two was from a TOC representative who is planning to bring their question wording more closely in line with that of the NPS. [6]

"The only time NPS hasn't correlated with TOC results has been when question was different . . .   I hold it *(NPS)* in high regard. A lot of work has gone into the development of it since 1999. I perceive it to be a valuable and important survey. It shows … consistency. In a very sceptical environment, I've never had any criticism of its integrity." (Other)

---

[6] The NPS uses semantic (i.e. verbal) scales in its satisfaction questions. Some TOCs use semantic scales in their own surveys and some use numeric scales (of varying lengths). This was the only point at which any discussion might have arisen about the relative merits of numeric and semantic scales, but it appeared to be an issue of very limited interest. Most TOCs seemed content to continue to employ whichever method had originally been chosen, for the sake of historical comparability. None made any criticisms of the scales used in the NPS.

"The NPS figures sometimes vary from ours but I think that's more down to questionnaire design. Over time we are bringing them closer together so that will harmonize - so I would expect to see a greater correlation." (TOC)

2.10 Other comments on the NPS

One theme which emerged at several points was that most stakeholders regard as one of the greatest values of the NPS its consistency of approach - across all TOCs and over time. As a result, although this Review seemed to be regarded as a useful exercise, many consultees clearly hoped that it would not result in any changes which undermined the consistency of the NPS.

"I do want the consistency maintained. You've got to occasionally update things and change with the times, but one of the things that has been to the benefit of the survey is its consistency." (Other)

"Its strength is in its history." (Other)

"One of the things that has been to the benefit of the survey is its consistency." (Other)

"Don't change it unless it really makes sense." (TOC)

This is not to say that all changes should be barred, e.g. one change which was felt to have been for the better was the extension of the fieldwork period from 3 to 10 weeks.

"The survey wave is now ten weeks which is pretty good (i.e. 20 weeks per annum)." (Other)

Another theme which emerged at several points was that some stakeholders feel that the level of information provided about the NPS has been less than they would like. This was reflected, for example, in criticisms of the limited amount of methodological data published (as already mentioned), in the limited awareness of extra analysis and database options (as already mentioned), and in the lack of advance information about changes to the questionnaire or about the extra sections occasionally added to the questionnaire.

"We don't know what's coming up there and it's not highlighted very well in the report that we get." (Other)

"There's so much in there - but no attempt is made to produce a fact sheet saying 'this is what you can get from the NPS, this is what you can do'." (Other)

In this general context, improved communication would be beneficial, and it was also suggested more than once that some type of user group would be a helpful forum.

"A consultation group where members of the TOCs can get together and we can talk about the survey that's just been done or proposals for the next questionnaire - little workshop groups - so we can spread the knowledge - getting that input there. Also being able to use a lot more of the data there." (TOC)

2.11 Overall assessment of the NPS

When asked how happy they were with the NPS generally, most of those consulted gave positive responses.

"It's fine. It's easy to understand. I don't have any problem with it." (Other)

"I think it's very good - it's a good vehicle." (Other)

"Overall, very satisfied." (Other)

"The executive level staff definitely buy into it." (TOC)

"As a broad indicator of general satisfaction we consider it as a valuable tool which has relevance to us both externally (since it tells the public how we are doing in the context of other operators) and internally because it gives us an objective method of assessment . . . we are very happy that we are surveyed." (TOC)

The majority view can perhaps be summed up by saying that - while there were minor criticisms of some specific aspects of the NPS - most of those consulted regarded the NPS as a unique and important benchmark measure of key elements of rail performance, and one that is regularly used as valuable input to their decision making processes.

Technical review – sampling and weighting

3.1 Introduction

A thorough review was conducted of the both the sampling and weighting methodology using survey documentation and other materials provided by Passenger Focus and Continental Research. In undertaking this review we have considered:

Around the topic of Sampling:

(a)   the nature and provenance of the information provided to Continental to draw interviewer shifts

(b)   the methodological robustness of the sampling procedure employed

(c)   the correspondence between the fieldwork procedures and the method used for sampling

(d)   Potential distortions in the sampling/fieldwork methodology

Around the topic of Weighting:

(a)   The overall efficacy of the current weighting procedure

(b)   The design effects of weighting (on statistical robustness of sample)

(c)   Potential redundancy among weighting variables

(d)   Potential additional weighting variables

"Making the sample more representative" is usually cited as the main justification for very complex weighting schemes. However, it is not so well known that there is an inverse relationship between the statistical robustness of the sample and the complexity and magnitude of the weighting employed at analysis.

It is usually desirable, to ensure that the unweighted sample is as representative as possible across all possible cuts of data, to try to keep weighting to a minimum. This minimises the "design effects" of weighting which in turn allows a greater degree of statistical robustness to be attached to the analysis.

We have also included supporting information on margins of sampling error (based on analysis of the dataset for Wave 11 of the NPS).

3.2 Sampling rationale

To begin it is worth restating the outcomes which are required from the sampling process. The objective of the NPS Sampling Methodology is to provide a representative sample of passenger journeys on each TOC. The target sample sizes are pre-set for each TOC based, we understand, on the size and complexity of each (most have NPS sample sizes of between 1000 and 1500, but range from about 300 for the small Island Line to about 2000 for the much larger SouthWest Trains). The purpose of the sampling frame is to produce parallel, representative samples of journeys on each TOC[7].

The sampling design currently employed can be thought of as a two-stage cluster sample, where the first stage consists of selecting clusters of individual stations, and the second stage consists of clusters (within selected stations) by time of day and day of week. Within the sampling frame, the Sampling Points selected are generally three hour fieldwork shifts scheduled at specific stations, days of week and times. Within these fieldwork shifts, the Primary Sampling Units are passengers making journeys.

Assuming that the questionnaire return rates can be engineered by Continental to remain approximately constant across fieldwork shifts, the sampling method needs to ensure that the fieldwork shifts selected for each wave of the NPS produce a suitably representative sample of passenger journeys on each TOC by:

- Station
- Day of week
- Time of day

In this way, sampling can be assessed at two levels: (a) achieving a representative sample of clusters (stations and times) and (b) achieving a representative sample of journeys within those clusters. The first level (a) is achieved via the sampling methodology employed for selecting fieldwork shifts. The second level (b) is achieved via the fieldwork methodology employed to select respondents making typical journeys within the shifts.

Critical to this process is:

- The integrity and provenance of the data used to represent the universe of trips by station, TOC, day of week and time of day. The quality and appropriateness of this data directly affects the quality of the sampling frame.

- The definition of a "journey". For instance a journey could be defined in various ways: (i) a complete one way trip, (ii) a complete round trip (outbound and return) or (iii) a trip leg – only one component of a trip (either of the previous types) before or after which the customer may have had to interchange on the system. The methodology employed needs to ensure that the particular definition used in constructing the sampling frame matches the definition assumed in the fieldwork methodology and questionnaire.

---

[7] Setting specific target sample sizes for each TOC can lead to big distortions in any analysis across TOCs, requiring weighting at analysis stage. High deviations in the distribution of target sample sizes from the true distribution across TOCs require greater weighting at analysis (see section on weighting).

- The use of an appropriate sample frame and selection procedure for the clusters of sampling points – the stations at which fieldwork shifts are to be conducted and the days/times of the shifts.

- A flexible fieldwork methodology which is sensitive to peaks or troughs in journeys made across fieldwork shifts.

Also, it is important to demonstrate that the sampling methodology employed is, as far as possible, independently managed and free from possible manipulation by TOCs. As the NPS is used to assess TOC performance, this is integral to the survey's credibility.

3.3   Universe data - number of journeys

The universe data used to construct Continental's sampling frame came from two sources. For numbers of journeys per station it is the CAPRI/Lennon database, the rail industry's central ticketing system[8]. This database is an objective source, independently compiled and therefore not subject to potential manipulation. The figures include interchangers and therefore the implication is that a journey is any "leg" of a trip. According to this database, around 14% of "legs" across the whole rail network result from interchanges rather than new trips. From this point onwards when we use the term "journey", this will mean "one leg of a trip". This is ultimately the Primary Sampling Unit which has been assumed.

We have not seen any evidence which leads us to doubt the quality of the CAPRI/Lennon database of journeys by station. It is the main source available and, as demonstrated by Continental's sampling methodology for shifts at stations as a whole, reconciles well with other data provided by the TOCs. The only issue we were not able to clarify was the extent to which the TOCs managing specific stations were able to alter these independent estimates. Continental's note on their revised Sampling Design for Wave 11 onwards indicated that some TOCs substituted the initial estimates with estimates of their own, though detailed records of these changes were not kept. This might be worth revisiting to establish the rationale for any changes and the extent to which the substituted data had better provenance than the CAPRI/Lennon data, in order to rule out possible manipulation[9].

The CAPRI/Lennon database of journeys is, we understand, provided in aggregate over the year and is not reported by time of day or day of week[10]. Therefore a separate source of data is needed to split the journeys by time of day and day of week.

---

[8] CAPRI was used from 1999 until 2003, when it was superseded by the Lennon database system.
[9] Even if this is thought unlikely, as is the case, putting such checks in place can only improve the credibility of the sampling plan. At the very least, records of such changes should be retained and made available for inspection as part of the essential technical description of the survey design.
[10] We understand Lennon data is not analysable by time of day; although it may be possible to extract Lennon information by individual date, we do not know how practicable it would be to use this as a source of day of week analysis.

3.4   Universe data - day and time

The source which is currently used to determine the split by day of week and time of day is, in our view, the weakest part of the current process. Rather than by using an independently verifiable source as in 3.3, a split was generated by time of day and day or of week using data from Waves 1-8 of the NPS. This is highly circular and carries forward any biases which may have existed in the previous waves. The estimates used by Continental which apply a generic split to City Centre and Other Stations, for weekdays (on average) vs weekend days (on average)[11] are shown below:

Volume Estimates for City Centre and for Other Stations

Centre	%	%	%
of day	weekday	weekend	Total
- 1000	8.02	0.33	8.35
- 1300	19.48	15.88	35.36
- 1600	22.01	5.91	27.91
- 1900	25.32	0.37	25.69
- 9999	2.52	0.16	2.68
	77.35	22.65	**100.00**
r Stations	%	%	%
of day	weekday	weekend	Total
- 1000	48.73	0.51	49.24
- 1300	27.93	10.78	38.70
- 1600	5.98	0.79	6.77
- 1900	4.99	0.04	5.03
- 9999	0.26	0.00	0.26
	87.88	12.12	**100.00**

Source: Continental Research.

Neither CR nor Passenger Focus were able to provide us with documentation on how the estimates were originally produced for Waves 1 to 8. An analysis of Wave 13 NPS data demonstrates that these estimates must be incorrect. On an *a priori* basis, most journeys on the rail network must be round trips (i.e. include an outbound and a return journey). We expected to see a ratio of approximately 50:50 between outbound and return journeys, but on NPS Wave 13 the split was 64:32 in the weighted sample[12], a ratio of exactly <u>2 to 1</u>. This cannot be correct, but the situation has evidently existed since the NPS began:

Wave:	1	2	3	4	5	6	7	8	9	10	11	12	13
	%	%	%	%	%	%	%	%	%	%	%	%	%
OUTWARD	68	67	66	66	66	67	66	68	64	63	63	62	64
RETURN	26	27	28	27	28	28	28	29	33	34	34	34	32
ONE WAY	5	5	5	4	5	4	5	2	3	3	3	3	3
DK/NS	1	1	2	2	1	1	1	1	1	1	1	1	1

Source: NPS weighted data, CR, 2005.

Without information on how the original NPS shift pattern was developed, we cannot determine how this anomaly arose. However, it is reasonable to assume that two of the original priorities were to

---

[11] We believe it may be acceptable to treat all weekdays as average weekdays for this purpose, but that it would be preferable to target weekend days separately (as Sat v Sun may differ more than e.g. Tue v Wed).
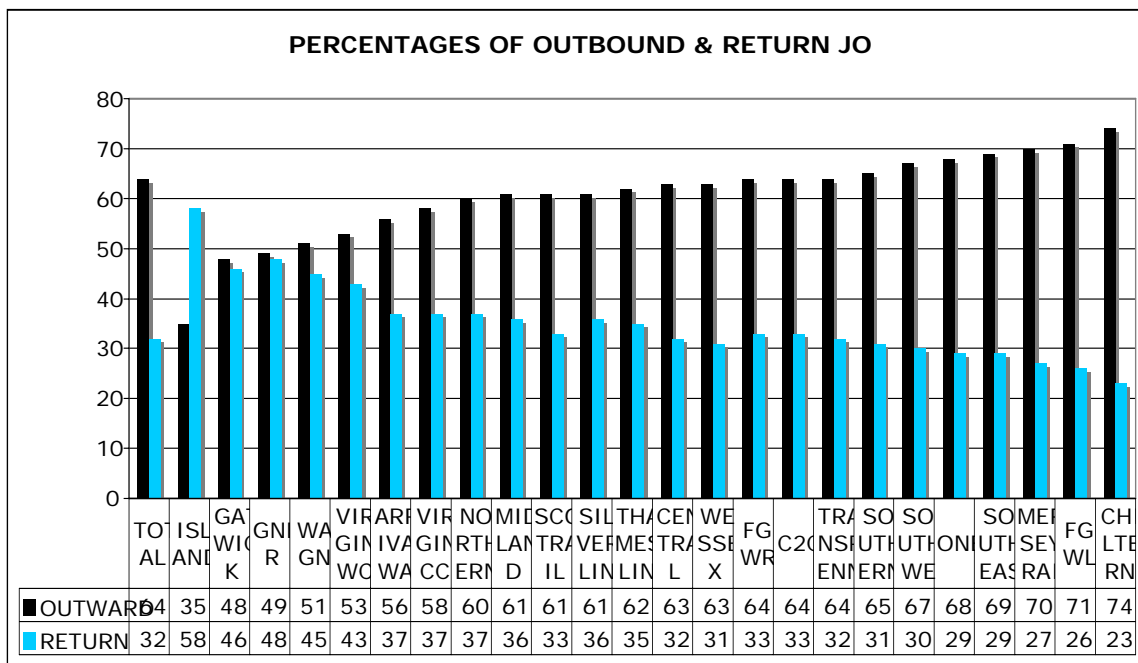
[12] The unweighted equivalent is similar at 62:34.

achieve (i) a representative ratio of peak v off-peak traffic, and (ii) a consistent and stable design over time.

Whether or not that was the case, analysis of Wave 13 suggests that the NPS sample does include a reasonable ratio of peak time journeys <u>in total</u> (i.e. a.m. + p.m.) to off-peak journeys (as far as we can judge without further work).

However, the shift pattern adopted to achieve this over-represents the morning peak traffic significantly (i.e. typically outward from home) and under-represents the afternoon peak traffic significantly (i.e. typically inbound return) and has done since since Wave 1, and the consistent approach to shift sampling has allowed this to persist.

Also, the bias evidently varies by TOC. Analysis of Wave 13 shows that 20 of the 24 TOCs had outward and return percentages which differed by 10% or more. Only three had roughly balanced percentages (Gatwick Express, GNER and WAGN):

PERCENTAGES OF OUTBOUND & RETURN JO

	TOTAL	ISLAND	GATWICK	GNER	WAGN	VIRGIN WC	ARRIVA WA	VIRGIN CC	NORTHERN	MIDLAND	SCOTRAIL	SILVERLINK	THAMESLINK	CENTRAL	WESSEX	FGWR	C2C	TRANSPENNINE	SOUTHWE	SOUTHONI	SOUTHEAS	MERSEYRA	FGWL	CHILTERN
■OUTWARD	64	35	48	49	51	53	56	58	60	61	61	61	62	63	63	64	64	64	65	67	68	69	70	71	74
■RETURN	32	58	46	48	45	43	37	37	37	36	33	36	35	32	31	33	33	32	31	30	29	29	27	26	23

Source: Analysis of NPS Wave 13 weighted data, CR, 2005.
Figures do not total to 100% because "one-way trip only" and "DK/NS" replies are not shown.

One TOC (Island Trains) had an inverted bias, with returns outnumbering outbound. However, Island Trains is anomalous: because it is a small network with correspondingly few passengers (especially outside the summer period, which is not covered by the NPS), and the Island Trains fieldwork procedure differs from that used on other TOCs, this result may be due to separate factors.

The main observation to take from the above table is that all but four of the TOCs listed exhibit a highly significant bias towards outward passengers.

We went on to analyse the outward:return ratios of NPS journeys for each hour from 7am to 10pm. The proportions of outward and return passengers change throughout the day, with outward peaking above 80% in the three hours between 7am and 9:59am, and the return percentage peaking above 80% in the four hours between 4pm and 7:59pm:

OUTBOUND AND RETURN % BY HOUR

——— OUTWARD   —O— RETURN


	0700-0759	0800-0859	0900-0959	1000-1059	1100-1159	1200-1259	1300-1359	1400-1459	1500-1559	1600-1659	1700-1759	1800-1859	1900-1959	2000-2059	2100-2159
OUTWARD	96%	94%	89%	77%	65%	53%	39%	29%	23%	12%	11%	14%	14%	18%	17%
RETURN	3%	3%	8%	18%	30%	42%	55%	66%	72%	85%	87%	83%	83%	80%	79%
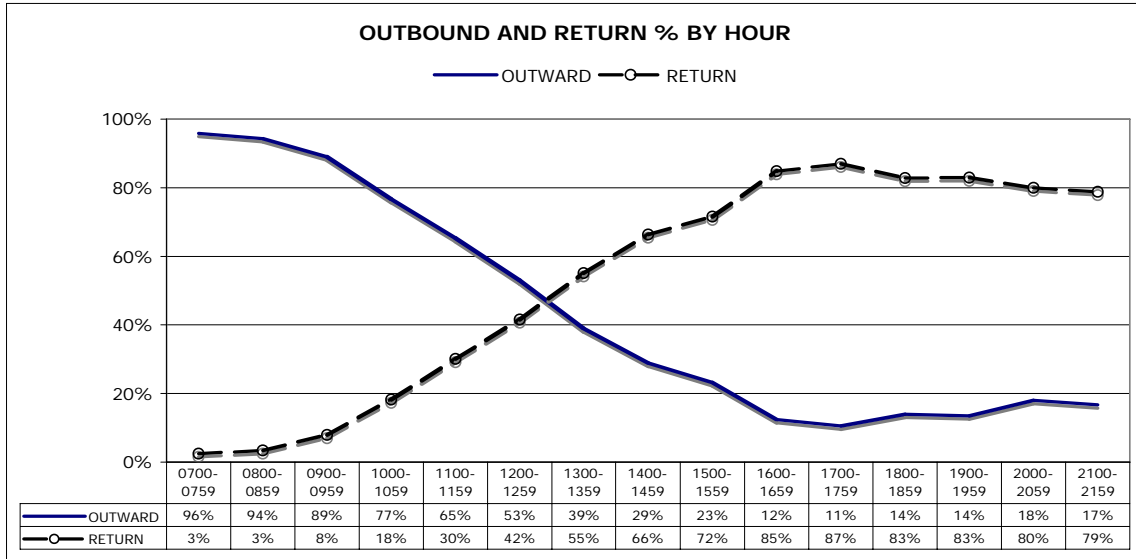
Source: Analysis of NPS Wave 13, CR, 2005.
Note: the balance of 100% is made up of one way trips (average 3%) and those who did not reply (av. 1%).

To be representative, the NPS sample should include correct proportions by time of day for both city centre and other stations. However, our analysis of Wave 13 shifts (by hour) suggests that about 70% of NPS fieldwork hours take place before 1pm:

SHIFT HOURS WORKED BY HOUR (NPS Wave 13)
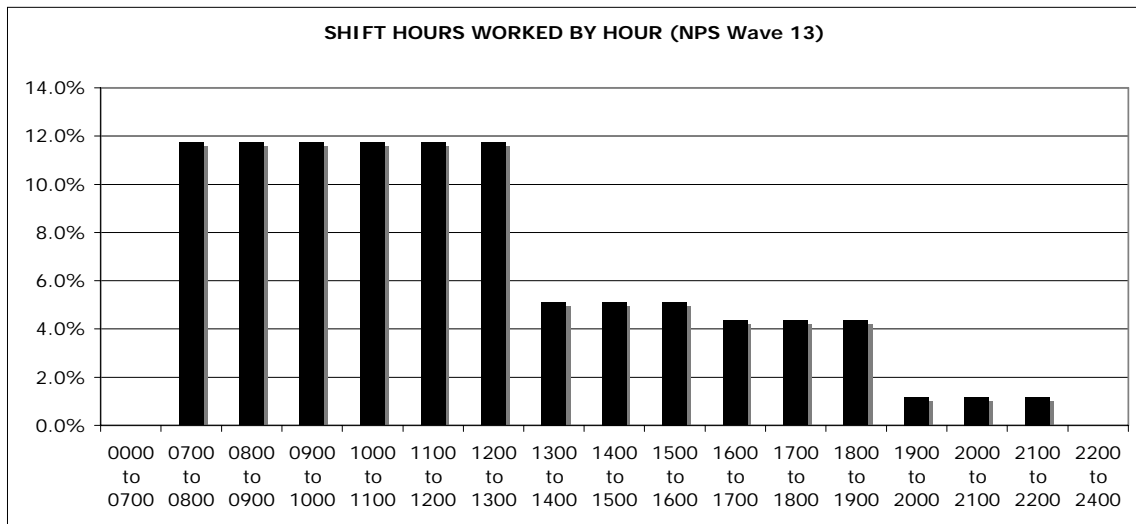
Source: Analysis of shifts, from NPS Wave 13 Sampling Plan, CR 2005

This shift pattern is inherently likely to oversample morning departures, and analysis of Wave 13 journeys by hour of departure confirms that this has occurred.

Morning (and thus also outward) journeys are being significantly oversampled and the afternoon (and thus also return) journeys are being significantly undersampled:

NPS JOURNEYS BY HOUR OF DEPARTURE

A bar chart titled "NPS JOURNEYS BY HOUR OF DEPARTURE" showing percentages by hour of departure:
- Before 0700: 0%
- 0700-0759: 14%
- 0800-0859: 16%
- 0900-0959: 9%
- 1000-1059: 9%
- 1100-1159: 11%
- 1200-1259: 9%
- 1300-1359: 6%
- 1400-1459: 6%
- 1500-1559: 5%
- 1600-1659: 5%
- 1700-1759: 6%
- 1800-1859: 4%
- 1900-1959: 1%
- 2000-2059: 0%
- 2100-2159: 0%
- 2200 or later: 0%
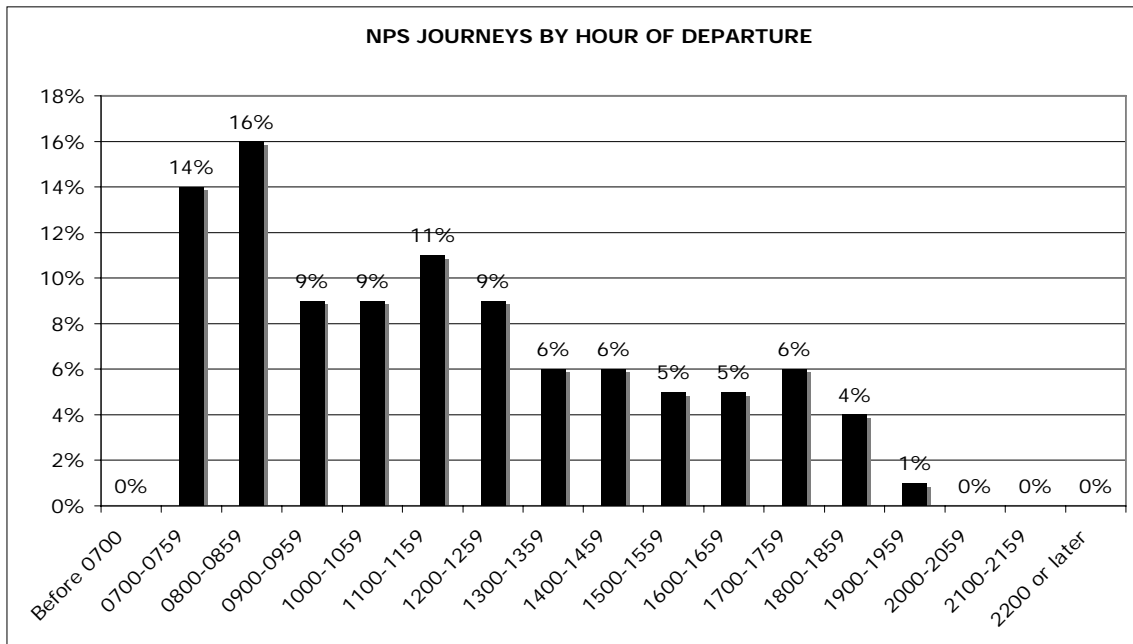
Source: Analysis of NPS Wave 13 unweighted data, CR 2005

As this table shows, 39% of NPS sample journeys departed in the peak outward period of 7am to 9:59am but only 16% in the peak return period of 4pm to 7:59pm. Further investigations were carried out, comparing data from the NPS with estimates of counts by station obtained from the London Area Transport Survey (figures from which were provided by Passenger Focus), and with Rail Planner data (from a special analysis of national rail schedules for the same period as the NPS Wave 13 fieldwork, commissioned from Travel Infosystems). Due to timetable constraints, we were unable to look at more than the overall data, plus data for a number of selected stations. Although the pattern varied somewhat between individual stations, the general situation we had identified was confirmed: the NPS significantly over-represents morning/outward journeys and significantly under-represents afternoon/return journeys. We also examined the effects of this bias on other aspects of the sample profile. It may be contributing to the under-representation of under-35s (and appears to affect the ratios of commuter/business/leisure travellers). Correcting the am/pm ratio may help to correct the known response shortfalls among young travellers (and may help improve other aspects of the sample profile) thus possibly reducing the need for weighting, which is a valuable goal.

Also, NPS fieldwork allocated to the 7pm to 10pm shift appears to contribute very little to the sample (and next to nothing after 8pm), and this situation should be reviewed. [13]

We have not examined the distribution of shifts or response rates by fieldwork week -ideally these should be consistent - but we have no reason to suppose that these vary in a way that affects the NPS results (although it is expected to be the case that the last few days of the fieldwork may be occupied largely with "booster" shifts).

---

[13] The extension of interviewing hours beyond 7pm from Wave 8 onwards had a negligible impact on the outward:return ratio (as only 1% of Wave 13 questionnaires relate to departures after 7pm). The main shortfall in return trips evidently arises between 4pm and 7pm.

We also looked at the possible effects of the am/pm bias on the NPS satisfaction ratings.

Q27 Overall satisfaction with journey.

	TOTAL	Outward	Return	One Way	DK/NA
Sample size	27581	17164	9444	817	156
	%	%	%	%	%
VERY SATISFIED	34	34	34	37	38
FAIRLY SATISFIED	46	46	47	46	48
NEITHER SAT. NOR DISSAT.	12	12	11	11	9
FAIRLY DISSATISFIED	6	6	5	4	3
VERY DISSATISFIED	2	2	2	2	1
NO OPINION/ DON'T KNOW	-	-	-	-	-
Summary:					
SATISFIED (V+F)	81	80	81	83	87
DISSATISFIED (V+F)	8	8	7	6	4
Standard Error	0.238	0.306	0.402	1.304	2.741

Source: Analysis of NPS Wave 13 (unweighted), CR, 2005.

At the national level, outward and returning passengers do not appear to differ significantly in their "overall satisfaction with journey" scores. Because of this, we believe that action taken to correct the am/pm imbalance is unlikely to affect the overall satisfaction score significantly at the national level.[14] Also, most of the individual evaluation questions on the NPS appear unlikely to be affected significantly by correcting the am/pm imbalance - but there are two exceptions. Taking the "good" scores for each factor among outward and return passengers, averaging them and then comparing them with the Wave 13 average (and assuming that this algorithm gives an approximation of what will happen if the sample design is changed to reflect the correct ratios of outward and return journeys) we can say that two ratings are likely to be affected: "car parking" (expected to drop by 5 points) and "bicycle parking" (expected to drop by 2 points):

Projected effect of balancing outward v return journeys

	FROM CURRENT NPS DATA			Projected
	TOTAL	Outward	Return	Rating Change
Sample size:	27581	17164	9444	
"very good" + fairly good"	%	%	%	
Facilities for car parking	50	52	38	-5
Facilities for cycle parking	47	49	41	-2

Source: Analysis of NPS Wave 13 data (unweighted), CR, 2005.

The reason is the difference in the ratings the outward and returning passengers gave to the two factors - which in turn is probably due to the fact that far more return journeys start in city or town centres (where there are usually fewer parking facilities than in most suburban stations).

At the national level, none of the other 30 main evaluation factors we examined appeared likely to vary by more than the normal confidence limits (e.g. plus or minus 1%). We cannot say whether or

---

[14] This would be much less likely if the total (am+pm) peak to off-peak ratio were seriously wrong.

not this is also likely to hold good at the TOC level without repeating the exercise for every TOC, which was not possible in the time available.

To sum up, the data upon which fieldwork shifts are allocated to specific timeslots is deficient. Not only is the original source of the data unknown, but it clearly leads to a split of outbound vs return journeys which is significantly skewed towards the former for the majority of TOCs. The journey distribution by time of day is also inconsistent with data from the LATS survey and Rail Planner, both of which are believed to be reliable benchmarks. The <u>Monday to Friday</u> shift pattern <u>within TOC</u> should be revised to reflect, for (a) city centre and (b) other stations, a more representative mix by hour of departure. Crucially, it must reflect the large differences in patterns between the morning-outbound-suburban traffic and the afternoon-return-city traffic. To give a typical example:
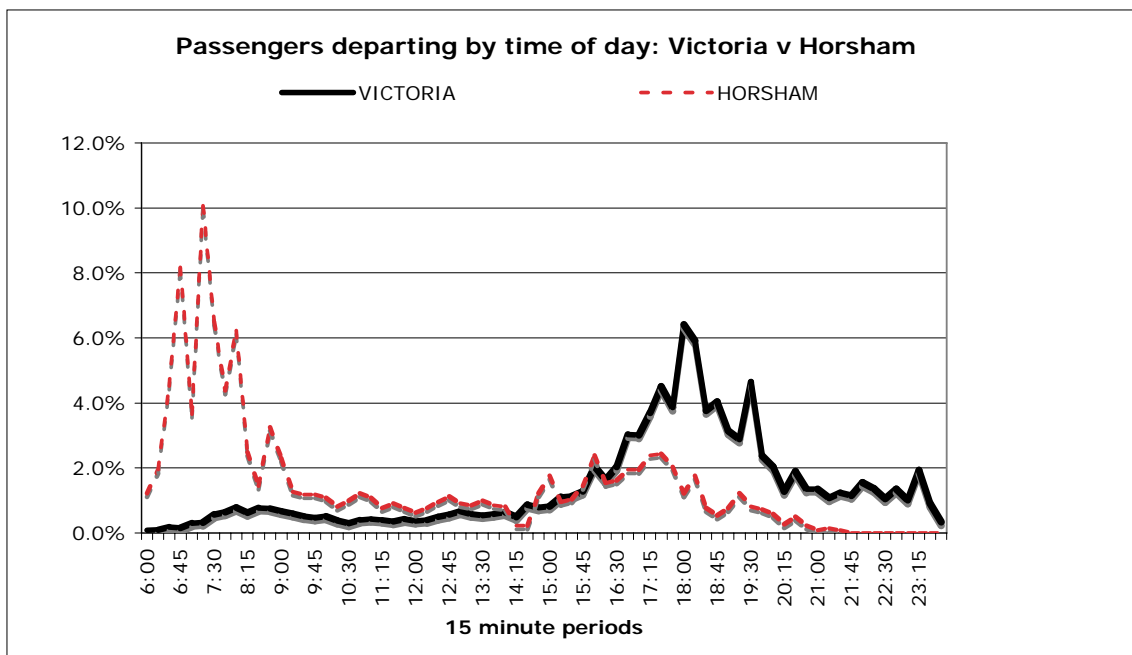
Source: Weekday passenger counts, LATS, 2001.

LATS (London Area Transport Survey) can contribute valuable input to this process, at least for London & SE area stations (and a few outside that area). Equivalent data on a national basis is expected to be available later this year from the NRTS (National Rail Travel Survey). We understand that both LATS and NRTS cover Monday to Friday traffic only, but this accounts for the bulk of the traffic. Rail Planner data may also be of use in shift planning (and covers all days of the week).

We believe that the same time of day target assumptions, and the same process logic, can be used for shift selection for each weekday (as Monday to Friday all appear to have very similar timetables and broadly similar traffic patterns). The process should aim to include approximately equal numbers of interviews for each of the five days. However, in view of the timetable differences between Saturdays and Sundays evident from Rail Planner data and differing passenger volumes by day and by hour each day as suggested by NPS, we believe that Saturday and Sunday should <u>each</u> have separate time of day targets for (a) city centre and (b) other stations within each TOC (rather than being treated as a single "Saturday+Sunday" target as at present).

We believe these changes will greatly improve the representativeness of the NPS sample.

3.5   Definition of journey

Another (smaller) deficiency noted in the current design is that there may be a mismatch between the definition of journey used to define the "journey universe" (see Section 3.3) and the definition applied by the fieldwork methodology. The concerns are:

(a)   In the majority of cases, questionnaires are distributed only to passengers beginning their journey at a specific station (as in many cases access cannot be gained to platforms where passengers interchange). The implication of this is that customers joining the rail network for the first (or only) leg of their journey are over-sampled (in relation to interchangers) at stations with high proportions of interchangers.

(b)   When (hard to find) interchangers do happen to receive questionnaires, they are treated in the analysis as if they have just begun their train journey at that station.

(c)   If results at stations with a large number of interchange passengers are sampled based on total passenger numbers (i.e. including many interchangers) any weighting may then further over-represent non-interchange passengers.

(d)   Extra analysis we carried out indicates that 15% of (weighted) journey legs recorded on NPS Wave 13 involved the respondent taking one train (or more) on later leg(s) of the journey[15]. However, the NPS questionnaire does not record whether the respondent has made any train journey(s) immediately prior to arrival at a station where questionnaires are distributed. If the proportion is as high as the 15% for subsequent interchanges, then of the order on 30% of passengers may be on journeys involving interchanges.

Given that interchangers are therefore likely to represent a significant proportion of passengers (and may have a journey experience which differs from the norm), consideration should be given to these issues when reviewing future questionnaire design (e.g. include a "prior interchange" question as well as the existing "subsequent interchange" question), and when reviewing shift allocations and weighting.

---

[15] NPS Wave 13 shows 75% of them as starting the current leg between 7am and 1pm, but this figure has probably been somewhat distorted by the outbound:return anomaly.

3.6 Sampling frame

In the light of the problems identified with the "time-of-day" universe data used in the sampling methodology, the process for allocating fieldwork shifts to timeslots (at selected stations) should be reviewed (see Sections 3.4 and 3.7).

However, the process employed by Continental for allocating the number of shifts to specific stations (but not shifts to time slots) appears to be very robust. In particular:

(a)   Applying electronic timetable data[16] to station journeys to estimate the proportion of journeys attributable to specific TOCs produces aggregate estimates which appear to reconcile well with TOC's own estimates of their total journeys (the adjustments needed to total TOC journeys derived at station level are very small).

(b)   Drawing shifts in proportion to these reconciled estimates ensures that a single sampling frame can be used to meet TOC quota targets and produce a representative shift pattern for each TOC by station (in aggregate).

The only NPS Report we have seen which contains detailed technical information on the sampling process is the "NPS Overview Report". However, this is not produced after every Wave, and has been given very limited circulation (we understand that, in the past, copies were sent only to the SRA). As a result it is unlikely to be readily evident to many NPS users that, for example, Wave 12 (for which no Overview Report was produced) used effectively the same Shift Sample as did Wave 11 (i.e. a largely identical set of station/day/time combinations), rather than merely the same Shift Sampling process (while Wave 13 used a fresh sample). This is not an inappropriate approach, but such facts should be made more readily accessible.

---

[16] Rail Planner is an electronic timetable of UK rail services published by Travel InfoSystems.

3.7 Fieldwork methodology

Continental's fieldwork shifts are typically 3 hours long (97% of Wave 13 shifts). Continental's sampling methodology assumes that on <u>average</u> around 16 completed questionnaires are returned per shift. The number of questionnaires actually handed out varies between about 20 and 70 per shift depending on passenger volumes (which vary by station, day and time). We understand that interviewers are given a supply of questionnaires for each shift based on the number expected to be required for that shift.

Within each shift, the design aims to achieve a representative spread of passengers about to board trains at that station over the duration of the shift.[17] There is no specific sampling methodology at each station.[18] Interviewers are asked to distribute questionnaires evenly across the three hours of the shift, and to move around between platforms.

We have no reason to believe that the existing procedure is not adequate. However, if there are any limitations in the execution of this procedure, they may be exacerbated by the fact that 96% of shifts arranged in the four discrete waves, each of 3 hours, starting respectively at 7am, 10am, 1pm and 4pm (the 7pm wave can be disregarded for this purpose as it accounts for only 1% of shifts and 1% of NPS journeys). We believe it would be safer to stagger the shifts (e.g. so that some start at 7am, some at 8am, some at 9am and so on through the day).

3-Hour Shift Blocks by Time of Day (NPS Wave 13)
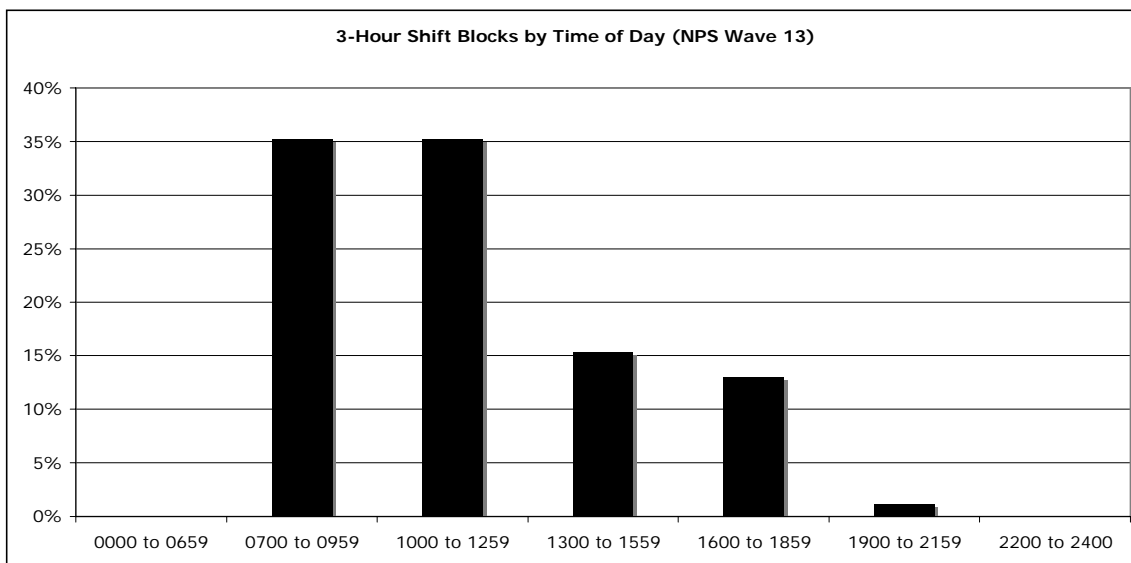
Evidence is that at most stations there are major swings in passenger volume from hour to hour (and quarter hour to quarter hour). Staggering shifts would reduce the risk of any bias due to uneven distribution within shifts, compared to the present design.

---

[17] As mentioned before, many interchangers may be omitted from this process as they may congregate at parts of stations which are less accessible to interviewers.

[18] For example, no specific quotas are set on Age, Gender, Journey Purpose or any other such variables – but as there are no sources which could ever conceivably provide reliable target numbers <u>per station</u> per hour per day, this is not a plausible option anyway.

However, even if interviewers are distributing questionnaires completely evenly throughout the shift, they are still in fact liable to under-represent peak journey flows. This is illustrated conceptually below.

Changes in journey volumes during a three hour shift
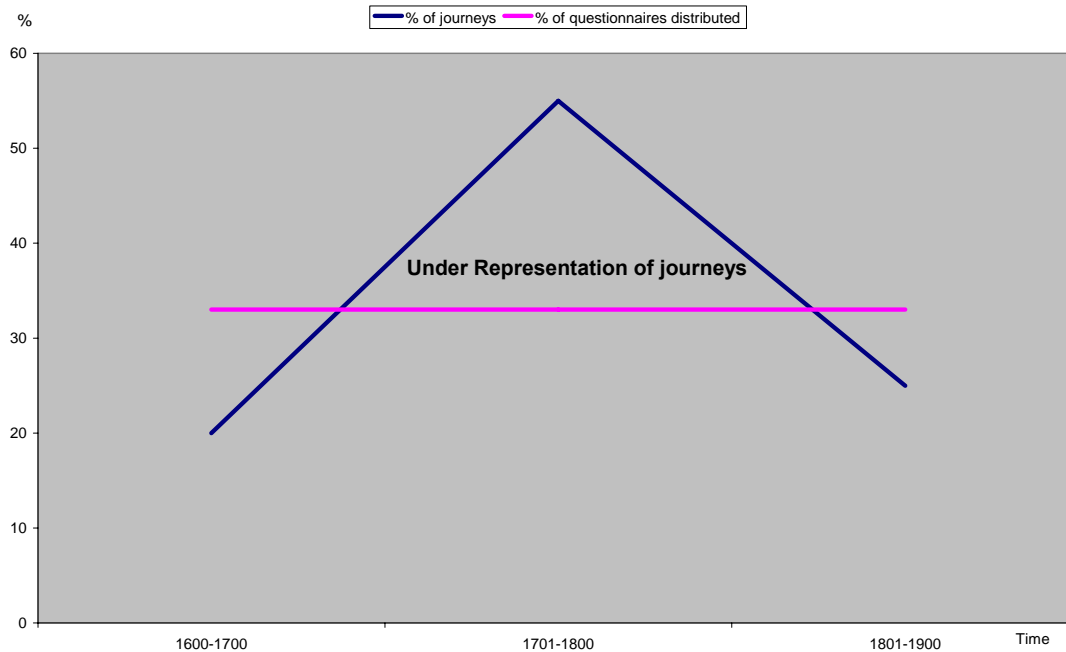
As we will never know in advance how many passengers will arrive at each station in each future hour it is always going to be impossible to arrange for interviewers to match the rate at which they distribute questionnaires from hour to hour in proportion to the numbers of passengers about to depart.

One way we <u>can</u> reduce the impact of this issue is to draw the shift sample at the level of two-hour slots, or 90-minute slots or even one-hour slots (rather than the current 3-hour slots). Sampling at this finer level of granularity would mean that the sample is more representative of the peaks and troughs in journey volumes, particularly during peak commuter periods.

Also, reducing shift duration (for any given sample size) would correspondingly increase the number of shifts (i.e. sampling points): e.g. halving the shift length to 90 minutes would allow a doubling of the number shifts. This would improve the quality of the NPS sample very significantly. In effect this would (i) greatly improved coverage of stations (i.e. far more stations would be sampled - especially smaller ones) and at the same time (ii) give greatly improved time of day balance at stations of all sizes (i.e. more shifts, on more days, at more times of day). We recommend very strongly that this is done.[19]

---

[19] Without further work we cannot say what the exact cost implications would be, but we seriously doubt that they would be adverse enough to offset the very large benefits re: the quality of the sample of smaller stations, and the representativeness of the time of day sample generally. Indeed, it should be possible to raise the cost-efficiency of the design: e.g. the electronic timetable could help in drawing a more cost-effective sample (in the past some 3-hour shifts at smaller stations have produced <u>nil</u> questionnaires). Also, some of the proposed short shifts could possibly be grouped in the interests of fieldwork efficiency: e.g. 3 consecutive 1-hour shifts at 3 <u>adjacent</u> small stations would still be <u>very</u> much better than 1x 3-hour shift at 1 small station (and not much more costly); similarly at <u>larger</u> stations 3 consecutive 1-hour shifts (each with its own shift target) would be better than 1x 3-hour shift (and would have an almost identical cost).

3.8 Differential response rates

Response rates[20] vary considerably by passenger Age and (to a much smaller extent) by Journey Purpose (Commuter, Business, Leisure). Differences by Gender are negligible:

Response Rates by Age, Journey Purpose and Gender[21]

Age	Response %	Purpose	Response %	Gender	Response %
15-34	30	Commuter	39	Male	36
35-54	42	Business	41	Female	39
55+	50	Leisure	35		

Source: Response analysis from a sample of NPS Wave 8 Fieldwork Reports, CR, 2003.

The difference in response rate by age - specifically the lower response rate in the under-35 age group - is a concern. However, setting quotas for questionnaire distribution by age group is impractical (distribution by age varies by station and time of day making it impossible, given the lack of reliable universe data on this demographic variable, to set realistic quotas).

However, correcting the age profile may affect the journey purpose profile, because we know that the two are interrelated:

	16-25 %	26-34 %	35-44 %	45-54 %	55-59 %	60-64 %	65+ %
Commuter	54	58	53	45	35	20	6
Business	6	14	19	23	22	15	7
Leisure	40	28	28	32	42	65	88

Source: NPS Wave 13 (weighted), CR 2006.

Following on from the proposed changes to the shift sampling methodology, further work will be needed to understand how best to respond to this issue. It would be helpful, prior to making changes, to obtain an analysis of response rate by age interlocking with journey purpose. This would help ensure that any age-weighting will achieve the correct balance by age and journey purpose.

---

[20] Number of completed questionnaires used in analysis divided by number distributed.
[21] Source: Continental Research email to Passenger Focus.

3.9   Weighting rationale

The purpose of weighting the data is to correct for any known distortions in the distribution across TOCs and across other key variables within TOC. These distortions are caused by a combination of:

- The setting of target sample sizes for each TOC
- The replacement of shifts for a specific TOC at a suburban station with a shift at a main line station. We know that this occurs in the sampling process and accept that this is a legitimate strategy provided changes are documented, distortions are kept to a minimum and appropriate weighting is applied at the analysis stage.[22]
- Differences in response rates among certain groups (e.g. age and journey purpose)

Within the NPS the weighting also has a second purpose: to gross data up to the annual journey estimates within TOC and across the rail network as a whole.[23]

Whilst weighting can have a beneficial effect on the analysis in that it forces the profile of the data to be "more like the real world", care must also be taken to ensure that weighting is not excessive. Weighting adversely effects the margin of error of estimates (percentages and mean scores) obtained from the survey through increasing the design effect of the survey. Therefore lighter, less complex weighting enables the research users to have more confidence in estimates produced by the survey.

The goal should be to simplify the weighting process as far as possible and harmonise the weighting scheme applied within the TOCs. The information used for weighting also needs to be more representative of the market than estimates produced by the unweighted survey data in order to justify the introduction of a greater design effect (with the loss of statistical robustness this entails).

---

[22] These changes are documented in a "Field Report" on each NPS Wave.
[23] This does not affect the design factor for individual TOCs.

3.10 Current weighting scheme

The current NPS method is to weight separately within TOC, using rim weighting. Each TOC is rim weighted separately by

- Journey Purpose (Commuter, Business, Leisure)
- Day of Week (Weekday, Weekend)
- Station Strata (Small, Medium, Large, Very Large)

GNER is additionally rim weighted by Class of Service (First Class, Other).

The distribution across all TOCs is then simultaneously weighted and grossed-up according the total number of journeys per annum on each TOC.

The following sections discuss the design effects of the current weighting.

3.11 Design effects of weighting

Rim weighting within TOC runs a computer based algorithm which produces a unique weighting factor for each cell defined by the levels of the weighting variables. The complete set of these cells forms a weighting matrix which can be reconstructed to investigate the design effects of weighting.

Details of the weighting process are provided for Wave 11 in Continental's methodological note. We have carried out a full analysis for this Wave, in which the TOCs were broken up into 28 units (some TOCs were separated further due to differences in profile). The number of weighting cells can be calculated as follows:

TOCs (other than GNER):
27 (TOCs) x 3 (Journey Purpose) x 2 (Day of week levels) x 4 (Station Strata) = 648

GNER:
1 (TOC) x 3 (Journey Purpose) x 2 (Day of week levels) x 4 (Station Strata) = 48

Total Matrix[24] = 696 cells

The number of TOC categories varies from Wave to Wave, resulting in different numbers of cells. However, the weighting design is always the same in principle.

Most of the analysis generated from the NPS is at the level of individual TOCs therefore the main concern is the margin of error for estimates within TOC. However, some analysis (Category Analysis and Station level analysis) is undertaken across TOCs. We have investigated the effects of weighting at all of these levels of granularity.

The design effect at each level of granularity can be assessed by the effect on weighting on "effective sample size". The level of statistical confidence increases with larger sample sizes. As a rule of thumb, given any arbitrary sample size you have to quadruple the sample size in order to halve the margin of error. A full table of confidence intervals (the margin or error around estimates derived from the survey) is provided in Section 3.21, along with a more complete explanation.

Weighting actually increases the margin error for specific subgroups. We can quantify the increase in the margin of error due to weighting by estimating the drop in sample size which would produce an equivalent widening of the margin of error in the unweighted sample. Using "effective sample size" instead of actual sample size to look up the confidence interval takes account of the effect of weighting at a particular level of granularity. Under any weighting scheme where one or more weights are not equal to one, the effective sample size due to weighting is always less than the unweighted sample size. Generally speaking, the further the weighted sample profile from the unweighted profile, the bigger the design effect, and hence the bigger the drop in effective sample size and the greater the widening of the margin of error around survey estimates.

Strictly speaking, this analysis assumes that our sample is a simple random sample of journeys and that there are no anomalies in the sample frame design. We know that this is not a simple random

---

[24] GNER class weighting has been ignored for this purpose and is looked at separately in Section 3.18. Average cell size on Wave 11 was 34 questionnaires overall, or 37 if GNER class weighting is ignored.

sample and that there are anomalies in the design by time of day. Therefore the margins of error presented can probably be taken as minimum errors (other distortions in the sampling are likely to widen these somewhat further).

Note that the confidence intervals stated in the following sections are "worst case" confidence intervals assuming we are interested in assessing the level of confidence around an estimate of about 50% for some quantity from the survey (e.g. satisfaction with some measure). The closer an estimate is to 50%, the wider the confidence interval. For a fuller appreciation of the margin of error around estimates nearer 0% or 100%, see Section 3.21.

3.12 Design effects overall

For the entire sample, the weighting scheme employed in Wave 11 results in a reduction in effective sample size from 25,596 to 14,439, a reduction of 44% in the effective sample size. This is mainly due to the large weighting factors that have to be applied to the individual TOC samples in order to represent them in the correct proportion over the entire rail network.

We anticipated a large design effect when aggregating across TOCs, given that quotas for the TOCs are set artificially to achieve a target base size, rather than to produce a more representative distribution in the sample. Because of the large sample size at aggregate level, this only represents a small increase in the width of the 95% confidence interval from ± 0.6 percentage points (a width of 1.2 percentage points) to ± 0.8 percentage points (a width of 1.6 percentage points).

In other words the whole confidence interval widens by less than half a percentage point (0.4), which is negligible and is therefore not a matter for concern.

This design effect could be narrowed by adjusting the TOC's target base sizes to make them more in line with the actual number of journeys on each TOC, but we appreciate this may be a sensitive issue for some TOCs and feel on balance this is not necessary, given the very large sample size for the entire network overall.

3.13 Design effects within TOC

The design effect within each TOC (or "virtual" TOC)[25] together with the worst-case confidence interval (for an estimate of 50%) is summarised below.

Design Effect by TOC (ordered by loss of effective sample size)

TOC	Unweighted Base	Effective Sample size due to weighting	Effective Sample reduction due to weighting	95% Confidence Interval width (unweighted)	95% Confidence Interval width (weighted)	95% Confidence Interval increase due to weighting
				(percentage points)	(percentage points)	(percentage points)
Island Line	255	113	-56%	12.2	18.4	6.2
Silverlink metro	461	230	-50%	9.2	13.0	3.8
One - Anglia	443	289	-35%	9.4	11.6	2.2
Arriva Trains Wales	776	519	-33%	7.0	8.6	1.6
Wessex Trains	512	370	-28%	8.6	10.2	1.6
Gatwick Express	503	374	-26%	8.8	10.2	1.4
ScotRail	1034	775	-25%	6.0	7.0	1.0
Virgin West Coast	1015	761	-25%	6.2	7.2	1.0
WAGN (GN only)	695	530	-24%	7.4	8.6	1.2
One - Great Eastern	734	568	-23%	7.2	8.2	1.0
Midland Mainline	1075	833	-23%	6.0	6.8	0.8
One – West Anglia	436	343	-21%	9.4	10.6	1.2
Merseyrail	540	426	-21%	8.4	9.4	1.0
Central Trains	1524	1208	-21%	5.0	5.6	0.6
Arriva Trains Northern	787	642	-18%	7.0	7.8	0.8
Virgin CrossCountry	1140	953	-16%	5.8	6.4	0.6
Southern	1655	1402	-15%	4.8	5.2	0.4
Trans Pennine Express	1038	889	-14%	6.0	6.6	0.6
GNER	1078	930	-14%	6.0	6.4	0.4
First Great Western	1185	1039	-12%	5.6	6.0	0.4
Silverlink county	645	569	-12%	7.8	8.2	0.4
First North Western	514	455	-12%	8.6	9.2	0.6
c2c	1039	924	-11%	6.0	6.4	0.4
Thameslink	1260	1133	-10%	5.6	5.8	0.2
Chiltern Railways	1108	1003	-9%	5.8	6.2	0.4
FGWL	1042	948	-9%	6.0	6.4	0.4
South Eastern	1542	1447	-6%	5.0	5.2	0.2
South West Trains	1560	1509	-3%	5.0	5.0	0.0

Source: Analysis of NPS Wave 11

For most TOCs the effect of weighting is modest. Only 8 of the 28 (TOCs or virtual TOCs) above have their confidence interval increased by more than one percentage point after weighting. The (real) TOCs which are of greatest concern are Island Line, Wessex Trains and Gatwick Express, all of which have a confidence interval with a width greater than 10 percentage points after weighting. These TOCs tend to have a combination of lower sample size and larger design effect due to weighting - and it is this combination which leads to the widest (poorest) confidence intervals.[26]

The adjusted weights[27] applied to some of these TOCs (or virtual TOCS) are shown in the following Charts, together with a summary of the main variables contributing to the design effect.

---

[25] For sampling purposes, the NPS design breaks some TOCs into sub-sections (referred to in the Overview Reports as "virtual TOCs"): e.g. in Wave 11 Silverlink was split into "Silverlink Metro" and "Silverlink County" and "one" was split into "one Great Eastern" and "one West Anglia". This very useful concept was introduced to allow the NPS sample design to allow for like-for-like comparisons with historical data even after some sections of the network have been transferred from one (real) TOC to another between Waves.

[26] Even if their sample sizes cannot be increased, proposed sample design changes (e.g. reduction in shift length - see Section 3.7) would be expected to improve this situation significantly.

[27] For this comparison weights were rescaled to produce the same weighted as un-weighted base size.

Magnitude of the standardised weights within TOCs (1)
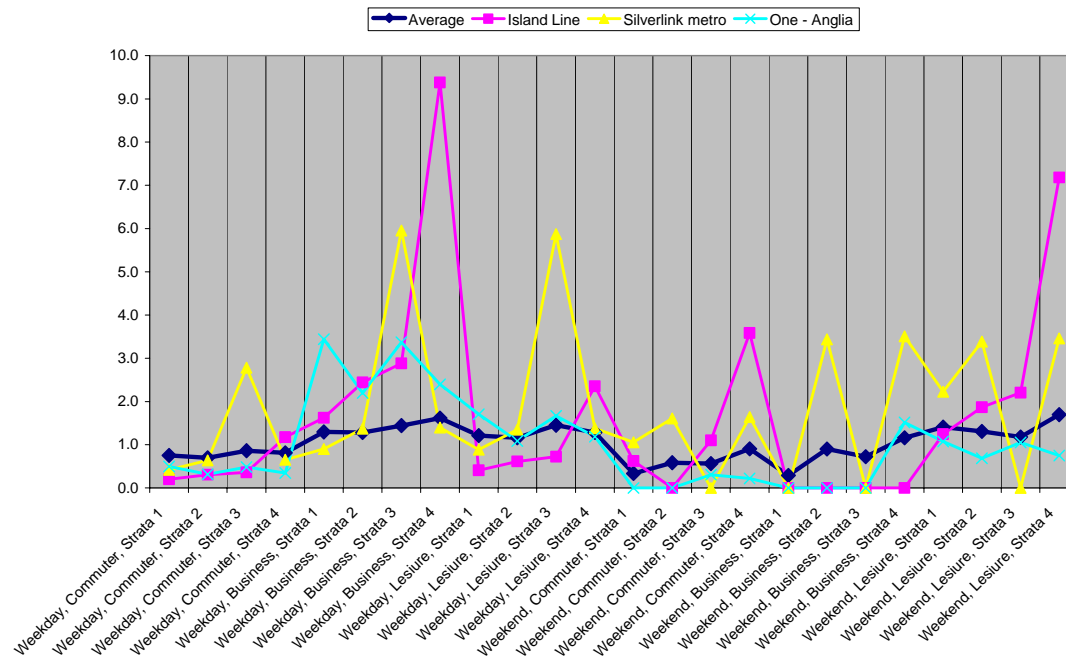
Source: Analysis of NPS Wave 11
NB: Zero value = no data (no questionnaires apply).

Design Effects caused by:

Island Line
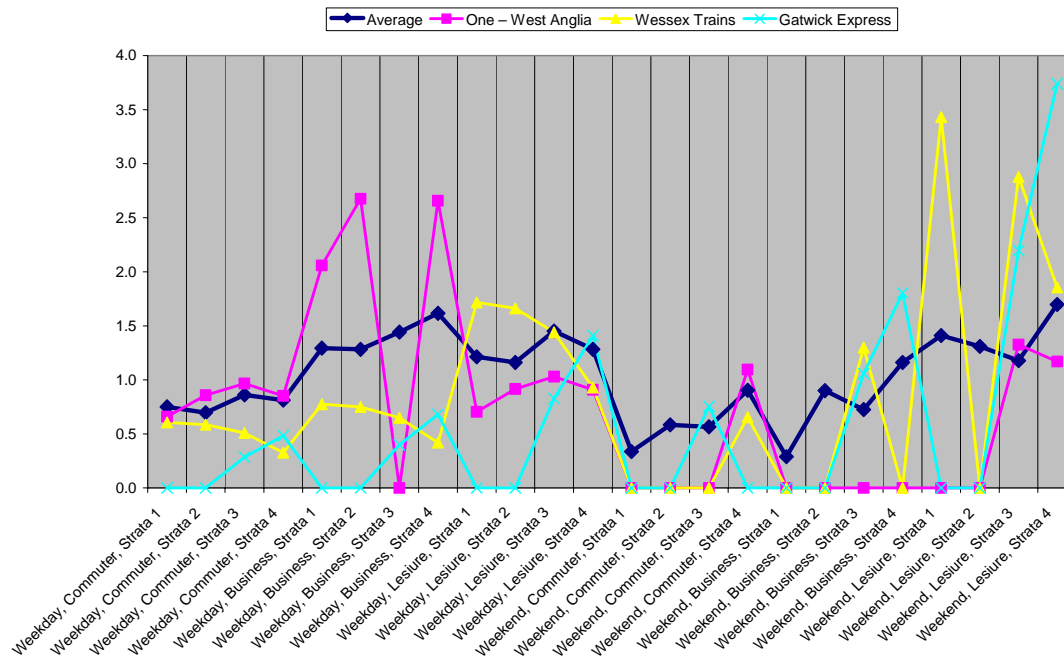Upward weighting of Very Large stations (Stratum 4) and Business Trips, particular during weekdays.

Silverlink Metro
Upward weighting of Large stations (Stratum 3) during Weekdays and Business and Leisure journeys during the weekend

One Anglia
Upward weighting of Business trips during Weekdays

Magnitude of the standardised weights within TOCs (2)

Source: Analysis of NPS Wave 11
NB: Zero value = no data (no questionnaires apply).

Design Effects caused by:

One – West Anglia
Upward weighting of Business journeys during weekdays

Wessex Trains
Upward weighting of Leisure journeys particularly during the weekend

Gatwick Express
Upward weighting of very large stations (Stratum 4) at weekends

Although we have not investigated other TOCs due to time restrictions, this analysis demonstrates that there are some significant differences between the distribution of journeys in the sample and the journey universe, described by the weights applied at analysis. The weighting by station strata and journey purpose accounts for the largest discrepancies in these examples.

3.14 Design effects within category

For completeness, the design effect was investigated for the TOC categories in aggregate. Category averages are used as yardsticks for individual TOCs, therefore it is useful to understand how this level of reporting is affected by weighting.

Design effect at Category Level

Category	Unweighted Base	Effective Sample size due to weighting	Effective Sample reduction due to weighting	95% Confidence Interval width (unweighted)	95% Confidence Interval width (weighted)	95% Confidence Interval increase due to weighting
				(percentage points)	(percentage points)	(percentage points)
London & SE	12620	8775	-30%	1.8	2	0.2
Regional	7483	4239	-43%	2.2	3	0.8
Long Distance	5493	4358	-21%	2.6	3	0.4

Source: Analysis of NPS Wave 11

In fact, the impact on margin of error is minimal, due to the larger base sizes at category level. Regional Operators are affected most by weighting, although the widening of the confidence interval for this group is still less than one percentage point.

3.15 Design effects within station

As we understand that some station-level tables are also produced from the survey, we also investigated the design effect at one of the biggest mainline stations – London Victoria.

Design Effect at London Victoria

Region	Unweighted Base	Effective Sample size due to weighting	Effective Sample reduction due to weighting	95% Confidence Interval width (unweighted)	95% Confidence Interval width (weighted)	95% Confidence Interval increase due to weighting
				(percentage points)	(percentage points)	(percentage points)
London Victoria	666	287	-57%	7.6	11.6	4

Source: Analysis of NPS Wave 11

At this station the design effect is large in magnitude (a loss of greater than half in effective sample size). The combination of this, together with the smaller base size at station level has led to a widening of the confidence interval by four percentage points, taking its width to nearly 12 percentage points in total.

This level of robustness is a lot poorer than for other cuts of the data investigated (e.g. by TOC). The significant design effect is due to the large differences in weighting applied to the three TOCs operating services from Victoria – South Eastern, Southern and Gatwick Express, as well as an up-weighting applied to Business Journeys at the station. The large impact on the confidence interval is not helped by the smaller sample size at station level (although the sample size for Victoria is still greater than for many smaller TOC totals in the survey).

We have not investigated other mainline stations due to time constraints, but suspect that the pattern would be similar. Data reported at Station level is therefore likely to have a lower level of statistical robustness than for other cuts of the data.

3.16 Design effects across all TOCs

The design effect for the subgroups defined by the weighting variables was estimated over the entire sample, for all TOCs[28]. The purpose was to shed further light on which variables are most adversely affected by the current weighting scheme. The results are highlighted in the table, sorted by order of design effect. For simplicity, in this instance, we only show loss in effective sample size due to weighting.

Design effect for weighting subgroups across all TOCs

Segment	Effective Sample reduction due to weighting
Weekday, Business, Strata 4	-62%
Weekday, Business, Strata 1	-60%
Weekend, Business, Strata 4	-59%
Weekend, Business, Strata 1	-59%
Weekend, Business, Strata 3	-54%
Weekend, Business, Strata 2	-53%
Weekday, Business, Strata 3	-52%
Weekend, Commuter, Strata 3	-52%
Weekday, Commuter, Strata 4	-51%
Weekend, Lesiure, Strata 1	-49%
Weekday, Business, Strata 2	-49%
Weekend, Commuter, Strata 4	-48%
Weekday, Lesiure, Strata 4	-43%
Weekday, Commuter, Strata 1	-37%
Weekday, Commuter, Strata 3	-37%
Weekday, Lesiure, Strata 3	-37%
Weekend, Commuter, Strata 1	-37%
Weekend, Lesiure, Strata 3	-37%
Weekday, Lesiure, Strata 1	-36%
Weekend, Lesiure, Strata 4	-35%
Weekend, Lesiure, Strata 2	-34%
Weekend, Commuter, Strata 2	-33%
Weekday, Commuter, Strata 2	-30%
Weekday, Lesiure, Strata 2	-26%

Source: Analysis of NPS Wave 11

It is clear from this table that the design effect is greatest for the Business travellers subgroup, followed by Commuters. The pattern by day of week (weekday vs. weekend) and by Station Strata is much less clear. This analysis suggests that there is a much greater mismatch between the journey purpose data captured in the survey and the journey purpose split (by TOC) used for weighting, compared with Station Strata or Weekday.

---

[28] Excluding GNER which is weighted differently from the other TOCs; this is looked at separately.

3.17 GNER weighting by class

The design effect of weighting additionally by Class of travel for GNER only was investigated. The results are shown in the table, sorted by order of design effect.

Design Effect of Weighting by Class (GNER)

Segment	Effective Sample reduction due to weighting
Weekday, Business, Strata 4	-10%
Weekday, Business, Strata 2	-10%
Weekday, Business, Strata 1	-9%
Weekend, Business, Strata 4	-9%
Weekday, Commuter, Strata 1	-8%
Weekend, Commuter, Strata 4	-7%
Weekday, Commuter, Strata 4	-5%
Weekend, Business, Strata 2	-5%
Weekday, Lesiure, Strata 4	-5%
Weekend, Lesiure, Strata 4	-4%
Weekday, Lesiure, Strata 2	-3%
Weekend, Lesiure, Strata 1	-3%
Weekday, Lesiure, Strata 1	-3%
Weekend, Lesiure, Strata 2	-3%
Weekday, Commuter, Strata 2	-3%
Weekend, Commuter, Strata 1	0%
Weekend, Commuter, Strata 2	0%
Weekend, Business, Strata 1	0%

Source: Analysis of NPS Wave 11

The design effect due to weighting by class for GNER is minimal (resulting in a less than 10% reduction in effective sample size for each subgroup of GNER). The merits of this weighting variable are discussed later, but we can conclude that it does not have an adverse effect on the margins of error of estimates obtained for GNER.[29]

---

[29] Although if the ticket class ratios used as weighting targets are incorrect, that would be expected to alter GNER's average satisfaction and rating scores.

3.18 Reducing design effects

Design Effects of weighting can be reduced and hence Confidence Intervals narrowed (improved) by:

(a)   Increasing the sample sizes for the affected TOCs.

This can be expensive and does not address any underlying distortions created by the sampling and weighting process. However, a larger initial sample size will lead to a lower (=better) overall confidence interval and a lower overall design effect due to weighting, provided there is faith in the overall sampling and weighting process.

(b)   Minimising and/or trimming the weighting applied at the analysis stage.

This is a dual process of ensuring that the sampling process produces as correct a distribution of journeys as possible within TOC and ensuring that there is real confidence that the weighting applied produces a "better" distribution of journeys than achieved in the unweighted sample. In summary, the goal should be to trade off improvements in the distribution of journeys in the sample against complexity of the weighting process. There are no hard and fast rules, but the design effect tends to increase rapidly as additional variables are included in the weighting process.

We would strongly recommend considering (b) before (a) as simplifying the weighting process does not have an adverse impact on budget.

3.19 Review of Existing Weighting Variables

We now consider each of the weighting variables in turn (other than weighting by TOC - see section 3.13).

Journey Purpose

This is the most contentious of the weighting variables. Journey Purpose weightings are based on estimates. For most TOCs, they are currently derived by making generalisations on data held on ticket type sold by each TOC, i.e. it is assumed for the purposes of the NPS weighting process that:

Season Tickets = Commuters
Full price singles/returns = Business
Discounted/Cheaper tickets = Leisure

This generalisation clearly does not apply in all cases, as can be seen by the following analysis of unweighted data from NPS Wave 12.

Ticket Type by Type of Journey (Wave 12: unweighted)

Journey type:	Commuter	Business	Leisure
Ticket Type	%	%	%
Season Ticket	51	2	3
Full Price Ticket	26	56	30
Discounted Tickets	22	42	66
	100%	100%	100%

According to the NPS, only half of commuter journeys are on a Season ticket and only just over a half of Business journeys are on a full price ticket. It is very clear from this analysis that weighting journey purpose using universe data on ticket type will produce very significant skews in the analysis. Ticket Type is a very bad approximation to journey purpose. This also explains why we found a large design effect for this group - the big discrepancy between the two sources of data on journey purpose requires heavy weighting at analysis, resulting in larger design effects. The mismatch almost certainly also produces a less realistic profile than obtained from the unweighted sample.

We recommend that this weighting should be removed (once the shift allocation process is revised satisfactorily). A possible rationale for including this weight would be if there exists a differential response rate by Journey Type. However, our analysis shows that only small differences exist in response rate. It is also our opinion that these can be more effectively addressed via other strategies e.g. (i) a review of the sampling strategy for allocating shifts to particular times of day and/or (ii) weighting using information collected on differential response rate among these groups. Strategy (i) is the one we would recommend pursuing. Strategy (ii) only has additional merits if there is clear evidence that differential response rates among these groups are having an adverse impact on sample distribution.

An alternative suggestion, of weighting the Ticket Type proportions in the survey using the Ticket Type universe data, was proposed in 2005. While this has the merit of matching like data with like, we do not believe that there is yet a sufficient case for recommending applying this additional level of weighting, for the same reasons discussed above – namely that any bias is better addressed via changes to the sampling strategy. An improved sample design re: shift allocation by time of day should produce a more representative result by journey purpose.

Day of Week
Theoretically, the sampling frame for the survey should ensure that approximately the correct split should be applied to weekday v. weekend journeys (for city centre and other stations). However, as documented previously, there are deficiencies in this part of the design. Assuming that these deficiencies will be corrected, there is still a case for weighting by Weekday vs. Weekend within TOC if Passenger Focus believes that:

- The weekday to weekend journey split differs significantly for each TOC. This appears to be the case given the splits assumed for the weighting of Wave 11.
- The information on the split used for weighting is better than the more general split applied during sampling and gives a more representative distribution within TOC. This depends on the sampling process adopted in future.

If there is not a strong case for either of the above then there is not a strong case for weighting by this variable. However, if there is still a case for believing these points, then this weight should remain for the sake of consistency.

Station Strata
There is a very strong case for continuing to weight by Station Strata within TOC. There are occasions when some TOCs may have a shortfall in their quota as a result of:

- Imperfect allocation of shifts at particular stations (the number of shifts only approximates the number of interviews that will be obtained at a specific station)
- Shortfall in the balance of TOCs at specific stations
- Booster interviews for specific TOCs, which occur at the end of fieldwork, may introduce mix changes which require correction

The four strata used currently should produce a reasonable correction for these anomalies. This is the variable for which there is the strongest justification for weighting. The effect of weighting by Station Strata can be minimised by ensuring that adjustments to the fieldwork schedule do not overly distort the distribution by TOC at different sized stations (strata 1 to 4) in the unweighted sample. Minimising distortions will help reduce the larger design effect due to weighting at mainline stations, such as London Victoria.

Class of Travel
Weighting by class is bound to affect GNER's scores, so there would have to be an extremely strong argument to justify it. Logically, the weighting by class of travel for GNER, if valuable, should be applied to all TOCs; if not, to none. Without knowing why this exception was introduced for GNER it is difficult to comment further, except to say that perhaps the sampling improvements we have

suggested will make it unnecessary.

3.20 Review of possible additional weighting variables

A review of weighting would not be complete without considering whether there is justification for introducing additional weighting variables. Several ideas have been floated such as whether to weight additionally by:

- Gender/Age/Journey Purpose
- Station
- Routes
- Category (Long Distance/Regional/London SE)

Taking these in reverse order:

Category

There can be no merit in weighting by category, if as we understand the categories consist of mutually exclusive subsets of TOCs. Given that the data is already weighted at the level of individual TOCs (a much finer level of granularity) there is no justification or need for weighting by region.

Routes

Although weighting by route gives a much finer level of precision than weighting by TOC, the survey is currently not designed for analysis at this level of granularity. Therefore in our opinion this would be a step too far. It would also create a much more complex weighting matrix with an order of magnitude difference in the number of weighting cells, which is likely to cause a steep increase in the design effects of weighting. Passenger Focus would additionally need to consider whether the information available for weighting at route level was reliable enough to justify the loss in effective sample size resulting from more complex weighting. We believe that this is unlikely to be the case and therefore recommend rejecting this additional variable.

Station

Although the sampling frame is constructed at station level, its overall purpose is to derive samples which are representative by TOC. The primary unit at which the results are analysed is by TOC. For this reason we are not persuaded that further weighting by station (either across or within TOC) is likely to improve on the existing strategy of weighting by station strata.

Age and/or Gender and/or Journey Purpose

There is no case for weighting by gender – males and females have similar response rates.

Although journey purpose distribution (commuter, business or leisure) is already a contentious issue due to the sample not adequately representing time of day and its existing treatment in the weighting process (section 3.19) there is little evidence of the need to weight by this variable in future. The response rates are broadly similar.

There may be a case for weighting by age. Based on response analysis the younger age groups are less likely to return completed questionnaires.

However, weighting to correct for non-response bias by age may alter the sample profile by journey purpose (see Section 3.8). Changing the distribution of either variable through weighting will affect

the distribution of the other. We propose therefore that this be reviewed in the light of any changes to be made to the sampling methodology in future. Further work should also be conducted on non-response bias (e.g. examining interlocking response rates by age and journey purpose) before any demographic weighting is reconsidered.

3.21 Confidence Intervals

Any estimate of a population percentage made from a representative sample has a margin of error associated with it, known as sampling error. This results from the fact that not all units in the sample (in the case of the NPS, all journeys on the rail network) were included in the survey. A mathematical formula exists for calculating the error we would expect around these estimates, which for percentage estimates is dependent on the sample size and the size of the estimate itself. As a rule of thumb, quadrupling the sample size halves the sampling error; the sampling error is also greatest for estimates of 50% and least for estimate at the extreme (tending towards 0% or 100%).

We can express the sampling error for any sample estimate in the form of a "95% confidence interval". This means that 95 times out of 100 we would expect that the true population value lies within a specific percentage range around the estimate we have made from the sample, though with a higher probability of being near the sample estimate itself than at the extremities of the interval. Typical 95% percent confidence intervals for the NPS are shown in the table over page.

For example if in our sample for South West Trains (the TOC with largest NPS sample base size), we find that, for 80% of journeys, people are very satisfied with a particular aspect of the journey, then 95 times out of 100 we would expect between 78.0% and 82.0% of universe journeys to have this score (80% ± 2.0 percentage points). Another way of expressing this is to say that the confidence interval around the estimate has a width of 4.0 percentage points.

If the same situation applied to Island Line (the TOC with smallest NPS sample base size), and we find that, for 80% of journeys, people are very satisfied with a particular aspect of the journey, then 95 times out of 100 we would expect between 72.6% and 87.4% of universe journeys to have this score (80% ± 7.4 percentage points). Another way of expressing this is to say that the confidence interval around the estimate has a width of 14.8 percentage points.

Strictly speaking, the confidence intervals shown in the following table assume that Simple Random Sampling is used, which means that every journey on the network is equally likely to be selected for the survey. This is not the case for the NPS, which uses a mixture of stratified and cluster sampling. Other sources of error due to the design effects of the sampling process are difficult to estimate, but the confidence intervals shown below can be thought of as minimum levels.

The following table shows the 95% confidence limits (i.e. 2 standard errors) for the sample as a whole and for a range of TOCs, for survey results at a range of levels from 10% to 90%. The calculations were based on the dataset for NPS Wave 11.[30]

---

[30] Wave 13 database was not available when this calculation was done, and therefore Wave 11 (the autumn Wave from the previous year) was used for this purpose.

Confidence Intervals

Subgroup	Efective Sample Size	Size of Estimate				
		10% or 90% ± %	20% or 80% ± %	30% or 70% ± %	40% or 60% ± %	50% ± %
All Journeys	**14439**	0.5	0.7	0.7	0.8	0.8
TOC						
South West Trains	1509	1.5	2.0	2.3	2.5	2.5
South Eastern	1447	1.5	2.1	2.4	2.5	2.6
Southern	1402	1.6	2.1	2.4	2.6	2.6
Central Trains	1208	1.7	2.3	2.6	2.8	2.8
Thameslink	1133	1.7	2.3	2.7	2.9	2.9
First Great Western	1039	1.8	2.4	2.8	3.0	3.0
Chiltern Railways	1003	1.9	2.5	2.8	3.0	3.1
Virgin CrossCountry	953	1.9	2.5	2.9	3.1	3.2
FGWL	948	1.9	2.5	2.9	3.1	3.2
GNER	930	1.9	2.6	2.9	3.2	3.2
c2c	924	1.9	2.6	3.0	3.2	3.2
Trans Pennine Express	889	2.0	2.6	3.0	3.2	3.3
Midland Mainline	833	2.0	2.7	3.1	3.5	3.4
ScotRail	775	2.1	2.8	3.2	3.5	3.5
Virgin West Coast	761	2.1	2.8	3.3	3.5	3.6
Arriva Trains Northern	642	2.3	3.1	3.5	3.8	3.9
Silverlink county	569	2.5	3.3	3.8	4.0	4.1
One - Great Eastern	568	2.5	3.3	3.8	4.0	4.1
WAGN (GN only)	530	2.6	3.4	3.9	4.2	4.3
Arriva Trains Wales	519	2.6	3.4	3.9	4.2	4.3
First North Western	455	2.8	3.7	4.2	4.5	4.6
Merseyrail	426	2.9	3.8	4.4	4.7	4.7
Gatwick Express	374	3.0	4.1	4.7	5.0	5.1
Wessex Trains	370	3.1	4.1	4.7	5.0	5.1
One – West Anglia	343	3.2	4.2	4.9	5.2	5.3
One - Anglia	289	3.5	4.6	5.3	5.7	5.8
Silverlink metro	230	3.9	5.2	5.9	6.3	6.5
Island Line	113	5.6	7.4	8.5	9.1	9.2
Region						
London & SE	8775	0.6	0.8	1.0	1.0	1.0
Regional	4239	0.9	1.2	1.4	1.5	1.5
Long Distance	4358	0.9	1.2	1.4	1.5	1.5
London Victoria	287	3.5	4.6	5.3	5.7	5.8
General Samples						
	50	8.4	11.2	12.8	13.7	14.0
	100	5.9	7.9	9.0	9.7	9.8
	150	4.8	6.4	7.4	7.9	8.0
	200	4.2	5.6	6.4	6.8	6.9
	250	3.7	5.0	5.7	6.1	6.2
	300	3.4	4.5	5.2	5.6	5.7
	400	2.9	3.9	4.5	4.8	4.9
	500	2.6	3.5	4.0	4.3	4.4
	750	2.1	2.9	3.3	3.5	3.6
	1000	1.9	2.5	2.8	3.0	3.1
	1500	1.5	2.0	2.3	2.5	2.5
	2000	1.1	1.4	1.6	1.8	1.8
	5000	0.8	1.1	1.3	1.4	1.4

NOTE: Data based on NPS Wave 11.

4.   Conclusions and recommendations

Conclusions and recommendations expressed in this section are based on the evidence of the results of the foregoing Consultation Exercise and Technical Review plus our study of the documentation and analysis available to us, and include judgements and opinions evaluated in the light of our collective experience of many years of survey research.

4.1   Gold Standard Uses and Key Objectives

Technically speaking the NPS is a trip-based transactional survey - a survey of rail journeys. Currently the NPS is carried out twice a year (10 weeks in "spring" and 10 weeks in "autumn") and covers journeys commencing after 7am and before 10pm, based on interviews with passengers in GB aged 16 or older.

The "gold standard" use of the NPS is to measure, for each franchised TOC (and for the GB "heavy" rail network as a whole) the passengers' levels of satisfaction and their evaluations of each of a large number of components of performance.

The data produced is used (i) to measure changes in these ratings over time for each TOC, and for the network overall, and (ii) to compare the ratings of any one TOC with any other TOC, or group of TOCs, or the network overall.

Given that the questionnaire is already a fairly long one, adding further objectives which would require adding significantly to questionnaire length, would be likely to prejudice the achievement of the "gold standard" goals.

It is to be expected that the survey will produce reliable and representative results for major journey subgroups (peak v. off-peak; commuter v. business v. leisure; station size; station type - i.e. city centre v. other; weekday v. weekend; a.m. v. p.m.; long distance v. regional v. London/SE; etc.) and major passenger types (gender, age etc.) for each TOC individually, and at the national level.[31]

Given that the survey is an important yardstick of TOC performance, it should be technically sound and free from bias, and seen to be so.

These goals require that (i) the survey can be demonstrated to be representative, (ii) changes to the design which are important or essential (such as those we recommend are necessary for the sample to be seen to be sufficiently representative) should be implemented carefully in order to maintain comparability of results over time, and (iii) other changes from Wave to Wave should be relatively limited.

---

[31] It has not been regarded as an objective for the NPS to produce representative results for sub-TOC level areas such as an individual PTE area or a city (although the sample size for Greater London would certainly be large enough to allow this) or results for individual stations (only the very largest termini would be likely individually to have a sample size sufficiently large for independent analysis) and even for the largest of these examples it will not reliably do so unless the sample design is re-structured to meet that objective.

4.2 Self-completion method and response rates

The NPS fieldwork method is to distribute self-completion questionnaires to travellers at stations before they board trains. Respondents are asked to complete the questionnaire in relation to the train journey they are about to commence and are given a reply paid envelope in which to return the questionnaire.

The self-completion method is generally satisfactory for the purposes of the NPS.

The alternatives are less satisfactory. The **on-line method** is reasonably cheap but there is no way of which we are aware that a satisfactory sample of journeys could be achieved using the on-line method; **telephone interviews** would be more expensive, and would also face similar sampling problems; **face-to-face interviewing** (as opposed to face-to-face questionnaire distribution) would raise the cost of the NPS very substantially, and would be very difficult to carry out at peak times.

The main weakness of the self-completion method in general is that response rates are often very low (e.g. 10% or less) and this allows scope for substantial non-response bias. The NPS response rate over the last few waves has been between 36.2% and 41.2%% (in terms of questionnaires processed divided by those handed out[32]), which is remarkably high for such a long self-completion questionnaire for completion of which no incentive is offered.

This suggests that respondents have a relatively high level of interest (or self-interest) in the topic of the survey.[33]

NPS Response Rates[34]

NPS Wave:	10	11	12	13
Distributed:	70788	72476	80981	82985
Returned:	29170	26221	29423	30497
Response rate:	41.2%	36.2%	36.3%	36.8%

Source: Waves 10 to 13 (unweighted), CR, 2005

However, a response rate around 38% still allows scope for significant non-response bias.

Interviewers handing out questionnaires are asked to record the respondent's name and phone number (which are needed to allow quality control checking). On Waves 8 to 12 they were asked to record also the gender, observed age, and journey purpose (for Wave 13 journey purpose was replaced by ticket type). By matching responses with interviewer record forms it is possible to estimate response by type of respondent.

Such analysis has been carried out on a sample of forms on at least two occasions:

---

[32] This takes no account of any travellers who refused to accept one.
[33] Indeed, one criticism which has been made is that people who have strong feelings for or against the journey are more likely to respond than those who have less strong views. This may be true, but (a) if so the extremes may cancel each other out, and (b) the survey result is still likely to provide an effective measure of changes in the balance of views over time. Such criticisms can be muted if successful efforts are made to raise the response rate (and analyse its effects).
[34] The NPS Overview Report for Wave 11 gives the response rate as 34.5%, not 36.2% as shown in the above table. We were unable in the time available to investigate this discrepancy.

NPS Response Rates	Wave 8 Spring 2003	Wave 12 Spring 2005
Among	%	%
Male	36	37
Female	39	39
15-34*	30	28
35-54	42	43
55+	50	52
Commuter	39	n.a.
Business	41	n.a.
Leisure	35	n.a.
TOTAL	38	38

Source: Analysis of samples of fieldwork reports, CR, 2005
* Wave 12 = 16-34

Clearly, there are significant response biases by age (less so by journey purpose or gender). CR suggested adding survey weighting by gender and observable age, but we understand this was rejected by the SRA as it was thought likely to disturb the historic series (the NPS was stable in these terms - the weighted NPS age and sex profile figures did not vary by more than 1 point throughout Waves 2 to 8).

We recommend that weighting the NPS results by age should be reconsidered.

Wave 13 NPS procedure requires interviewers to record the gender, observed age and ticket type of all individuals who are given a questionnaire. We strongly recommend that this information should always be data-entered. The information can then be analysed against questionnaires received, and response rates can be calculated for each Wave.[35]

Without this exercise, it will be impossible to monitor response rate thoroughly by age, which is known to be a deficiency of the current approach, which is therefore a known source of potential bias (and one which may change over time and route). If it is then confirmed that it is necessary or worthwhile to proceed with weighting by age, the process should aim to restore in the final weighted sample of journeys the age profile recorded when questionnaires were issued.

Indeed, we strongly recommend taking this process a stage further and analysing response thoroughly by gender and ticket type - as well as by age. The shift number on the record form also allows response analysis by station size stratum, station type (i.e. city v. other), time of day, even individual station, or shift, or interviewer - or any combination of the above - or indeed many other factors. There would be additional costs, but they would be marginal to the total cost of the NPS.[36]

We recommend that such a thorough response analysis should be undertaken, as it could suggest ways in which NPS response rate, survey design and weighting all could be improved, which will

---

[35] In the past we understand only samples were analysed, and only occasionally.
[36] The data entry process will greatly facilitate the analysis of response rates within any interlocking cells.
We do not believe has been done so far, and could provide information critical to future NPS design.

increase the quality of, and confidence in, the data produced by the NPS. If this process helps to suggest ways in which the NPS can provide more representative results with less weighting, then the <u>effective</u> sample size of the NPS will increase <u>without any increase in actual fieldwork</u>.

Whatever detailed decision is taken about the above processes, in any case an adequately detailed response analysis should definitely be produced after each NPS wave.[37]

Although continuous measurement of the NPS refusal rate[38] may be impractical, we recommend that an ad hoc study should be carried out on this issue. Without any information on this factor it is difficult to establish the true value of the apparently high NPS response rate.

Finally, there is no doubt that the longer the interview, the lower the response rate. On Wave 13, the addition of the Central Trains supplement to the NPS questionnaire (see Appendix F) was accompanied by a response rate 4.7% lower than in the rest of the sample :

Wave 13 NPS areas:	NPS TOTAL	Central only	All other areas	DIFFERENCE Central v rest
Distributed:	82985	15666	67319	
Returned:	30497	5154	25343	
Response Rate:	36.8%	32.9%	37.6%	-4.7%

Source: Wave 13 response data, CR, 2006.

It is very difficult to justify any additions to the NPS questionnaire which result in such a significant drop in response rate (whether in one region, or overall).

---

[37] There is little or no information about response rates in most published NPS Reports. Even the NPS Overview Report for Wave 11, for example, contains only one response rate figure (34.5%).
[38] The NPS response rate is calculated by dividing the number of completed questionnaires returned by the number handed to passengers. The refusal rate in this instance would be the proportion of passengers approached who refused to accept a questionnaire and are not included in these response calculations.

4.3   Distribution of questionnaires

Many rail passenger surveys (e.g. some of those commissioned independently by TOCs) are conducted by distributing and collecting self-completion questionnaires on board trains. This method appears to work very satisfactorily for short questionnaires, although it may be much more problematic at peak times or on short journeys, and does lead to a high level of sample clustering (i.e. many respondents are on the same train).

However, the method used by the NPS, of distributing the questionnaires (and prepaid return envelopes) at stations prior to a train's departure, allows interviewers to distribute them to people travelling on a number of trains (hence less clustering), without the interviewer having to leave the station. This is a more cost-effective use of fieldwork time, especially at busier stations. It also has the merit of allowing respondents to complete the questionnaire at a time of their choosing when the journey is over (e.g. later at work or at home) which is an advantage when dealing with crowded journeys and with long questionnaires.

We believe that the present NPS methodological approach to this is satisfactory.

4.4   Semantic scales

There are many different views as to the most appropriate scaling technique to use to measure customer satisfaction, and there are many different approaches. They all measure perceived performance (rather than absolute performance such as timetable data on numbers of departures or punctuality). We are not aware of any conclusive evidence which suggests, categorically, which is the most desirable method. Given the lack of empirical evidence, there has to be an extremely strong reason for organisations to change a scaling technique they are already using, because to do so will inevitably result in a break in the valuable trend data.

Satisfaction Scales and Evaluative scales

Satisfaction scales are probably the most commonly used method, and this is the format used on the NPS satisfaction questions[39]. Satisfaction scales usually employ a semantic scale. The NPS employs a very widely used version i.e.:

Very satisfied/Fairly satisfied /Neither satisfied nor dissatisfied /Fairly dissatisfied /Very dissatisfied

There are various other versions of this scale, both semantic and numeric (e.g. "rate it on a scale of 1 to 10 where 10 is best"), and of various lengths. Satisfaction scales have been used since customer satisfaction surveys began.

Evaluation scales are another very commonly used method, and this is the format used on the NPS rating questions[40]. Evaluative scales also often employ a semantic scale. Again, the NPS employs a widely used 5-point version i.e.:

Very good/Fairly good /Neither good nor poor /Fairly poor /Very poor

Evaluative scales are becoming increasingly popular. As with the satisfaction scale, there are other versions of this scale, both semantic and numeric, and of various lengths.

The difference between the 7-point and the 5-point version of the two semantic scales above is that the 7-point scale usually adds an "extremely" position at either end of the scale. Both the above scales may also use a numeric scale (e.g. 1 to 7 or 1 to 10). Changing to a scale of 7-points or more (whether semantic or numeric) can produce slightly greater discrimination than a 5-point scale, but in most cases the differences in levels of discrimination are relatively minor.

Some argue that numeric scales are better than semantic scales. In our experience, people naturally respond semantically (e.g. "it was quite good"), and it can be argued that semantic scales require less explanation (which is especially useful in a self-completion context), and are easier to understand - for both the respondent and the client (e.g. "good" means "good" - but what does "7 out of 10" mean - is it better or worse than "good"?).

---

[39] This format is used for Q19 (journey characteristics) and Q27 (journey overall) (NB: Wave 13 numbers)
[40] This format is used for Qs 12 (tickets), 15 (station), and 20 & 21 (train) (NB: Wave 13 numbers)

Other Scales

As well as the above satisfaction scales and evaluative scales, there are several others: e.g. expectation scales, improvement scales, acceptability scales, agree/disagree scales and choice of "levels" of performance.

Some of these have limitations: for example expectation scales attempt to measure performance against expectations ("better than expected", "worse than expected" etc.) but give no indication of the level of performance, nor indeed whether performance is satisfactory. Also, this scale can be of diminishing value over time (e.g. if your performance improves to a high level in year one and remains there, your results on this question will improve in year one but fall in year two if performance is now "as expected"). Improvement scales have a similar problem. Acceptability scales do not evaluate quality of service or level of performance, merely what is "acceptable" (which is a weak descriptor). Agree/disagree scales are widely used both for attitude and rating purposes and are generally regarded as effective but offer no advantage in this context over the evaluative scales now used by the NPS. Option level choice is a technique in which a range of service standards are described (different ones for each parameter, e.g. the train was: 1 minute late, 2 - 4 minutes late, 5 or more minutes late) and respondents are invited to say which one was closest their own experience. This approach can be useful in certain situations but is unsuited to the needs of the NPS (as every rating question would be different - and longer).

Technically, the NPS fulfils a requirement for trip-based transactional research[41]. The satisfaction scales and evaluative scales employed on the NPS are appropriate for the purpose. Both scale types are very widely used, and both are generally regarded as effective. Based on the Consultation Exercise, the current NPS questionnaire is perceived by the majority of users of the NPS to be comprehensive and effective, and the NPS rating results are regarded as a credible, reliable and useful measure of trends. The differences between the merits of the two scales used on the NPS - and the merits of the best of all the other scales - are relatively minor.

The most important issue from the point of view of measuring service delivery is the trend data. It is the consistency of approach that is important in terms of tracking change.

We are certain that no benefit could arise from changing the scales in use on the NPS which would conceivably compensate for the inevitable loss of comparisons with historical data.

We strongly recommend that the NPS continues to use the existing scales.

---

[41] I.e. it is a survey of journeys at a particular time and not a survey of the population at large. The question of inter-modal comparisons (which is a general issue which often arises in discussions of scaling in many industries) would be very much better served by a population-based survey designed specifically to investigate this question, and which can cover users of all the relevant modes - and can also cover non-users of each mode. In contrast, harmonizing scaling on a range of trip-based transactional surveys (each with different sample profiles) would be of limited value (as results of each could not be projected to produce comparable population totals), and none would cover non-users.

4.5 Changes to the questionnaire

Based on our review, and feedback from the Consultation Exercise, we believe that all of the key satisfaction parameters are already included, and relatively few changes to the questionnaire need to be considered. None of the proposed changes is major.

Prior Interchange

The NPS questionnaire includes the following question about interchanges:

> Q2a   Did you continue your journey by train after getting off at *(station named at Q1b)*?
>
> > yes   - go to Q2b
> > no   - go to Q3

We recommend that a parallel question should be added re: the respondent's journey <u>to</u> the station at which the questionnaire was distributed. This will establish whether the respondent arrived at the interviewer's station by train (i.e. is already an interchanger). This will be of value in checking and analysing the NPS results and is likely to be of value in planning future surveys. For this purpose we need only duplicate Q2a above, but re: the prior leg (as Q2a already covers the subsequent leg).

The suggested addition is very short, very simple and is relevant to all. We consider this addition to questionnaire length is insignificant.

Prior Interchange and Other Connections

A more ambitious alternative to the above alternative, which should be considered seriously, is to expand on both the above questions (i.e. prior and subsequent). On some Waves, questions are already added to cover other modes of transport used on the journey before the surveyed leg (e.g. Wave 10 Q33) and after it (Q37). We believe that a strong case can made for retaining similar questions in all future waves.[42] The best solution might be to expand the existing (Wave 13) Q2a above, e.g. to a format such as this:

> Q (extra Q)   Which modes of transport did you use to get to the station where you were handed the questionnaire? (TICK ALL THAT APPLY)
> > LIST: train, bicycle, motorbike, car, bus/coach, train, etc. (e.g. as Wave 12 Q33)
>
> Q (new 2a)   Which modes of transport did you use after getting off the train at *(station named at Q1b)*? (TICK ALL THAT APPLY)
> > LIST: train, bicycle, motorbike, car, bus/coach, train, etc. (e.g. as Wave 12 Q33)

As well as providing the same information re: interchangers as the shorter version above (which will be useful for technical survey reasons - see Section 3.5) this slightly longer version would allow the NPS to provide, for the first time, useful data on the rail passenger's <u>complete</u> journey, including all major transport modes, public and private.

As the suggested addition is still fairly short, fairly simple and is again relevant to all, we do not consider this addition to the questionnaire length is a problem.

---

[42] But <u>not</u> the other Wave 10 extra questions: Q34 (re: how else they would like to have got to the station) or Q36 (re: other features they would like).

Cleanliness of Trains

We have noted that there are three questions on satisfaction with cleanliness of train at Q20 and at Q21. All three use the same scale:

Very good/Fairly good/Neither good nor poor/Fairly poor/Very poor

Q20   How would you rate the train you boarded for that journey in terms of: Cleanliness

Q21   Specifically thinking about the cleanliness of the train you boarded for that
journey, how would you rate it for:
a - The cleanliness of the inside of the train
b - The cleanliness of the outside of the train

We understand that the first version of the question (Q20) was regarded as insufficiently precise and Q21 was therefore added - but Q20 was not dropped. However, a brief examination of the Wave 13 results shows that Q20 appears overwhelmingly to be measuring <u>interior</u> not exterior cleanliness, as the Q20 "very good + fairly good" percentage is within 2 points of the Q21 equivalent "inside" percentage for all but two of the 26 TOCs in Wave 13, whereas it varies from the "outside of the train" result by up to <u>16 points</u>. Unless the Q20 version is essential for contractual reasons it is difficult to see why the Q20 version should be retained. If it is dropped, consideration should be given to moving the Q21a and Q21b items into its place at the top of the Q20 battery rather than presenting them as an unnecessarily separate question.

Apart from this there are no obvious candidates for elimination (based on the Consultation Exercise, most of the questions and most of the rating scales were used by most of the users, and all of them were important to some of the users).

Overall Satisfaction

The overall satisfaction question is currently located at Q27 (after six pages of questions, including all the satisfaction and rating questions re: journey, station, and train):

Q27 Taking into account just *(station name)* station where you boarded the train and the actual train travelled on after being given this questionnaire, how satisfied were you with your journey today?
[Very/Fairly satisfied/Neither satisfied nor dissatisfied /Fairly/Very dissatisfied]

The question has been raised as to whether this question should be placed elsewhere in the questionnaire. There is a certain amount of debate among practitioners of customer satisfaction research on this issue. Probably the most widely held view is that if such "overall satisfaction" scales are asked at the end of the questionnaire, then preceding questions may influence response, and we believe this to be true (certainly, it has been shown in very many studies that question order and list order affect response).[43]

---

[43] It should be noted that question order is slightly less of an issue for self-completion surveys than for administered interviews because the respondent may browse the whole questionnaire before completing it.

In the case of the NPS, we believe that Q27, the overall satisfaction question, should stay where it is. Almost all of the questions prior to Q27 relate to the journey experience of <u>every</u> passenger[44] so little or no differential bias is likely to be introduced. Also, all the questions prior to Q27 have hitherto remained largely unchanged, and therefore for the sake of consistency of the time series, we recommend that Q27 stays where it is.

Even so, if Q27 remains where it is, it must be remembered that it will remain important that in future there are no significant increases in numbers of questions (or major changes of content) <u>prior</u> to Q27, and in particular that no questions are added which might introduce differential bias. Any occasional sets of questions (e.g. the access to network questions as on Wave 10, or the security questions as on Wave 13) or questions aimed at particular subgroups, should continue to be placed <u>after</u> Q27.

Overall Length
We understand that Central Trains were given permission by the SRA to add (at their cost) an 8-page extra section to the NPS questionnaires issued in their own area (taking the total questionnaire length from 12 to 20 pages - see Appendix F). The response rate in the Central area was 4.7% lower than in the rest of the sample (see Section 4.2). We would advise against repeating this. Given that the <u>standard</u> NPS questionnaire might reasonably be regarded as being long already, we recommend strongly that it should not be allowed to become significantly longer - either overall - or in any one region.

Other Possible Changes or Additions
The question order of the NPS questionnaire might well be slightly different if we were starting again from scratch (this is a hallmark of many continuous research questionnaires which have evolved over time) and it might be possible to improve the wording in a few places. However, there are no questionnaire issues which appear serious enough to absolutely require change.

E.g. there is a line in Q15 re: "The **attitudes and helpfulness** of the staff" (at the station) and one in Q20 re: "The **helpfulness and attitude** of staff on the train". It is not an important point, but why do these two questions not use exactly parallel phrases?

Similarly re: additions to the questionnaire, there are no obviously major gaps, but there are no doubt several minor items which might be suggested, for example:

Q6 asks if passengers were travelling alone or not, and whether they were accompanied by children (of various ages) or not, but fails to ask how many people in total were in the party. This could be a useful addition (e.g. if trying to estimate total passenger numbers) but is not a critical issue.

It would be very simple to add a short question asking how long the journey leg took from start to finish. Satisfaction may vary by duration, and this question would allow the factor to be analysed and better understood. Also it would also allow direct comparisons between TOCs on their <u>actual long distance journey ratings</u> (much better than comparing only "Long Distance TOC" averages, when so many of them are clearly "mixed").

---

[44] The main exceptions are Q8b and Q8c which are asked only of those with long term illness or disability.

4.6   Sampling plan

4.6.1 TOCS and stations

As stated earlier (Section 3.6), we believe that the process employed by Continental Research for allocating the number of shifts to specific stations appears to be very robust. We have no recommendations for specific changes to this aspect of the NPS design.

However, we do recommend that the standard of documentation should be improved. For example, any changes made to the volume estimates generated from the CAPRI/Lennon database should be documented and the reasons recorded (e.g. why are the substitute estimates regarded as better). Without this it is difficult to confirm that changes were made in an objective fashion and that integrity of the methodology has been preserved.

4.6.2 Shift patterns

This is an area of weakness in the current NPS sampling method. We recommend that changes are necessary to three elements:

a)   the mix by time of day and day of week
b)   shift length and phasing
c)   documentation

a)   Time of Day and Day of Week

The original NPS shift pattern appears to have been designed to produce a reasonable mix of peak v. off-peak journeys, which it broadly appears to do. However, the design falls short re: time of day and thus also by direction of travel (outward v return) and the sample is badly skewed in both respects, i.e. over-representing outward/am/peak and under-representing return/pm/peak (see Section 3.4). The use of previous NPS results as the basis for planning future NPS shift patterns is a circular solution which has allowed this anomaly to persist. We cannot be certain until this issue is resolved but it appears to have had only limited effects on ratings. However, the am/pm bias varies by TOC and further work will be necessary to establish the precise effect on scores.

In any event it will definitely be necessary to change the way fieldwork shifts are allocated (within station) to day of week and time of day. Fortunately the National Rail Travel Survey (expected later in 2006) will provide the best source so far available of universe data on passenger departure volumes by station, hour and day, and this will allow the new shift allocation scheme to be drawn up with much greater precision and confidence than has been possible before.[45]

If, after further work, it appears that there is risk of a discontinuity in the NPS trend results when the new shift allocation method is introduced, this can be minimized if the existing sample structure is replaced gradually (e.g. by changing a quarter of the shift pattern from the old to the new method in each of four consecutive waves) or possibly by switching over immediately but weighting to smooth the results.

---

[45] We understand that NRTS covers only Monday to Friday, but this accounts for the majority of traffic. Rail Planner data plus TOC by TOC estimates will be used for the Saturday and Sunday design. Rail Planner could also help improve the efficiency of the shift pattern at smaller stations.

b)   Shift Length and Phasing

We strongly recommend two other changes to the shift pattern: phasing them rather than starting them all simultaneously, and (much more important) using shorter shifts, e.g.:

current 3-hour shift, approx: 1800 shifts = 30,000 interviews at 650 stations
new 90-minute shift, approx: 3600 shifts = 30,000 interviews at <u>about</u> 1000 stations[46]

We believe this change is very unlikely to create any discontinuity in the survey ratings, but will result in <u>very</u> significant improvements in the representativeness and reliability of the survey data at the station level and the time of day level (see Section 3.7).

c)   Documentation

The standard NPS TOC reports contain little or no technical detail, and the NPS Overview Report (which does) has received very limited distribution. We recommend that the accessibility and content of NPS technical documentation should be improved.

4.6.3 Range of hours

The range of departure hours now covered by the NPS is 7am to 10pm. This is a wider range than was covered in early NPS waves. Volume outside this range is relatively small and is expensive and problematic to survey, and we do not propose any extension of the range. However, the stated range may be somewhat misleading as only 1% of NPS shift hours occur between 7pm and 10pm (from CR Wave 13 shift data) and only 1% of NPS journeys depart after 7pm (from NPS Wave 13 unweighted data) whereas we would expect the real total to be somewhat higher[47]. Also, for whatever reason, the NPS response between 8pm and 10pm is negligible. The cost and value of surveying the early and late hours should be examined, and a decision taken re: the range of hours over which a <u>representative</u> sample should be obtained. It would be useful in this context to have an analysis of the response rate by hour for weekdays versus weekends.

4.6.4 Interchangers

The definition of journeys used in the sampling methodology (i.e. "legs of trips including interchangers") differs from that employed in practice in the fieldwork methodology (those <u>beginning</u> journeys at sampled stations are included in survey, but interchangers are known to be under-represented) and this fact may introduce distortions. The issue should be addressed. We have already recommended adding a questionnaire item to obtain better data on this subject (see Section 4.5). Further analysis will then be possible. At that point it will be possible to evaluate alternatives. One future possibility is that the interchangers are removed from the universe data used to compile the sampling frame (this effectively alters the definition of journeys in the sampling frame but maintains consistency between the sampling frame, fieldwork methodology and questionnaire), but analysis of the data may suggest other solutions.

---

[46] Because some of the extra shifts will become extra shifts at stations already covered, doubling the number of shifts will not double the number of stations to 1300 (the number covered could be adjusted by varying the proportions of shifts at stations of different sizes, but this would require corrective weighting).
[47] Based on LATS, 4.5% of passengers leaving London & SE stations depart between 7pm and 8pm alone.

4.7   Weighting

The Passenger Focus brief included the request that we look for possible improvements to the weighting of the survey, the better to represent passengers using each TOC. In many situations, less weighting is better weighting (as this can limit design effects, and thus preserve effective sample size).

We have therefore looked at ways in which the sample design can be optimized in the first place. If the NPS sample by time of day (etc.) can be made more representative, and response rate bias can be limited, then less weighting may be required.

It is not possible to produce full and definitive recommendations on the "best" weighting options until such sampling issues have been resolved. However, it is possible to suggest what we regard as likely future directions.

4.7.1 Station size stratum

Station size is a key factor in the journey universe and any distortions in the sample could have significant implications. There is a strong case for retaining this weighting variable. Weighting by station size within TOC is required to correct for any distortions in the sampling frame arising from whatever cause.

4.7.2 Journey purpose

The present method of weighting journey purpose using ticket type data as a proxy should be abandoned. There is clearly a mismatch between these two variables. Proposed changes to the sampling methodology may remove the need for this weighting (see Sections 3.19 and 3.20). If not, there is a better alternative (see 4.7.7).

4.7.3 Day of week

The data expected to be available from NRTS is likely to remove the uncertainties about sampling and weighting issues re: weekdays, but Saturdays and Sundays (which have different patterns) will remain a problem and will need to rely on the TOC's own volume estimates. If, as seems likely, there are wide variations in weekend volume profiles by TOC, then weighting of weekday v weekend (and, we would argue, weighting of weekday v Saturday v Sunday) will probably continue to be necessary and appropriate.

4.7.4 TOCs

The fact that each TOC is given a target NPS sample size requires weighting to allow for this in producing correct national totals. This is appropriate and should continue.

However, our Technical Review has indicated that certain TOCs have larger design effects than others, e.g.: Island Line, Wessex and Gatwick Express. There appears to be a greater mismatch between the shifts sampled and the expected distribution (from weighting) for these TOCs but we have been unable to resolve this in the time available. This requires further study.

Also, the anomalous weighting of GNER should be reconsidered.

4.7.6 Other weighting issues

The NPS sample is designed to produce representative data at the TOC level and the national level. Some users (e.g. PTEs), and possibly many more users in the future, may wish to analyse the data at sub-TOC level. This is not a level at which the survey sample was designed to operate, and it is not a demand which can be answered by weighting. It would require significant changes to the sample design: such changes are not impossible to consider, but it must be recognized that they would introduce significant complexities to the TOC level building blocks which form the basis of the existing NPS sample design. In particular, to be effective, they would almost certainly require a substantial increase in the number of local stations sampled per NPS wave, and in the case of smaller areas they would also require a substantial increase in sample size, which would greatly increase cost.

It is theoretically possible to weight NPS data by stations or by routes. In practice the NPS is already weighted by station stratum within TOC. It is uncertain how much benefit could be gained by weighting individual small stations (but if such weights are large, the design effect would increase and effective sample size would fall). Without a reliable source of universe data on route volumes, and given the enormous complexity of the rail network, there is even greater doubt about the practicality of weighting by route. However, the proposed reduction in shift length could improve the quality of the sample at station level (and thus indirectly also at route level) enough to make such weighting issues irrelevant.

4.7.7 Future weighting

Once it has been decided how the sampling design should evolve (see Section 4.6) and if results are available from the detailed response analysis we have recommended (see Section 4.2) and other work (e.g. Section 4.7.4), it will be appropriate to conduct a further review of the weighting. Meanwhile, we believe that it is likely that it will be decided at that point that it is sensible to weight by station size, TOC, and day of week (as above) and possibly by age (or age within gender) as well.

Subject to the effectiveness of any changes to the shift allocation design and procedures, it may be recommended that it is advisable also to carry out weighting by time of day.

We are hopeful that the changes we have recommended will result in weighting by journey purpose becoming unnecessary, but if there are perceived to be deficiencies in the journey purpose profile at that stage, it is possible that weighting of ticket type by ticket type (rather than weighting journey purpose by ticket type proxy) may be reconsidered as a solution, as ticket type volume data is one of the few independent sources of universe data.

4.8   Technical documentation

The national rail network is enormously complex. The number of internal parameters within which significant variations in traffic patterns may arise are numerous. Very substantial variations in the volume, purpose and direction of travel (and the profiles of the passengers) may arise by station, route, hour of the day, day of week and date.

As the subject of research it therefore presents enormous challenges. It is unsurprising if the design of a survey intended to present a representative picture of the traffic on this network is likely to be complex.

In such circumstances we recommend that it would be appropriate for the NPS reporting process to include an above-average standard of detailed methodological documentation sufficient to allow an NPS user to make an informed judgement on the validity of the results produced.

In our view this means that methodological documentation should be available which includes adequate details of procedures for sampling, weighting, fieldwork and fieldwork checking, analysis of the sample profile and response rates, and a copy of the questionnaire.

The NPS Overview Report comes closest to meeting this specification. However, it is not produced after every NPS Wave, and has been given very limited distribution. We recommend strongly that it should be produced for every Wave and should be made much more widely available. It should also include more detail on certain technical aspects of the survey design and performance, e.g. response rate data (in the Wave 11 Overview Report only one national response percentage is given).

4.9   Cost-effectiveness

We have not been able to make any recommendations which set out specifically to reduce the cost of the NPS. The survey already employs the self-completion method, which is a relatively low cost option and is suited to the objectives of the NPS. The survey frequency and the TOC sample sizes are predetermined, and these set the overall scale of the research.

However, it is to be hoped that the changes to the sampling, weighting and response analysis procedures which we have recommended will improve the reliability and representativeness of the NPS. To the extent that these changes reduce the design factor and increase the effective sample size, this will produce a measurable improvement in the survey's cost-effectiveness.

If this improvement is sufficiently large, it could possibly even allow the overall NPS base size to be reduced slightly without loss of accuracy, resulting in cost savings.

On the other side of the cost-benefit equation, we believe that much greater use could be made of the NPS, which could enhance the benefit derived (see 4.10).

4.10 Development and communication

Several relevant issues emerged from the Consultation Exercise. A general theme which links some of these issues is the evidently patchy knowledge of the NPS (e.g. questionnaire content, technical details, analysis options, etc.) among some of those consulted. Part of this may be accounted for by the fact that some of the consultees make very intensive use of the survey, while for others it is one of a number of sources scanned. At the same time, the history of the NPS appears to have been one in which most of the communication has been with the intensive users and little appears to have been done to create user-friendly reports which would be more widely accessible.

We believe that a great deal more valuable use could be made of the NPS data, both by existing consultees and by a great number of other people.

"It's amazing how much is in it." (TOC)

There are various ways in which we believe the data can be exploited more intensively, communicated better, and used more widely and effectively. These could include:

- improving the standard and range of reporting generally
- exploiting the existing data more thoroughly
- improving communication with users
- improving awareness among users and potential users
- supplementing the existing data

We suggest that these goals could be furthered in a variety of ways:

a) Standard NPS Reports

We recommend that the existing NPS TOC Reports and NPS Consultees Reports continue in roughly the present format, as they contain the chief top-line information, and in a familiar format. However, we suggest that the reports should be expanded to include:

(i) additional tables giving a basic analysis of other key data collected regularly[48].
(ii) a more detailed explanation of methodology, response rates etc.[49]
(iii) a copy of the current questionnaire.[50]

Also, given the issues surrounding the relationship between journey purpose (and ticket class) on the one hand, and satisfaction ratings on the other, we feel that greater attention should be paid to each TOCs performance <u>within</u> journey purpose and <u>within</u> ticket class, rather than concentrating too much on their overall average performance versus the equivalent average for their "TOC type" (long distance, regional, London/SE), as these TOC type distinctions are becoming less clear cut.

---

[48] TOC Reports are accompanied by a PDF set of extra tables but we believe Consultees Reports are not.
[49] At the very least, the NPS Overview Report should be available to all NPS users.
[50] Currently distributed electronically, but not bound into TOC Reports or Consultees Reports.

b) Extra NPS Reports

The NPS already contains a great deal more data than it is practical to include in the standard reports above. We recommend that from time to time special reports are produced based on special analysis and dealing with particular issues in more detail (including a narrative commentary and description of the methodology). Based on the range of data already collected by the NPS, this allows a wide scope of topics. This would incur extra cost (extra analysis is cheap so the main cost would be the narrative report) but could be of great value: it could raise awareness of the scope of the NPS, exploit its databank, and be of assistance to the rail industry in continuing to improve their standards of performance for the benefit of all passengers. We feel this is more likely to be achieved if user-friendly reports are available, rather than just raw data.

c) Reports on Extra Questions

Some NPS waves have included extra questions (e.g. on security issues on Wave 13). We have not seen the format in which these results were published, but some consultees seemed unaware of their existence. We believe that there are several other such subjects which could usefully be added to the NPS every few years (but not every wave). Again, we feel this is more likely to be achieved if formal, user-friendly reports are available.

d) Analysis of the Importance of Specific Scales

The issue has been raised as to the relative importance of the various scores measured by the NPS. Some work has been done in the past on this issue, but we recommend that it is made a more thorough and more visible exercise.[51] Correlation analysis of existing NPS data can be carried out cheaply (e.g. costing hundreds of pounds rather than many thousands). There can often be great benefit in analysing customer satisfaction correlations, particularly by level (e.g. among those with high scores v those with low scores, as positive and negative drivers usually differ) or by passenger type (as different categories probably differ in their priorities). This will be more valuable if commentaries are produced (rather than just raw data) and regularly published. This would help improve understanding of the NPS data, increase confidence in the results, and allow importance to be monitored continuously (so that trends are more likely to be detectable).

e) Supplementing the Data

In our experience such correlation analysis work is likely to be a valuable exercise. However, this is not able to answer all the questions which might be raised about importance, and occasional ad hoc research (e.g. qualitative research looking in depth at individual scale questions and related issues) will continue to be worthwhile. There are various subjects which might merit specific studies (either qualitative or quantitative). To the extent that they improve understanding of the relationship between the performance of the rail industry and the satisfaction of its passengers, or help enhance the design of the NPS, they can contribute to the overall goals of Passenger Focus.

f) Extra Analysis

---

[51] After each NPS Wave a single-sheet multivariate factor analysis summary is sent to PF (without commentary) which shows results for GB and for each TOC. These indicate that, typically, 51% of all factor variance is accounted for by punctuality alone. However, this is now a "known known". We believe it would be sensible now to examine the equivalent result for those journeys which are on time. It would also be valuable to examine the correlations between individual scores and overall satisfaction.

A few users already exploit the NPS through extra analysis. Much more use would probably be made of these options if they were more widely known. A further extension of the Extra Analysis facility might be to provide sortable access to the verbatim comments made on the NPS questionnaires.[52] For example, TOCs could sort for respondents who said "very poor" on a particular question and then look for comments on <u>why they rated it so low</u>. We understand that Continental Research is able to provide this facility (or soon could be).

We are assuming there is no significant <u>extra</u> cost to Passenger Focus as we imagine all such extra analysis (or copies of the database) are or would be paid for by the user at cost. In fact, in other similar situations we believe that it has been found reasonable to charge cost plus a margin to contribute to the cost of producing and maintaining the database.

In any case, we recommend that it would be prudent for Passenger Focus to produce guidelines re: the use of extra analysis in terms of the reliability of certain forms of data, sample sizes etc., and rules re: copyright and publication.

g) A Guide to the NPS

Some NPS users have regular contacts with Passenger Focus and Continental Research and exploit the NPS heavily, while others we consulted do not. Knowledge of the NPS among a wider audience is probably very low. We believe that producing a user-friendly "guide to using the NPS" would be a very worthwhile exercise. It should include guidance on confidence limits (see Section 3.21).

h) NPS User Group

We believe that some sort of NPS user group could also perform a useful service in disseminating information about proposed changes to the survey, extra questions etc. and obtaining feedback from key users prior to finalisation.

i) NPS Newsletter

Finally, without wishing to add indefinitely to Passenger Focus's costs, we suggest that an NPS newsletter (perhaps 2 or 3 times a year) could be a useful exercise, drawing attention to any new or forthcoming NPS surveys or reports and their expected publication dates etc.

j) Cost Recovery

Most of the suggestions made in this section would add to costs - some to a greater extent than others. We have already suggested ways in which revenue might be generated (or at least costs recovered) in the area of extra analysis. Perhaps further costs could be recouped by charging for copies of the new reports and newsletters we have suggested.

---

[52] We understand that TOCs receive a CD-ROM of verbatim comments to Q45 only. We believe that PF should consider extending this facility to include sortable online access to open-ended comments elsewhere in the questionnaire.

4.11 The Scope of the NPS

Other TOCs

It was suggested to us during the consultations that the NPS should cover all TOCs, i.e. including Hull Trains and Heathrow Express - and possibly also Eurostar, plus the DLR and other light rail systems. We cannot comment on the non-research issues which may be involved, but certainly Hull Trains and Heathrow Express could be added without technical difficulty to the NPS as they both operate to stations which are already liable to be included in the NPS sample (although there would be added fieldwork costs).[53]

Fieldwork Period

NPS fieldwork now covers 10 weeks in spring and 10 weeks in autumn, i.e. 20 weeks per annum. This is an improvement on the original NPS design which covered only 6 weeks per annum - but the present plan still leaves 32 weeks of the year unsurveyed (including the whole summer, when patterns are different - for some TOCs more than others).

We believe there would be no need for significant added costs if the NPS were to change to covering up to around 48 weeks per year, i.e. two waves each of 24 weeks.[54] Exactly the same sample design and sample size could be used for a wave of around 24 weeks as for one of 10 weeks (all that is required is to spread the fieldwork shifts evenly over the longer period). The same reporting frequency could be maintained (2 per annum[55]), but the survey would now cover almost the entire year.

Discontinuity may be a concern. We have no information as to how it was dealt with when the NPS changed from 3-week waves to 10-week waves. If it is a concern (and it may well be) it will be possible (a) to introduce the change in stages (e.g. extending waves from 10 to 15 weeks initially, and so on) and/or (b) to continue producing data for the "old" (shorter) fieldwork periods for purposes of comparison as well as for the full extended period (although the base size for the shorter period alone will be smaller and will contain fewer "station/shifts").

The only other concern we can imagine is that each half year wave will include data relating to performance up to about 6 months ago rather than up to only 3 months ago as at present. If this is a problem it can be solved as stated at (b) above - i.e. analysis can be produced based only on the most recent 10 weeks of fieldwork (as at present).

The benefit is that the extended NPS (i) will be able to give a picture of the whole year-round situation, and (ii) will also be capable of being analysed by any of the four quarters or four seasons as required, none of which is the case now.

---

[53] We understand that Heathrow Express may be included in Wave 14. Including Eurostar or light rail would raise a number of research design problems, as well as adding to fieldwork costs.

[54] I.e. excluding Bank Holidays and adjacent days, as fieldwork costs then would very probably be higher.

[55] In fact top line quarterly analysis data could be produced at marginal extra analysis cost. Analysis of periods shorter than 10 to 12 weeks for a single TOC would be less reliable. However, assuming the sample design is evenly structured throughout the year, rolling 6- or 12-month average data per TOC (or nationally) could be produced at any time with no loss of accuracy compared to previous waves.

Appendices

Appendix A - the project team

The project team comprised:

Peter Bartram, FMRS
Mary Bartram, FMRS
Gary Bennett, MMRS
Richard Roberts-Miller, FMRS

All parts of the project were carried out by members of the project team. All are members of the Market Research Society, and adhere to the Market Research Society Code of Conduct. The following pages set out each member's background and experience.

Peter Bartram – PB Consulting

A senior marketing research consultant with an understanding of service industry marketing and management needs:

- **Experience in marketing and general management,** delivering advice which is relevant to operational and strategic realities
- **Technical expertise** in quantitative and qualitative research of all kinds, ensuring effective design and execution
- **Broad knowledge of the UK and global market research industry**, ensuring that suppliers are chosen and used cost-efficiently

Offering advice, project management, and technical appraisals, with special focus on difficult/complex/sensitive markets and product fields, and projects related to customer service quality, employee motivation, delivery channels, corporate positioning/branding, product development, communications and all marketing management issues.

Extensive experience across all industry sectors, but mainly specialising in work for service-oriented organisations. Projects have included:

- **Travel and Transport**: Service quality measurement and product/service development research for tour operators, airlines, shipping lines, road, rail and other transport organisations
- **Professions and Public Sector**: Client satisfaction measurement for leading accountancy firms, expert witness assignments for leading law firms, service quality measurement for healthcare and educational institutions, local councils and other public sector organisations
- **Communications**: Communications effectiveness studies for publishers and other media, film industry studios, IT companies, utilities, industrial distributors, and many others

Leading clients have included Thomas Cook, The Orient Express, The Consumers' Association, Regional Railways North-East, Avis Car Rental, British Airways, American Express, Reed Exhibitions, Eurailpass, BUPA, PPP, Automobile Association and many others.

Senior-level expertise is derived from previous career experience as managing director of Harris Research and chairman of Applied Research & Communications, president of City Research Group, and other senior positions with NOP, American Express, and Thomson Newspapers. A former Chairman of the Market Research Society, convenor of many seminars and educational workshops and author of many papers and textbook contributions on research and its applications.

Mary Bartram

Mary Bartram is a graduate of London University and has had a long career in market research spanning more than 35 years. She has worked on all sides of the industry: she has held directorships in major market research agencies in the UK (NOP, RSGB and BMRB); she has headed the market research departments of leading advertising agencies on both sides of the Atlantic; and she was Head of Market Research for the English Tourist Board for 4 years, and for British Airways for 7 years.

In the latter position, she spearheaded the customer service research programmes which have underpinned the transformation and success of the airline over the last two decades, and she has been generally acknowledged as a leader in establishing the research techniques now used so widely in both the public and private sectors in the UK.

For the past ten years she has worked as an independent consultant, working with a wide range of clients including government departments and agencies.   For instance:

- she has provided extensive advice to British Telecom on the design and execution of their customer satisfaction tracking research programmes;
- she recently conducted an extensive review of TfL's customer satisfaction research programme, covering both London Underground and London Buses;
- she has designed continuous tracking research programmes for leading banks, for Guinness, and through Speedwing, the consulting arm of British Airways, for airlines in Africa, Asia and elsewhere;
- she has acted as an expert witness for a leading law firm;
- she was a member of the 'Panel of Research Experts' which advised the Consumers Association on their market research surveys conducted for publication;
- she has had a long term contract with the Ordnance Survey as their Market Research Consultant, helping to ensure that their services meet the needs and expectations of their various customer groups.

She is a Fellow of the Market Research Society, serving on its governing Council for 4 years, and being elected its Hon Secretary/Treasurer in 1987.

She has convened and chaired many committees and seminars related to customer service issues. She has had several papers published on such topics, and has been a member of the Market Research Society's Awards Committee.

Gary Bennett – Logit Research

Logit Research was formed in March 2004 by Gary Bennett, an applied Market Researcher with around 15 years experience in the UK Market Research Industry, including half a decade in a Director role. Logit Research is an Independent Marketing Research Consultancy, primarily focused around adding value to research through the use of advanced Multivariate Statistical and Modelling techniques. We provide a full-service consultancy to Market Research Agencies, Field & Tab and Data Analysis Companies. We also work with end-users of research, both directly and in conjunction with partners. Logit Research offers a portfolio of complementary tools and services and provide thoughtful and tailored advice to our clients. The statistical and modelling techniques we offer include:

- CHAID Analysis
- Latent Class Analysis
- Data Reduction
- Structured Equation Modelling
- Optimal Scaling
- Brand Mapping
- Discrete Choice Conjoint
- Multivariate testing
- Advanced Regression Modelling

We also provide ad-hoc consultancy around more "nuts and bolts" research issues such as sample frame design and weighting, drawing on around 15 years experience in the Market Research Industry, including three years managing on-going tracking studies on the transport network for TfL.

Gary's career includes over five years at MORI and more recently nearly two years at RS Consulting. Gary has also worked at Benchmark Research and as a client-side Market Researcher/Analyst at London Transport (now TfL). In the early 1990's he worked as an Independent Consultant on a team conducting Organisational Reviews.

Over recent years, Gary has specialised in Structured Quantitative Research Techniques and applied new implementations of Choice Modelling and Segmentation techniques in commercial settings. In 2002 he studied at the EPFL in Switzerland, under the tuition of among others, Professor Daniel McFadden, the Nobel Laureate who pioneered the use of Choice Modelling Techniques. Gary is a Full Member of the Market Research Society.

Our direct clients include Accenture, Shell, MORI, ICM, Mercer Human Resource Consulting, RS Consulting, Insight Research, Acumen, Intrepid Consultants, Evo Research, Sweeney-Pinedo, Benchmark Research, Infocorp, Hall & Partners, 2cv: research, Spring Research, Business Advantage, Prodata Partners and Facts International. Indirectly we have undertaken work on projects for organisations as diverse as BT, Motorola, EA Sports, DHL, Microsoft, ICI, Somerfield, Pepsico, Sony, Areeba, The Asset Skills Council, The Electoral Commission, The Ombudsman for Estate Agents, NHS, Unite (Student Accommodation Service).

Further details can be found on our web site: http://www.logitresearch.com/

Richard Roberts-Miller - RMA

Richard Roberts-Miller established RMA (Roberts-Miller Associates) in 1989, as a management, marketing and research consultancy specialising in travel and tourism. Since then, RMA has carried out a wide range of research and consultancy projects for national tourist boards, tour operators, travel agencies, transport companies (airline, cruise line, rail), hotel groups, cottage and villa companies, as well as many projects in other sectors. Roberts-Miller's background includes work on several major continuous surveys (e.g. UKTS, BNTS and HBI/HBS in travel, the NRS in media, and the single-source TGI). He has extensive experience in the area of customer satisfaction (including developing the Thomson "CSQ" system which has been widely imitated by other tour operators). RMA projects have included:

- quantitative and qualitative surveys
- consumer research surveys
- business to business research surveys
- international research surveys
- customer satisfaction surveys
- brochure and advertising research
- survey design consultancy
- marketing development studies
- branding exercises
- product launches and re-launches
- advertising planning and media buying projects
- management studies
- takeover and merger projects
- investment project evaluation studies

Roberts-Miller is a BSc Economics graduate of Southampton University. He is a Fellow of the Market Research Society, a Fellow of the Royal Geographical Society, a Fellow of the Tourism Society and a Fellow of the Institute of Travel & Tourism. He has delivered papers to MRS, ESOMAR, the Marketing Society and ADMAP on topics such as: research for travel, research for publishing, and effective market research buying. He served on the Jamaican Government's Marketing Advisory Committee for Tourism and their Tourism Advisory Council, and was presented with their "Blue Mountain Award" for services to Jamaican Tourism. Prior to establishing RMA, Roberts-Miller worked at the British Market Research Bureau (research executive), Times Newspapers (senior research executive), The Thomson Organisation (group research manager), Thomson Travel (research and planning manager), Thomson Holidays (marketing controller), Thomson Travel (board member), Thomson Vacations (President & CEO), Thomson Travel Inc. (President & CEO) and International Thomson (US) Inc. (Executive VP).

RMA (Roberts-Miller Associates)
PO Box 104 • Dorking • Surrey RH5 6YN
E-mail: richard@roberts-miller.co.uk
Phone: 01306-741368   Fax: 01306-741356

Appendix B – the review method

The review consisted of four elements:

Study of Documentation
Consultation Exercise
Technical Review
Conclusions and Recommendations

The work on the Study of Documentation was shared between the project team members, The Consultation Exercise was conducted by Richard Roberts-Miller and Peter Bartram (with input from the other project team members). The Technical Review was conducted by Gary Bennett (with input from the other project team members). The evaluation and reporting stage was co-ordinated by Richard Roberts-Miller, acting as project manager, but involved input from all four team members. Each element is described below.

Study of Documentation
The brief suggested that the NPS Review project should consider a range of written documentation, data and information relating to the NPS. This work continued throughout the NPS Review project. At the beginning of the project, early in November 2005, we were provided by Passenger Focus with a wide range of documentation relating to the NPS. Further documentation was provided by Continental Research. This material was reviewed by members of the Project Team. This in turn led us at various stages to request further information from Passenger Focus, CR and others, and to carry out extra analysis of the NPS and other data sources. This work was not reported separately but contributed to the design of the Consultation Exercise (Section 2), to the work done in the Technical Review (Section 3) and provided input to our Conclusions & Recommendations (Section 4). For a list of documentation reviewed see Appendix E.

The Consultation Exercise
Passenger Focus provided us with (i) a list of TOCs[56], in each case with contact details for one or more individuals who are named on the TOC Report circulation list (or were known to be users of NPS data) and (ii) a list of individuals at other organisations receiving NPS Reports, plus (iii) contact details for the NPS contractor, Continental Research.

In the case of the TOCs a purposive selection was made which was intended to produce a representative range of TOCs - i.e. to include: large, medium and small; long distance, regional, London/SE; and England (north/midland/south), Scotland and Wales. In the case of the other stakeholders we endeavoured to consult a representative range of the regional bodies, plus as many of the national organisations as possible in the time available, plus the NPS contactor.

All the consultations were conducted by Peter Bartram and Richard Roberts-Miller, using the discussion guide shown in Appendix D, at a time of each consultee's choosing, between the 7th and

---

[56] Excluding the following: light railways, metros and trams; Eurostar, Heathrow Express and Hull Trains (which have different operating agreements to those of the other "heavy" TOCs), and Northern Ireland (not covered by NPS).

21st of December 2005. There were no refusals (although in one case the individual nominated a substitute as having more involvement with the NPS). Some consultees asked if one or more of their colleagues with an interest in the NPS could attend as well: consequently six of the face-to-face consultations were with groups (each of between two and six individuals). All consultees agreed that their sessions could be recorded for analysis purposes only, and all agreed to be named in this report as having participated. The contents of the consultations have been treated as anonymous (it was explained to consultees that some of their remarks might be quoted in our report for illustrative purposes, but only on the basis of anonymity).[57]

In the time available we were able to complete 23 consultation sessions with a total of 37 individuals. 14 sessions were face-to-face and were conducted at consultee's offices (individual consultations typically lasted about 50 minutes although all the group sessions were longer than that) and 9 were by telephone (typically lasting 20 to 30 minutes). In addition, written comments were volunteered by two other NPS users. In total feedback was received from 39 individuals representing 17 TOCs and 13 other stakeholder organisations. The feedback was evaluated and a draft report on the findings was prepared and circulated to the project team, and formed the basis of Section 2 of this document. For a list of those consulted see Appendix C.

The Technical Review
The main part of the statistical analysis of the sampling and weighting was carried out by Gary Bennett, who drafted a report on the findings and circulated it to the project team. That formed the basis of Section 3 of this document.

Conclusions and Recommendations
The project team evaluated the draft report on the Consultation Exercise and the draft report on the Technical Review and agreed the final versions of the two reports, and also produced an agreed series of Conclusions and Recommendations. The end products are included in this document as follows:

Consultation Exercise - see Section 2
Technical Review - see Section 3
Conclusions and Recommendations - see Section 4

---

[57] Initially, consultees were offered the option of their remarks being confidential and anonymous (rather than attributed), not because the topic was thought to be contentious (it was not expected to be), but because this approach has been found more likely to encourage a frank and easy exchange of views. Some consultees preferred anonymity, while others would not have objected to being attributed (although a proportion of those wanted to have sight of the text before publication - which we did not believe was practical given the timetable) but none objected to the anonymity option, and the consultations proceeded on that basis. This also eliminated the problem of identifying individuals in the group sessions.

Appendix C – the consultees

TOC Consultees
Arriva Trains Wales, Guinaz Daud, Market Development & Analysis Manager (T, RM)
Central Trains, John Sadler, Head of Revenue Strategy (F+, RM)
Chiltern Railways, David Whitley, Marketing Manager (F, RM)
First Great Western [First Great Western, First Great Western Link], Jeremy Clarke, Customer Research Manager (F, RM)
First ScotRail, Kenny McAlpine, Research & Pricing Manager (F, RM)
GNER, Tom Hanham, Service Specification Manager (T, RM)
London Lines [WAGN, c2c, Silverlink], Daniel Ward-Murphy, Market & Customer Service Planning Manager (F, RM) +
Midland Mainline, Richard Todd, Marketing Strategy Manager (F+, RM)
National Express Group PLC, Caroline Teo, Group Head of Research* (F+, RM)
National Express Group PLC, David Bird, Group Customer Service Director (F+, RM)
South West Trains, Trevor Adcock, Customer Service Strategy Manager (F, RM)
Wessex Trains, Dave Wallington, Head of Business Analysis (F+, RM)
Also, National Express reported to us on feedback re: NPS from:
Gatwick Express, Andrew Conroy, Head of Customer Service
'one', Andrew MacPherson, Customer Services Director
ATOC invited input from TOCs (i.e. including those not in the above sample) and e-mail responses were forwarded from:
GNER, Jennifer Hurley, Insight Manager
Virgin Trains [CrossCountry Trains Ltd, West Coast Trains Ltd], Lizzie Seear, Market Research Manager

Other Stakeholder Consultees
Association of Train Operating Companies (ATOC), David Mapp, Commercial Director (F+,RM)
Association of Train Operating Companies (ATOC), Ian Smith, Head of Commercial Strategy (F+, RM)
British Transport Police, Sara Kleinschmit, Quality of Service Research Manager (T, RM)
Continental Research, David Chilvers, Chairman (F+, RM & PB)
Continental Research, Greg Berry, I.T. Director (F+, RM & PB)
Continental Research, Tom Stacey, Field Director (F+, RM & PB)
Department for Transport (rail group consultation), Peter Lepper, Passenger Benefits Manager (F+, RM)
Department for Transport (rail group consultation), Peter West, Franchise Manager Rail Service Delivery (F+, RM)
Department for Transport (rail group consultation), David Allsop, Rail Service Delivery, North & MML (F+, RM)
Department for Transport (rail group consultation), Sheelagh Swayne, Rail Service Delivery (F+, RM)
Department for Transport (research group consultation), Jacob Wilcock, Rail Statistician (F+, RM)
Department for Transport (research group consultation), Gillian Smith, Head of Social Research (F+, RM)
Department for Transport (research group consultation), Helen Bullock, Social Researcher (F+, RM)
Greater Manchester PTE, Sally Holgate, Research and Intelligence Officer (T, PB)
London Travelwatch (London Transport Users Committee), John Cartledge, Deputy Chief Executive (T, RM)
Merseytravel, David Jones, Head of Passenger Services (T, PB)
Network Rail Infrastructure Ltd, Andrew Hutton, Commercial Manager, Stations (T, PB)
Office of Rail Regulation, Sarah Straight, Director - Rail Markets, Passengers and Freight (F, RM)
Passenger Focus, Vinita Nawathe, Policy & Research Director (F+, RM & PB)
Passenger Focus, Peter Thompson, Passenger Research Manager (F+, RM & PB)
Passenger Focus, John Mooney, Passenger Link Director - Manchester (T, PB)
Passenger Focus, Mark Woodbridge, Passenger Link Director - London (T, PB)
South Yorkshire PTE, Christine Colley-Jones, Fixed Track Manager (T, PB)
Strathclyde PTE, Hazel Martin, Head of Rail (T, PB)
Welsh Assembly Government, Gareth Price, Rail Policy Manager (T,PB)

Key to items in brackets above:
F = face-to-face individual consultation
F+ face-to-face group consultation
T = telephone consultation
PB = conducted by P Bartram,
RM = conducted by R Roberts-Miller

Appendix D – the discussion guide

The same discussion guide was used as a broad agenda both for the face-to-face and the telephone consultations. In addition, in the light of responses to questions 2 and 5 in the initial consultations, in subsequent cases it was decided to provide a "checklist" summary of the questions included in the NPS, to be administered at that point either as a self-completion questionnaire or as a series of extra questions.

DISCUSSION GUIDE

CHECKLIST

NPS Review Discussion Guide   . f2f / tel .

Name of respondent:_____

Job title:_____

Organization:_____

Date:_____   Interviewer:_____

Explain anonymity options and that they will be able to decide at end of consultation.
Request permission to record conversation for analysis purposes only:YES   NO

USAGE

1. How (do you/does your org) use the results of the NPS?
2. Which of the NPS results do you use?
3. How important are these results to you/your org? *(PROBE: essential?/important?)*
4. When you evaluate the latest NPS results, do you compare them with previous results - or do you only look at
       the latest figures?
5. Are there any results you do not look at/use at all? *(now use QUESTION CHECKLIST form)*
6. What other passenger surveys do you look at apart from the NPS?
(PROBE: Is (any of) that research which was carried out specifically for your organisation?)
 (PROBE: What does it/do they give you that the NPS does not?)

OUTPUT

7. Are you happy with the way the NPS results are reported?
CHECK ON:
7a. report format
7b. number of copies (2 per TOC)
7c. report frequency (2 pa) *(PROBE: 2pa versus 1 pa + additional diagnostic studies?)*
7d. turnaround time (speed of delivery of results)
7e. coverage (scope of topics covered)
7f. level of detail of results reported/cross-analysis categories
7g. accuracy/reliability/credibility of results generally
Ask TOCs only:
7h. what would be the best comparison figure for your TOC:
       "Long Distance" average, "Regional" average, or "London/SE" average?

METHOD

8. Do you have any comments on or criticisms of the NPS survey design or method?
CHECK ON:
8a. interview method (self-completion), response rate (about 40%) & timing (2 waves p.a.)
8b. questionnaire design
8c. sample design
8d. weighting method

OTHER COMMENTS

9. How happy are you with the NPS generally?
10. Do you have any other comments on the NPS?

Finally: *Explain anonymity options and record respondent's decision (CIRCLE ONE):*
YES - my name can appear in the report and my remarks can be attributed
        YES - my name can appear in the report but I wish my remarks to be anonymous
        NO - I do not wish my name to appear in the report.

THANK YOU FOR YOUR CO-OPERATION

Record duration:

NPS Question Checklist   Name:_____

Please indicate how much <u>you and your organisation</u> use the information from each section of the NPS questionnaire (please tick √ one box for each section):

TRAIN DETAILS	use a lot:	sometimes:	rarely/never:
1a Departure time from named station			
1b Where got off train			
1c Use of rail replacement bus/coach			
2a Did journey continue on another train			
2b Final destination station			
2c Other stations where changed trains			
3 TOC operating train from named station			

YOUR JOURNEY TODAY	use a lot:	sometimes:	rarely/never:
4 Main purpose of trip			
5 Was this outward or return journey			
6 Party composition (alone, or with adults/children)			
7 Travelling with heavy luggage/pushchair/bicycle/dog.			

DISABILITY	use a lot:	sometimes:	rarely/never:
8a Do you have a disability or long term illness			
If "yes" to 8a			
8b How satisfied with how station met needs of D or LTI			
8c How satisfied with how train met needs of D or LTI			
8d Did you make advance arrangements with TOC for travel			
9 How satisfied with arrangements made to organise travel			

TICKET PURCHASE	use a lot:	sometimes:	rarely/never:
10 When/how/where/by whom booked			
11 If advance booked: when booked exactly			
12 Rate: advance booking service/info provided/range of tickets available			
13 Type of ticket used			
14 Type of railcard (if used)			

OPINION OF STATION	use a lot:	sometimes:	rarely/never:
15 Ticket buying facilities			
15 Provision of information			
15 Upkeep of station			
15 Cleanliness			
15 Facilities			
15 Availability of staff			
15 Attitude/helpfulness of staff			
15 Connections with other public transport			
15 Car parking facilities			
15 Bicycle parking facilities			
15 Personal security			
15 Overall station environment			
16 How familiar are you with station			
17 Did you ask staff for help/info			
18 How satisfied with their handling of request			

OPINION OF ROUTE	use a lot:	sometimes:	rarely/never:
19 How satisfied with frequency on route			
19 How satisfied with punctuality on route			
19 How satisfied with scheduled journey time			
19 How satisfied with connections			
19 How satisfied with value for money			
continued overleaf . . .			

OPINION OF TRAIN	use a lot:	sometimes:	rarely/never:
20 Cleanliness			
20 Upkeep and repair			
20 Provision of information			

20   Availability of staff
20   Helpfulness/attitude of staff
20   Space for luggage
20   Toilet facilities
20   Sufficient room to sit/stand
20   Comfort of seating
20   Space for bicycles
20   Ease of getting on/off
20   Personal security on train
21   Cleanliness of train inside and outside

DELAYS	use a lot:	sometimes:	rarely/never:

22   Did you experience delay (serious/minor/none)
23   What sort of delay (late start, late arrival, cancellation)
24   Length of delay (hrs/mins)
25   How well TOC dealt with delay
26   If poorly: Why rated poorly

OVERALL OPINION OF JOURNEY	use a lot:	sometimes:	rarely/never:

27   How satisfied with journey today
28   How often do you make that journey

FREQUENT USERS OF ROUTE	use a lot:	sometimes:	rarely/never:

29   How long have you been regular user of route
30   Rate typical seat availability
31a   How satisfied with ticket office hours on route
31b   How satisfied with frequency ticket checked on route

SECURITY ON RAILWAY	use a lot:	sometimes:	rarely/never:

32   Ever worried about personal security on train journey
33   Reason for concern on station
33   Reason for concern on train
33   Reason for concern on in station vicinity
34   Regularly see Police Officers on trains
35   Ever not travelled by train because of concerns about security
36   Have had contact with BT Police

INCIDENT WITNESS OR VICTIM	use a lot:	sometimes:	rarely/never:

37   How satisfied with BT Police - contact, actions, info, treatment
38   How satisfied with BT Police overall
39   Have you considered contacting police re: incident on the railway
40   Why did you not contact police

GENERAL INFORMATION	use a lot:	sometimes:	rarely/never:

41   Potential improvements preferred re: planning journeys
42   Have you claimed compensation for delay or complained last 6 months
43   If yes: how satisfied with handling of complaint/claim
44   If dissatisfied - why
45   Further comments about trip or rail service generally

ABOUT YOU	use a lot:	sometimes:	rarely/never:

46   Age
47   Sex
48   Working status
49   Occupation of chief wage earner
50   Use of internet
51   Ethnic group

end/

Appendix E – the documents reviewed

Documents reviewed in the course of this project included (but were not limited to) the following:

Examples of standard NPS Reports:
NPS Wave 12 - Spring 2005, Consultees Report, RPC/CR, 2005.
NPS Wave 12 - Spring 2005, TOC Report (example), RPC/CR 2005.
NPS Wave 13 - Autumn 2005, PTE Report (example), RPC/CR 2005.

Examples of NPS questionnaires and fieldwork documentation:
NPS Fieldwork Report Wave 12, CR, Spring 2005.
NPS Questionnaires Waves 10, 11, 12, 13, CR, 2004 & 2005.
NPS Wave 13 Questionnaire Distribution Instructions, CR, 2005.
NPS Wave 13 Respondent Record Form, CR, 2005.

Other relevant documentation:
Contract Relating to the Procurement of NPS, SRA, December 2002.
NPS Final Sampling Plan Wave 11, CR, 2004.
NPS Final Sampling Plan Wave 12, CR, 2005.
NPS Final Sampling Plan Wave 13, CR, 2005.
Frequently Asked Questions - The NPS, RPC, August 2005.
Future of Passenger Representation, Stakeholder Briefing Note, RPC, 2005.
Invitation to Tender, NPS 2003-2007, SRA October 2002.
Multivariate Report Wave 12 Summary, CR 2005.
National Rail Travel Survey, Background Document, DfT, 2005.
National Rail Trends Yearbook 2004-5, SRA, June 2005.
NPS - A Brief Guide, RPC, August 2005.
NPS 2003-2007 Market Research Proposal, CR, November 2002.
NPS Overview Document, CR, Autumn 2004.
NPS Overview Document, TORA, August 2002.
NPS Qualitative Research, CR, 2003.
Passengers Expectations and Priorities for Improvements, CR for SRA, 2005.
Review of DfT and Agencies' requirements for Attitudinal Evidence - phase 2 (draft), DfT, June 2005.
SRA NPS Sample design – Wave 9 – Autumn 2003, CR, 2003.

In addition, analysis was carried out on the following:
LATS Rail Survey Counts, Tranches 1 - 6, Provisional Data, DfT.
NPS Wave 11 database, CR, 2004.
NPS Wave 12 database, CR, 2005.
NPS Wave 13 database, CR, 2005.
Rail Planner 2005, Travel InfoSystems.

KEY:

CR = Continental Research
LATS = London Area Travel Survey
NPS = National Passenger Survey
RPC = Rail Passengers Council (renamed "Passenger Focus" in 2006).
SRA = Strategic Rail Authority
TORA = The Oxford Research Agency

Appendix F – the NPS questionnaire

Wave 13 (Autumn 2005) Standard Version (12 white pages)

Wave 13 (Autumn 2005) Central Trains Extended Version (20 pink pages)