

NATIONAL PASSENGER SURVEY

Report on a Research Review

December 2010

Prepared for: Passenger Focus
One Drummond Gate
London
SW1V 2QY



Prepared by: Roberts-Miller Associates
PO Box 104
Dorking
Surrey
RH5 6YN
Tel: 01306-741368
E-mail: mail@roberts-miller.co.uk



Peter Bartram, FMRS
Paul Harris, CStat, FMRS
Richard Roberts-Miller, FMRS

17th December 2010

Contents	Page
EXECUTIVE SUMMARY	3
BACKGROUND & OBJECTIVES	7
COMMENTARY	8
1 Study of Documentation	8
2 The Stakeholder Review	9
2.1 Our Approach	9
2.2 How the NPS is used and its impact on standards	10
2.3 Which parts of the NPS are used most and which least	12
2.4 Perceived limitations of the NPS	15
2.5 Comments on NPS questionnaire content	18
2.6 Comments on NPS frequency and seasonality	21
2.7 Comments on NPS design and method	24
2.8 Comments on NPS accuracy, reliability and credibility	29
2.9 Comments on NPS report format, content and timing	31
2.10 Comments on NPS cross-analysis	35
2.11 The NPS going forward	38
2.12 Overall Assessment of the NPS	41
3 The Technical Review	43
3.1 NPS Changes Since 2006	43
3.2 Sampling & Weighting Rationale	45
3.3 Data Sources for Journey Volumes	48
3.4 Sample Sizes Per TOC	50
3.5 Design Effects & Accuracy	54
3.6 Spread of Shifts Across Days & Hours	55
3.7 Balance of Outward and Return Journeys	57
3.8 Building Blocks	60
3.9 Future Weighting	61
3.10 Response Rates & Response Bias	63
4 Conclusions & Recommendations	65
4.1 Gold Standard Uses, Key Objectives, Limitations	65
4.2 Self-Completion Method & Response Analysis	67
4.3 Distribution of Questionnaires	68
4.4 Sampling Plan	69
4.5 Weighting	70
4.6 Technical Documentation	73
4.7 NPS Reporting and Analysis	74
4.8 Changes to the Questionnaire	76
4.9 Other Topics	80
4.10 Summary of Recommendations	81
APPENDICES	
A. The Project Team	83
B. The Review Method	87
C. The Consultees	89
D. The Discussion Guide	90
E. The Documents Reviewed	91

EXECUTIVE SUMMARY

THE REVIEW APPROACH

1. This NPS Review included a detailed technical evaluation of key aspects of the NPS method and design, and a consultation stage in which 44 NPS users representing 30 stakeholder organisations (21 TOCs and 9 others) gave us their views on the NPS.

TECHNICAL HEADLINES

2. Our technical assessment of the NPS is that it employs a highly cost-effective approach to achieving its main goal, which is to provide a consistent measure of key satisfaction parameters at the TOC level and the national level.
3. Survey design and reporting have improved since the 2005 review. While we are suggesting further incremental refinements to detailed aspects of the methodology, we find no major flaws in the current approach.
4. We must however mention at this point that response rates have been declining gradually: this must be monitored more closely and steps must be taken to reverse the trend.

STAKEHOLDERS' VIEWS

5. NPS is widely regarded by stakeholders as very important if not essential. It is the only measure of passenger satisfaction with rail that offers consistent comparisons between TOCs and over time. Typically, NPS results are very closely monitored by TOCs, and are used to prioritise where improvement is needed, and to indicate 'best practice' targets.

6. Most consultees are confident that the NPS helps to drive up industry standards (whether or not franchise targets are set). The fact that the NPS is widely published contributes strongly to this effect.
7. Typically, stakeholders regard the NPS as providing a unique benchmark at the national level, and a reliable, independent source of passenger satisfaction data at the TOC level (changes in NPS ratings for individual TOCs were said usually to be consistent with their own data).
8. Recent enhancements to the sampling design are perceived to have increased its reliability at the sub-TOC "building block" level.

USAGE & SAMPLE SIZE

9. We judge that the NPS is used more widely and more intensively than was the case five years ago. This is pushing the NPS to the limits of the capacity of its design and its sample size.
10. Such is the acceptance of the NPS data, that many TOCs (and some others) would ideally like the NPS to provide data at the route and even individual station levels, but this would require NPS sample size to be very much bigger, e.g. x4 or x8.
11. By far the most frequent request regarding the future of the NPS was that the sample size should be much larger.
12. Many TOCs conduct their own research to supplement the local detail and/or range of data and/or frequency they get from the NPS.
13. The scale, scope and methods of the research conducted independently by TOCs vary very widely according to the structure and priorities of their networks, but in virtually all cases their research is conducted only in their own areas.

REPORTS AND ANALYSIS

14. Comments on the NPS report formats and contents were varied. Some users are happy with the headline "glossy" report, while others go straight to the data (e.g. the TOC report tables, or the Reportal, or a database disk). Some users prepare a report that is tailored much more closely to their own organisation's priorities.
15. The description of the NPS methodology included in the current Overview Report is much more complete than was available in 2006. We have suggested a few further improvements that could be helpful.
16. Extra analysis is generally valued highly, whether from special runs via Passenger Focus, or from a copy of the database, or via Reportal (although some Reportal users would like it to be more user friendly).
17. The NPS presentations to users, given by Passenger Focus and Continental Research, are also valued highly. Anything that increases understanding of the scope of the NPS (and its limits) such as these presentations and the Passenger Focus special reports, appear to be of value in helping users to get more out of the survey.
18. The NPS contains a huge range of information, which we believe could be exploited to an even greater extent, through production of ad hoc reports on specific topics, which could be based (at low cost) on extra analysis.

QUESTIONNAIRE

19. The questionnaire is widely regarded as comprehensive. The regular questions should continue largely unchanged.

Some additions were suggested, but very few had widespread support. Many regard the questionnaire as already quite long and complex enough (and we agree).

However, we are suggesting a number of places where we believe it may be possible to make small changes or improvements to question wording.

20. Any significant additions would tend to affect the response rate adversely. In fact we recommend that the questionnaire is shortened. This will be difficult, as there are few regular questions which can be eliminated without impairing the survey's utility for some users. However, the ad hoc extra questions could perhaps be limited.
21. We suggest that a follow-up survey facility is established that would allow special subjects to be researched economically without adding long extra sections to the NPS (in Wave 23 it occupies two whole pages). This could be of great benefit to the NPS response rate.

OVERALL

22. The NPS is the only measure of passenger satisfaction with rail that offers consistent comparisons between TOCs and over time.

This Review has provided strong evidence of the ways in which the NPS continues to contribute 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.

BACKGROUND & OBJECTIVES

Passenger Focus (PF) is an independent public body charged by the Government to protect the interests of Britain's rail passengers (as well as England's bus passengers outside of London, coach passengers on scheduled domestic services, and tram passengers). PF is funded by the DfT but its independence is guaranteed by an Act of Parliament.

Passenger Focus places a strong emphasis on research, and is responsible *inter alia* for the National Passenger Survey (NPS). The NPS was commissioned in 1999 to measure passenger satisfaction with rail company performance on a consistent basis, so that comparisons could be made between the train operating companies (TOCs).

The first NPS wave was in Autumn 1999 and it has been run twice a year since then. The first seven waves were undertaken by The Oxford Research Agency, until the contract was offered at competitive tender in Autumn 2002. In December 2002, Continental Research was appointed to run the survey from Spring 2003 until Spring 2007. In early 2007, Continental Research was reappointed to run the survey until Spring 2010, and this was subsequently extended to Spring 2011.

Prior to re-tendering the NPS, Passenger Focus wished to carry out a thorough review of the NPS. The objectives were to identify if changes or improvements to its design and methodology were necessary in order to maintain its utility to stakeholders and to determine if the results could 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.

As part of this process, RMA (Roberts-Miller Associates) was asked to submit proposals for carrying out (i) a consultation with NPS stakeholders and (ii) a technical assessment of the survey design and methodology. RMA was selected to carry out both parts of the review.

This document reports on whole project, which was carried out in October and November 2010 and included: an initial review of NPS technical documentation and data (described in Section 1); consultations with 44 individuals representing 30 stakeholder organisations (reported in Section 2); preparation of a technical review of the NPS methodology (Section 3); and the drafting of a series of conclusions and recommendations (Section 4). The Appendices list the project team members and their relevant experience and provide details of the review method.

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

COMMENTARY

1. STUDY OF DOCUMENTATION

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 PF, CR and others, and to carry out extra analysis of the NPS and other data sources. In addition, meetings and discussions with Passenger Focus and with Continental Research took place on a number of occasions.

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), and provided input to 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 STAKEHOLDER REVIEW

2.1 Our Approach

During October and November 2010, RMA (Roberts-Miller Associates) conducted consultations with representatives of 30 NPS stakeholder organisations: 21 TOCs and 9 other organisations. The list of organisations, and contact names at each, had been provided by Passenger Focus. In most cases our consultation was with the named contacts but in some cases the organisations suggested that different (or additional) people should participate. Most consultations were with individuals but some were with groups of two to five. All were assured that their comments would remain anonymous. 14 consultations were conducted face-to-face and 15 were by telephone. In one case an appointment could not be arranged, and their response was emailed to us. In total, feedback was received from 44 individuals at these 30 organisations:

TOCs

Arriva Trains Wales
ATOC
c2c
Chiltern Railways
CrossCountry
East Coast
East Midlands
First Capital Connect
First Great Western
First ScotRail Ltd
First Transpennine Express
Grand Central
London Midland
LOROL
Merseyrail
Northern Rail

NXEA

South West Trains
Southeastern
Southern
Virgin Trains

Other Stakeholders

British Transport Police
CDL
Department for Transport
GMPTE
Network Rail
Office of Rail Regulation
Transport for London
Transport Scotland
Welsh Assembly Government

In addition, these 24 organisations were invited to provide written feedback:

TOCs

First Hull Trains
Heathrow Connect
Heathrow Express
Wrexham and Shropshire

Merseytravel

Mott Macdonald
Napier University
National Audit Office
Nexus

pteg

Rail Safety & Standards Board
South West Observatory
South Yorkshire PTE
Steer Davies Gleave
Strathclyde Passenger Transport
University of Nottingham
University of West of England
West Yorkshire PTE

Other Stakeholders

ARUP
Centro
LEK Consulting
Line by Line Ltd
London Travelwatch
London Underground Ltd

For full details of the method see Appendix B. For a detailed list of consultees see Appendix C. The following pages report on the Stakeholder Review. Selections from feedback received during the consultations have been included as anonymous illustrative quotations (identified as from a "TOC" or "Other" stakeholder).

2.2 How the NPS is used and its impact on standards

We found that NPS users vary in the range of ways they use the data. National and regional bodies have differing perspectives; specialist users have a greater interest in certain questions than in others. However, the NPS is clearly used widely by those consulted as an important source of trend data and of benchmark data on passenger satisfaction with rail services. TOCs, in particular, study the results very closely, whether or not they have any contractual NPS targets.

"We use it to measure our performance, mainly against competitors . . . and also to monitor ourselves against franchise commitments." (TOC)

"It's the main comparison tool that we have. Also it helps identify issues that we need to do things about to improve the service." (TOC)

"We use it for an internal improvement programme, to help improve across the 31 factors." (TOC)

"We use it as a high level indication of customer satisfaction, and we use it to benchmark ourselves against the industry and against similar types of TOC . . . Before the NPS there was no national scoring . . . that's been the big advantage, the comparability across the country." (TOC)

"We use the data as part of our business planning process, in the formulation of our service plan . . . Then with each wave we check to see if there are any new issues emerging that need us to amend actions within the plan. We also look at our performance vis-à-vis others in the industry to see whether we are performing ahead of the curve or if there are others performing better from whom we can seek best practice." (TOC)

"A lot of work is put into trying to get the scores up, trying to meet the targets. It certainly has a value for us as a yardstick of how we are doing." (TOC)

"We review where we are against benchmarks, against other TOCs with similar types of operations . . . we look at best practice, what other TOCs are achieving." (TOC)

"We see NPS as very important . . . it benchmarks us against other operators . . . We wait with eager anticipation to see the results." (TOC)

Other users may vary in how they use the data, but typically regard it as valuable.

"It's a key source of a lot of things about passengers' behaviour and attitudes, with a good degree of drill down so we can actually examine a bit of the network or the element that we're looking for." (Other)

"It's always been very visible and important to us . . . it is an alert for where we may want to take further action or do further analysis or gain further insight." (Other)

Most NPS users, both TOCs and the others, regard the NPS as a factor in driving improvements in standards and in customer satisfaction.

"Because it's a national survey, because it's independent, because it's published as widely as it is, it is something that is very much a driving factor and something TOCs focus on in improving their results." (TOC)

"I think it is given a lot of weight within the industry. Stakeholders place a lot of emphasis on it, and as a result, operators want to see that their scores are going in the right direction." (TOC)

"Absolutely, yes. It's a very visible measure, the press reports on the figures, so it's in our interests to get good publicity from getting it right, rather than bad publicity from getting it wrong." (TOC)

"We use it for action planning that delivers customer improvements . . . This company pays a lot more attention to the NPS, and has done increasingly over the years. Let's face it, who wants to be at the bottom of the NPS table?" (TOC)

"We live for the day that it comes, because that's what decides our next course of action." (TOC)

"Yes. There's a lot of work going on, a lot of the activities are target led. The results do get to the people on the ground." (Other)

"Very much so. Because you can look at it and say 'we are failing in this area' so it's a good source of information on what you need to target . . . also, by publishing a league table there's always that peer group pressure, and it works." (Other)

"Absolutely, definitely . . . many managing directors live or die by it. I know one or two have contrasting views about the NPS, but on the whole many find it very valuable . . . once you've got a baseline - if everyone's going up and you're going down, there's a hell of a lot of pressure, whether it's in the contract or not." (Other)

"It's independent. And the same question is used across all the TOCs. And because the results are all out there then they can't hide." (Other)

"Yes it is very useful for driving up standards and it should continue to be used for doing so. The NPS is the best industry-wide measure of customer satisfaction. It is very useful for having a consistent measure across the TOCs. And it's useful for looking at trends over time." (Other)

"Yes. Through Passenger Focus PR, and by the comparisons - the reputation effects - and it's much more effective than fines." (Other)

2.3 Which parts of the NPS are used most and which least

Most consultees regard the NPS (as a whole) as very useful.

"I think it's very useful . . . within our own company the NPS is seen as a key KPI . . . at the director level." (TOC)

"It is valuable. Clearly it is pretty much the only measure where we can compare and contrast within the industry, but also compare performance of this business over time." (TOC)

"It is useful, absolutely, because it drives change." (TOC)

"Most definitely useful, you can look over the fence and see if someone who is doing something that you're not doing, and perhaps we could do a bit better." (TOC)

"Very useful. This is our major consultation with passengers. This is the one big representative survey with passengers. The NPS is crucial to us." (Other)

"For some things, NPS is pretty much the only measure we have." (Other)

"NPS is the best industry-wide measure of customer satisfaction." (Other)

When asked which of the NPS results were the most useful or valuable, detailed responses covered a fairly wide range.

"It's the overall satisfaction and value for money questions that get the most discussion." (TOC)

"I try to look at the longer term trends . . . you get a more realistic view of whether you are heading in the right direction or not." (TOC)

"The best part of the NPS is being able to compare it to like services in other TOCs." (TOC)

"The top three of interest to us are overall satisfaction, value for money and punctuality/reliability." (TOC)

"The ones we tend to concentrate on are punctuality and reliability, cleanliness of trains, cleanliness of stations, attitude and helpfulness of staff, staff availability, ticket buying facilities - and how we deal with delays." (TOC)

"There is a lot . . . security . . . rolling stock . . . attitude and helpfulness of staff . . . overall station environment." (Other)

"We use NPS to track progress on company initiatives such as . . . safety and security on trains and stations, on board initiatives. We also use it to track improvements in relation to disruption and how information is handled." (TOC)

"Overall satisfaction, and the ones on performance, punctuality, and value for money. Plus there's passenger information." (Other)

However, a large proportion of consultees said that most or even all of the NPS results were used and were regarded as valuable.

"Any attribute where we seem to be performing poorly, or that's shown a worrying trend, becomes important simply because we need to address it." (TOC)

"The headline score gives us a basic feel of overall benchmarking but it's the detailed scores that really help us understand areas to focus on and areas where there's weakness." (TOC)

"I tend to look at the high level stuff . . . but other people in the company will delve into the detailed elements they are more interested in." (TOC)

"It's all valuable really." (TOC)

"All of them really." (TOC)

"We use pretty much every element of it. The high level results I tend to keep an eye on because it gives me an indication of high level market trends, behavioural trends - having those high level results is incredibly useful . . . also, if we are looking at something specific, we can drill down." (Other)

"Probably pretty much all of them. Overall satisfaction is probably the headline one, but we look at all the others as well . . . We particularly focus on areas where we are trying to deliver improvements." (Other)

"It's all valuable. We find it extremely useful." (Other)

When asked if there were any NPS results they never look at, a similar view was widely expressed.

"No, not really." (TOC)

"No I don't think so." (TOC)

"No. Nothing at all. we use it all . . . every single question." (TOC)

"Actually I look at all of them." (TOC)

"We look at all of it." (Other)

"We literally look at all parts of it." (Other)

"No. I think frankly it is all useful." (Other)

However, in some situations the NPS is measuring elements of the journey experience which do not exist universally and/or which a TOC may not control.

Examples are: car parking (not all stations have car parks) other station facilities (as some TOCs manage few of the stations they use, if any), on board staff (some services are driver-only), and on-board luggage storage, catering and toilets (as local services may have none of these). In such cases these results may be ignored.

"Space for bikes is not really relevant to us." (TOC)

"Some we can't do much about. Connections with other forms of public transport: it's hard to know what you are measuring there . . . Ease of getting on and off: that's another one where it's difficult to know what we can do about that.¹ We try to focus our attention on those questions where we can take actions to make the figures change." (TOC)

"Station parking: a lot of the parking is not under the TOC's control . . . local authorities often provide parking." (Other)

It was also suggested that Wave 23 Q16: "How familiar are you with (this station)?" was of doubtful utility.

In cases where standard NPS questions are applied to elements of the journey experience which are not universal, this may result in some disconcerting replies.

"People at Charing Cross are happy with the car park!" (Other)

"On train toilets: 20% of passengers (*on trains with no toilets on board*) say they are satisfied with them. Quite what they have been doing God alone knows!" (Other)

Such anomalies could possibly be eliminated by minor changes to question wording and editing. However, this does raise a more general point, which affects any NPS questions relating to topics affecting only a proportion of journeys.² They can, if the wording and filters are not clear and effective, provide the respondent (and provide the NPS user) with unnecessary problems of interpretation; also, in any event, the relevant results will often be based on only a very small proportion of the NPS sample, and thus be liable to wider margins of error.

¹ Another consultee commented that it is very difficult to interpret the responses to this question, as dissatisfaction could arise from any of a variety of factors.

² E.g. asking about on-board catering or toilets (rare on commuter trains), or station catering, toilets or ticket sale facilities (rare at small stations), or handling of delays on a specific journey (most of which will not be affected by a delay).

2.4 Perceived limitations of the NPS

We asked consultees if they thought anything was missing from the questionnaire. Some consultees suggested adding extra questions (details of which will be covered in the Section 2.5), but none of the individual suggestions were mentioned by more than a few consultees. A more typical response was that NPS topic coverage is comprehensive, or that the questionnaire is long enough already.

"I don't think anything's missing. It's quite comprehensive." (TOC)

"Not that I'm aware of." (TOC)

"No. I think it's pretty comprehensive." (TOC)

"None that I can think of." (TOC)

"No I don't think so." (TOC)

"I think NPS pretty much covers it." (TOC)

"I think everything is there." (Other)

"I can't recall any . . . you can't go adding and adding." (Other)

"It's a big questionnaire already." (Other)

We asked also what they would like a future NPS to allow them to do, that is not possible now. Again there were several suggestions (which will be covered in the following sections) but once again, each was typically mentioned by only a few consultees.

The sole exception to this related to sample size. Many NPS users, probably most of them, would like the sample size to be increased substantially.

"You need a bigger sample." (Other)

"I think it is at the lowest limit of sample size . . . it's accurate overall, but . . . I want to drill down geographically, but to do that properly would require a huge amount of extra investment by Passenger Focus." (Other)

"We would be quite keen to get a better spread of the sample . . . to be able to break it down by operator." (Other)

This was a particularly widespread view among the TOCs.

"The sample is pretty small. Ideally I would like a bigger sample. Ideally I would like to have access to enough data to be able to drill down." (TOC)

"Sample size is a bit small. As we break it down further . . . you're only getting 20 or 30 interviews. It would be nice if they could go a bit further." (TOC)

"Why can't they increase the number they survey?" (TOC)

"It could probably benefit from a larger sample size." (TOC)

"The weaknesses we see are sample size, which currently is too low to allow meaningful segmentation." (TOC)

"When you go down to station level the sample size very quickly becomes too small to be of any practical use . . . we would like to see a greater sample size to give greater granularity in the figures." (TOC)

"When it's such a tiny number of customers telling you something it's quite hard to push a significant action through the business, so if you had a larger number of people telling you the same it would carry more weight and have more credibility. A bigger sample size would help us enormously." (TOC)

"I would prefer greater numbers of interviews . . . for some stations it's a very small number of people." (TOC)

"Sample size . . . (*at some small stations*) you get a just handful of questionnaires back. I fail to see how that can give me a balanced view of what customers think." (TOC)

"One of the weaknesses of the NPS is the limited number of people it will survey. When you get down to the smaller stations it may just be two people." (TOC)

The NPS was designed to produce reliable results at the TOC level (and of course nationally). The recent enhancement to the sample building block structure has improved the reliability of the NPS data at a large sector level (e.g. splitting a large TOC into 3 or 4 parts), but only a substantial increase in NPS sample size and budget (e.g. x4 or x8) would be able to produce significantly improved data below that level.

Evidently many TOCs regard the type of TOC-level information they get from the NPS as useful, to the extent that they wish to be able to obtain the same information at route or station level. Therefore, several TOCs conduct regular surveys of their own. In some cases they ask additional questions, or increase survey frequency, but a key objective in most cases was to be able to drill down to detailed local level.

"We want to see how stations are performing . . . We survey a lot more people, it covers more of our stations." (TOC)

"We supplement the NPS with additional research which allows us to go down into greater detail." (TOC)

"We already do our own survey which is very similar to the NPS but goes into more detail." (TOC)

"It has essentially the same sets of measures as NPS but it gives us more detail." (TOC)

One thing that the TOCs' own surveys do not provide is comparative data on other TOCs, as their surveys are conducted exclusively within their respective networks.

Other than the NPS, and the private surveys conducted by the TOCs, few other sources of rail passenger data were mentioned - apart from the NTS and operational statistics (which measure behaviour or performance, but not satisfaction).

2.5 Comments on NPS questionnaire content

As mentioned in Section 2.4, many consultees believe that the NPS questionnaire already measures all the key aspects of the journey experience.

At the same time, there were various suggestions, each from small numbers of consultees, regarding possible additions or changes.

Some of these related to a general issue already mentioned in Section 2.3 regarding the impact of stations not managed by a TOC, on a TOC's overall score.

"I question why someone, who is not at a station we manage, whether their opinion of the station should count towards our score." (TOC)

"Whatever goes on at the stations . . . we have little or no influence over. Each TOC has got some stations, or most TOCs have got some. Many of the big stations are still with Network Rail. Some of the small TOCs have no stations at all, but they still get a station score. I am not sure I understand the value of that." (TOC)

Given that there is now in fact a "station overall" score (i.e. ignoring the train), it seems reasonable that there should also be a "train overall" score (i.e. ignoring the station), as well as the existing "journey overall" score.

"Might be useful . . . to have an overall train average." (TOC)

"There's a station overall question: 'do you like the overall station environment?' but there's no parallel for the train overall." (TOC)

Among all the other suggestions, the largest group related to expanding the section on ticket purchase.

"There's a question on station ticket buying facilities, but there's no similar question that takes account of ticket buying done on line or by phone (which can be half of them in some cases). If you didn't buy your ticket at the station then where are you going to say whether you were happy with the ticket buying facility? I don't think the NPS really reflects that." (TOC)

"On the retailing question on how satisfied they were with the ease of buying ticket we don't know how they bought a ticket - did they get it from Trainline on a website, or at a station, or on board?" (TOC)

"It would be useful for me if we understood more about how a passenger plans their journey . . . for example if they had used the website. We talk about ticket purchase which is meant to refer to station ticket offices when increasingly more and more people are doing it on-line." (TOC)

"I would like it to cover more things about before you make that journey, how you found information, website-related, ticketing." (TOC)

"Ticket buying facilities: (*questions on*) ticket offices, on-line ticket sales, ticket vending machines." (TOC)

"The NPS talks about ease of purchase of ticket, which assumes that most tickets are purchased at the station, but . . . technology is changing." (Other)

"More about information services used before and during the journey . . . including use of a mobile phone for real time information." (Other)

"It's a bit pre-web in a way, re: web ticket procedure." (Other)

[It should be noted here that it was suggested that Wave 23 Q60 (re: respondent's use of the internet) has little apparent utility unless further details are provided.]

Another suggestion was that non-users' views would be interesting (but this could not be covered by the NPS, as the NPS interviews only passengers).

"It would be useful to have a similar survey to be asked of non-passengers. This is a survey that tells us what passengers think. It would be helpful to know what non-passengers think. What factors are stopping people using the train service?" (TOC)

"The only thing the NPS misses is those people who don't travel, and to know why they don't travel, but that wouldn't be possible." (TOC)

Other suggestions, each from just one or two people, covered a very wide range:

- Record respondent's home postcode
- Record train headcode
- More details for long distance TOCs (catering standards, wi-fi provision).
- On-board temperature (i.e. "Is it too cold? or is it too warm?")
- On-board toilets: "Was there one? Did you use it?"
- More details about customer services, enquiries or complaints handling.
- Frequency of on board ticket checking - "Did you have your ticket checked?"
- More information about the onward journey.
- Measure whether the "customer is happy using our website".
- Add a question on "net promoter score" or advocacy measure.
- Ownership of types of smart phones (iPhone, Blackberry, Android).

There was one suggestion that a lot more detail was needed on station food, drink and shopping facilities (but this would be relevant only to large stations). Another suggestion was that some important station parameters had not yet been identified.

"I think there are some questions missing from the station element but I don't know what they are . . . When we do our multivariate regression analysis or key drivers work on station satisfaction . . . it suggests that the questions that are there aren't describing overall (*station*) satisfaction . . . there are some nuggets that we are not getting . . . perhaps the look of the station approach . . . or the track?" (TOC)

There were a couple of suggestions to add questions that are already present on the NPS, e.g. one suggested adding "was there a delay?" (see Wave 22 Q28).

One criticised that very section of the questionnaire.

"I would like to see the 'information about disruption' questions changed as I feel they do not actually give us a great deal. For those TOCs who experience very little disruption . . . a very low sample size is asked the questions." (TOC)

Another wanted to look at what could soften the impact of price increases.

"Ticket price increases - is there any work going on to try to understand which attributes on the NPS would help support that cost increase - is there anything else that would give the impression of value for money that could offset the price increase e.g. longer trains, platforms, new trains?" (TOC)

One consultee argued that additional follow-up questions are needed on every parameter covered before the NPS data can be regarded as useful or actionable.

"It's not actionable data . . . if you get a low score the survey doesn't tell you why . . . so I have to take action fairly blind, or I have to commission more research myself . . . If they want to use it as a tool to drive improvements rather than as a tool for the DfT to beat the TOCs with . . . it should be useful . . . what could be useful is actionable data to really drive improvements . . . it would be better served by having deep dive questions, so you would ask what they think of (e.g.) toilets and then if anybody scored less than 'x', you would ask a follow up . . . the fact that it's not used for that suggests to me that honestly it's a degree of lip service rather than using it to drive customer experience." (TOC)

On the other hand, there are limits to what can be achieved.

"You don't want to make the survey so unwieldy that people don't want to fill it in." (TOC)

Also, should Passenger Focus even attempt it?

"It's not Passenger Focus's job to be doing the TOCs' job for them." (TOC)

Overall, there were relatively few criticisms of the current NPS questionnaire.

2.6 Comments on NPS frequency and seasonality

Most seemed happy with the NPS survey frequency of two waves per annum.

"Spring/autumn? That's fine." (TOC)

"I think it's a reasonable compromise." (TOC)

"Twice a year is OK, in fact it's fine." (TOC)

"We do ours quarterly . . . but twice a year for the NPS is fine." (TOC)

"Twice a year is about right. It's fine." (Other)

"Two a year is OK." (Other)

Many would also be happy if frequency increased, but few argued for it.

"Obviously the business would like it to be more frequent." (TOC)

"We'd be happy to see it increased to four p.a. to make it easy for us to look at the trends." (TOC)

"It would be nice if it was more continuous but the cost of doing that would probably be prohibitive." (TOC)

"Half a year is too long a gap, we'd like it quarterly." (TOC)

"Continuous would be an advantage." (Other)

Few would like to see frequency reduced.

"I wouldn't be in favour of that being changed." (TOC)

"You would lose some of the comparability." (TOC)

"If you did have one a year, using the NPS as a means of DfT battering the TOCs for poor performance would be a little unreal, a little unfair." (TOC)

"Doing it yearly . . . I think most of the TOCs would complain." (Other)

Some thought the current seasonal timing was acceptable.

"They've chosen those times to avoid Easter and peak holiday times . . . the times actually work quite well." (TOC)

"Good idea to avoid the summer. The system they have built is sensible." (Other)

But not everyone thought avoiding the summer was a good idea.

"By missing out the summer, those TOCs that have seasonal flows, you are missing out on a source of satisfaction and dissatisfaction - there are lots of these people . . . it's an omission." (TOC)

"You could do it in June and July, couldn't you . . . The industry always performs best just after the late spring." (TOC)

However, the most frequent complaint was that the current NPS waves are conducted at times of year when rail performance is challenged by adverse conditions.

"You've got the autumn survey carried out in leaf fall, and the 'spring' survey carried out in the depths of the winter. I'm not sure that's fair to the industry, is it?" (TOC)

"It's almost as if you picked the times when the rail industry will traditionally underperform - with January/February/March being the darkness of winter and the later wave being affected by the leaf fall season . . . this spring the customer responses were affected by the impact of the very cold snap, the severe weather, the snow . . . which will have made trains dirty, late, cancelled . . . I think the NPS does actually reflect what happens but when it's given to the media it doesn't explain why, they just give the bald statistics." (TOC)

"The spring wave kicks in a little early." (TOC)

"The spring survey is a concern, because it's not spring is it?" (TOC)

"It would be better if they were better spaced." (TOC)

"The surveys are not done throughout the year. They are done in two waves which are close together. You get an autumn one, and then just after that one allegedly in spring but it starts in January. They do tend to be lower scoring. The autumn wave can be skewed because the onset of leaf fall can vary by six to eight weeks. The spring one depends on exactly when the snow occurs. Yet we don't do anything from Easter to September. We don't actually measure anything in the summer when performance is much better. So the results can be misleading. I think the autumn wave should start earlier, say July. Then we would have a winter and summer one (with the existing 'spring' one being winter, and the current autumn one moved forward to be summer). If you can't do that (because of consistency or contract targets) you should have a third wave in the summer." (Other)

"The bad point from our point of view is that the spring samples are in January to March, that is probably the time we are going to get the worst of the weather. I am surprised they call it 'spring'." (Other)

Not everyone was as concerned about the seasonal timing: some were more concerned that consistency should be maintained and that like should be compared with like.

"It provides a consistency - and I wouldn't be in favour of that being changed." (TOC)

"It isn't an issue - because you compare autumn with autumn and spring with spring and that's perfectly robust . . . but it does feel strange." (TOC)

"Always compare autumn with autumn and spring with spring." (Other)

2.7 Comments on NPS design and method

When we invited comments on the NPS survey method and design generally, there were only a few spontaneous remarks.

"It's fit for purpose . . . for national survey purposes it's fine." (TOC)

"It's the best methodological approach." (TOC)

"Sample size could be increased."(TOC)

"I think it's reasonably cost-efficient." (Other)

"It's very good. It's a very professionally run survey, and generally I've got a lot of respect for Passenger Focus. I think they do a very good job." (Other)

Sample Size

There was more feedback when we asked specifically about sample size.

"A bigger sample size would help us enormously." (TOC)

"I would prefer greater numbers of interviews . . . for some stations it's a very small number of people." (TOC)

"When we look at parts of our area we have to be very, very cautious about the results and what they are telling us. Also, on satisfaction with stations, if that has dropped or improved . . . we want to try and look at which stations, and again sample size precludes us from being able to draw those kind of conclusions." (TOC)

"We would like to have better coverage of our stations." (TOC)

"(At small rural stations) you probably get 3 stations out of 50 surveyed and you get a just handful of questionnaires back for each. I fail to see how that can give me a balanced view of what customers think." (TOC)

"The actual sample structure is fine. It covers a reasonable cross-section of our stations, and they are the right stations, we've looked at that and it's fine . . . but we want a bigger sample and at more locations." (TOC)

"When it's such a tiny number of customers telling you something it's quite hard to push a significant action through the business, so if you had a larger number of people telling you the same it would carry more weight and have more credibility." (TOC)

"If you are going to cover operators within an area you need a bigger sample." (Other)

Sample Design

There were very few general comments on sample design.

"(The sample is) updated regularly enough to make it sensible." (TOC)

"They get a good mix of peak & off-peak and weekend & weekday." (TOC)

"It seems functional, it doesn't misrepresent anything in my view, it's just the sample size." (TOC)

"It's very complex. Is there a simpler, more open way to do it?" (TOC)

There were these three suggestions that the sample design should be altered to provide better data at the medium or small station level, or at route level.³

"We find the sample seems to be too centred on the main stations: it ought to be . . . putting more into the feeder stations . . . Most people are doing two journeys in the day, one from the local station into London, and one from London out again, which kind of suggests that the overall sample ought to be about the same (*each way*) . . . I think the samples . . . need to be farmed out to the feeder stations to a certain extent. " (TOC)

"It can cause a skew in the data depending on which stations are selected. Across the network as a whole it's fine, but when you start looking at one route versus another, there's inevitably a lot more volatility on a route that has a lot of small stations on it. I think a line-of-route consideration of the type of service people are on, rather than station within train operating company, needs some consideration. Performance is the biggest single driver of passenger satisfaction. One line-of-route might have poor punctuality, and another might have much better. When you are looking at large and small stations within a TOC, you need to make sure we are describing the right levels within the TOC." (Other)

"Big stations are already statistically significant. If we cut numbers at big stations and moved interviews to smaller stations, you would have more sampling points. It would be worth looking at . . . but bottom line is it's not small stations, it's the ones in the middle which are missing which are posing the biggest questions on investments, it's the commuter stations that have large numbers of originating passengers but are not big stations in their own right . . . it's the middle stations that are in need of investment . . . the ones that caused the most dissatisfaction were the middle tier." (Other)

³ The NPS was designed to provide data at TOC level, not route or station level. The recent development of 'building blocks' may have enhanced NPS sub-TOC data somewhat, but to achieve more reliable results at route or station level generally, or better coverage of smaller stations, would require a very substantial increase in sample size.

Building Blocks

We asked about the one aspect of the current sample design that was recently changed, i.e. replacing the 'virtual TOC' structure with one based on smaller 'building block' components, related to groups of stations and/or lines within TOCs. It was well received by many of the TOCs.

"That fits in with our network. It's good to have the breakdown. It has been a good tool for us." (TOC)

"Very, very, very helpful, yes - because I can split things down and give people realistic targets . . . I think it goes as far as it safely can, without the sample size getting too small . . . It's a good methodology. it works well. This year is the first time we have used the building blocks and I think that worked well." (TOC)

"I absolutely think you need that . . . it works perfectly for us." (TOC)

"That's a great idea." (TOC)

"Good idea. Absolutely. It's going to be useful for us." (TOC)

"The building block element is really useful. I can look at parts of (*our TOC*) and compare it with other (*similar*) operators . . . Building blocks are a good way of comparing at the moment." (TOC)

"I think that's good actually, it's the right thing to do." (TOC)

"It probably makes the figures represent more closely the breakdown of the customer base." (TOC)

"We find it gives very useful information." (TOC)

"We divided the network into blocks. They have certainly worked a lot better since we changed the methodology slightly." (TOC)

"It is good to see results by line. That helps." (Other)

However, any attempt to drill down comes up against the usual problem.

"Once you start breaking . . . TOCs down regionally . . . the sample becomes very small." (TOC)

Also, the building blocks that best suit the way that a TOC is organised may not be appropriate for all other purposes.

"The only problem with that is that I would like to be able to relate the building block satisfaction data to other data like punctuality, but because building blocks are (*built*) around stations it makes that quite hard." (Other)

A related suggestion was that all building blocks should be defined using a consistent set of rules or definitions.

Another comment suggested that the building block approach makes the sample design appear even more complicated than it did already.

"I must admit I sometimes get a little confused as to how the building blocks all fit together . . . I am not quite sure how . . . the survey's put together . . . so perhaps a bit more clarity . . . might be useful." (TOC)

Weighting

There were similar general remarks regarding the complexity of the weighting, but few substantive comments.

"It tends to be a bit confusing. It's not always clear how the weighting has affected the results." (TOC)

"They provide us with data which we have to compare with Lennon . . . the whole fares thing is so complex." (TOC)

"It was fine." (TOC)

"We are happy with the weighting." (TOC)

"I think they are making the right moves." (TOC)

"The team has been happy with it." (TOC)

"Not an issue." (Other)

Fieldwork Method

A few consultees regard face-to-face interviewing as inherently superior to the postal self-completion method used on the NPS.

"I think face-to-face is better." (Other)

However, others recognise the substantial cost advantage of self-completion, and few were unhappy with the NPS self-completion approach.

"I wouldn't have a problem with self-completion. If anything I would have a problem with changing the method." (TOC)

"Self-completion? I don't think that's a problem." (TOC)

"No problem." (TOC)

Also the NPS response rate was regarded as very satisfactory.

"For this kind of survey it's probably quite good really given the size of the questionnaire." (TOC)

"35% is very good for self-completion postal." (TOC)

"Fairly impressive actually." (TOC)

"Very impressive. Very good." (Other)

The minority who are less enthusiastic about the NPS self-completion fieldwork method would perhaps appreciate some evidence of its effectiveness.

"It might be nice to see a little bit of validation work." (Other)

2.8 Comments on NPS accuracy, reliability and credibility

Most consultees were happy with the reliability and accuracy of the NPS data.

"I've got no doubt about the accuracy of the results." (TOC)

"I don't think we've ever questioned the results at all . . . they look about right . . . we've never had cause to question any of the results." (TOC)

"I think it's about right. I've got no reason to suspect that it's not accurate. It seems very professional." (TOC)

"I don't have an issue with the accuracy and reliability of the survey, because I think it's representative of the numbers of people that have been asked, and it's measured against what we believe are the numbers of business and leisure customers that we have, so I think in terms of the accuracy of the information that you've got it's a fairly good representation of customers' views." (TOC)

When they compared the NPS results with those from their own surveys, while some individual figures might vary, the trends were usually consistent.

"We are fairly comfortable with NPS results. It matches the feedback we get. It's good to get the feedback from customers." (TOC)

"The results generally agree with what we find from our own survey so I wouldn't have any issues with that." (TOC)

"When I correlated it with other data the resulting graph is logical." (TOC)

"They are largely in sync, within 2% or 3%." (TOC)

"There are some disparities . . . but the same sort of trend." (TOC)

"Our results are similar." (TOC).

"I was asked to predict what our NPS results will be (*based on our own surveys*) and I usually get it right." (TOC)

"There's quite a reasonable relationship between our results and the NPS results. In overall terms at TOC level we would be reasonably confident." (Other)

"I think it's good. I think looking at . . . the stations they select, it's a useful tool to use . . . it is quite clear that the low scoring areas in the NPS are the low scoring areas (*on our own data*)." (Other)

"Generally the trends are similar." (Other)

As was the case at other points in the consultations, the desire for a large sample size was expressed once again in this context.

"I come back to wanting to see a larger sample to give a level of comfort. At times you see variations because so few small stations are surveyed." (TOC)

"I think I would be concerned about doing more (*building blocks*) on the current sample size." (TOC)

"Our concern would be at local level on a smaller base." (Other)

2.9 Comments on NPS report format, content and timing

There were mixed reactions to the various NPS report formats, largely reflecting the different ways different stakeholders wish to use the NPS.

The amount of data and potential cross-analysis available from the NPS is enormous, and far too much to be included in a report intended for wide public distribution. The published "headline" report was generally regarded as satisfactory in that context.

"The presentation of the NPS is simple because it needs to be understandable for journalists, it needs to be as easy as possible . . . but (*reality*) is more complicated." (TOC)

"The standard report is actually not bad at all for executive level and we can always drill into the data if we need to." (TOC)

"I don't think we have any reservations around the report." (TOC)

"They are fine. The graphs are quite clear." (TOC)

"I don't have a problem with it . . . I'm happy with the way it presents the information." (TOC)

"I skim the overall report just for the headlines. For me it's probably got more information than I normally need at that level. Only when I am looking at a particular subject would I go to a more detailed level." (Other)

"The way the charts are . . . easy to use." (Other)

"Very easy to understand." (Other)

The design was not necessarily to everyone's taste.

"I think the format could be better." (TOC)

"It's not the slickest looking report, but it does the job." (TOC)

Most were reasonably happy with the format and content. However, there were two minority schools of thought regarding the commentary. One was that the report should contain only facts - and no opinions.

"The commentary should be facts. I don't think it should be opinion based." (TOC)

Another view was that the report was too factual, and should include more "insight".

"It's very factual - it never gives you any insight - what would be good would be if you could overlay the report with verbatim comments . . . it's a very hard report, very data oriented, and it needn't be." (TOC)

"The report doesn't do very much . . . there's no insights. I am always surprised they don't say more, recommend more . . . they have one page of bullet points . . . One of the things we look for is help us understand things." (Other)

Some remarked on the attitudes taken in the commentary.

"The commentary is editorial rather than an impartial summary of the findings . . . I think if there was more of a constructive dialogue I think more could be achieved. On a personal level Passenger Focus are very constructive, it's just the political agenda." (TOC)

"The PR message will be prepared in advance." (Other)

At many TOCs, some users will scan the headline report, but then immediately dive into the detailed NPS data.

"We use it as a headline, to see out of the 31 questions how many have we improved on, but we tend to then break it down to the building blocks and look at the stations." (TOC)

Many TOCs rapidly produce their own report for internal circulation. Given that they are certain to be picking out the sections of data that relate to their TOC and its priorities, they will find it easier to extract data from an Excel spreadsheet than from a PDF.

"Everything's easier in Excel than in a PDF report . . . For getting it out across the business. Clearly there are some people in the business who want to see the whole. But just so the document can be easily dissected." (TOC)

"We tend to make limited use of that to be honest. I produce our own reports. We tend to work directly with the spreadsheets rather than taking too much out of the report." (TOC)

Attitudes to the various volumes and data formats will depend on how the individuals plan to use the data. Some users are happy with a headline report and may only rarely refer to the TOC report, consultees or overview reports.

"It's all very detailed. Probably too detailed." (TOC)

"I'm comfortable with it . . . it works for me . . . and there's more detail in the consultees volume." (TOC)

"The consultees report is a big one, that is heavy going, but the individual TOC ones are reasonably straightforward, they're pretty clear, the use of colour, use of bar charts, I've never had any problem with those." (TOC)

"I quite like to have one report that I can file away and look up when people want to look at the scores, but I do have a copy on disk. I have never had any comments from the team. I think they're OK with it." (TOC)

The frequency of publication of most NPS reports is twice a year.⁴ This is seen as satisfactory as it reflects the fieldwork timing.

Of much more interest (mainly to the TOCs) is the speed of publication.

"Provided the results come out quick enough, I think twice a year is sufficient." (TOC)

"Twice a year is fine - but it takes three months to get the results. In this day and age, I think you ought to be able to get the results in six weeks." (TOC)

"If that could be quicker that would be beneficial." (TOC)

"The only issue we have is that the results that come about around June time, it would be nice if they could come out a bit sooner." (TOC)

"We would like to see the reports quicker . . . the survey that just started in September - we won't really see the results until Christmas, if not January. That's a concern. It's quite a lengthy delay. We'd like it sooner." (TOC)

"They seem to have a very long time to get the data set. If you think that fieldwork for . . . autumn started at the beginning of September and we won't get the data until the middle of December, and it won't get published and you won't be able to use it widely in the business until at some point in January . . . I would like to see an improvement in the time between the end of fieldwork and getting the first data in." (TOC)

"There's not much gap between getting the results and the next wave starting. You get your results and you have little or no chance to respond . . . is there any way the results could come out so we could have an impact on the next wave?" (TOC)

"It takes a long time . . . you get it just before the next wave of fieldwork starts, so if you did want to adjust your action plan it doesn't give you an awful lot of time." (TOC)

"It takes so long to get the results." (Other)

"It is a long time actually." (Other)

⁴ The overview report is published annually.

The fact that the NPS is now classified as Official Statistics is perceived by some to have made the situation worse.

"It takes a long time, I understand why it does, but that is a bit of a niggle . . . it takes about three months then we're in embargo for ages, and then eventually it's released . . . it can be an issue." (TOC)

"There's a group of people, mainly the board, who are allowed to see the results, and for the rest it is embargoed for four weeks. It would be nice to get it out to the rest a bit more quickly . . . I am not sure why there's a need to embargo it for four weeks." (TOC)

Not everyone was as concerned.

"That has been getting longer . . . but we can live with that." (TOC)

"The key is to make sure that what you are putting out is 100% accurate. I guess the turn round times are pretty reasonable." (TOC)

"It's reasonable for that type of work, and the quality of presentation is very good. So it takes time to do it." (Other)

"It's not something Passenger Focus can do anything about." (Other)

"It's just something we have had to adjust to." (Other)

"No problem. A shorter time will get a more anodyne response from TOCs because they have had less time to prepare their response." (Other)

"There's a lot of whinging goes on." (Other)

2.10 Comments on NPS cross-analysis and extra-analysis

Some people thought the NPS reports contain more than enough cross-analysis.

"It's probably more than we'd ever need in the printed reports." (TOC)

"Probably no need. We'll still need to produce our own report." (TOC)

"You don't want to get too complex." (Other)

Others would have liked more detail, but tended to be fairly relaxed about it because they know they can get extra analysis through Passenger Focus or through Reportal.

"If we need further information we talk to the local Passenger Focus link." (TOC)

"We could do with it down to station level in the report, but we can get that through Reportal." (TOC)

"The printed report tells us as much as it can. It doesn't always give you the level of detail you would like, but you can go and drill down with Reportal." (TOC)

"I think that's fine because I take the raw data as well - we can look at different ways of cutting the data." (TOC)

Several consultees were spontaneously complimentary about the ease and speed of access to extra analysis through Passenger Focus.

"They are very good at turning around requests." (TOC)

"In terms of turnaround time of individual requests I've got nothing but praise." (TOC)

"Fantastic, you tell him what you need and he will do a little report for you." (TOC)

"The guys at Passenger Focus have been really, really helpful. I'm very impressed actually with that." (TOC)

"We have found Passenger Focus to have been responsive if we want to drill down in more detail." (Other)

Presentations were also mentioned as having been helpful.

"We had a presentation from Passenger Focus which we found very helpful." (TOC)

"Continental present to us and they focus on the (KPI) drivers. I find that useful." (TOC)

The Reportal facility received mixed reviews. Some were positive.

"I've always found it fairly straightforward. Like anything, it takes a couple of goes to get comfortable with it, but it seems to work fairly well." (TOC)

"The usability of the Reportal is pretty good, and its a good idea." (TOC)

"I've used the NPS portal. That's been very useful." (TOC)

"The portal's good as well. I don't use it myself but the guys that use it are perfectly happy with it . . . we get lots of requests for information and they can turn round information very quickly." (TOC)

Rather more were critical.

"I have tried numerous times and either the passwords don't work or the pages don't load . . . I found it quite user unfriendly. All you can get out of it is basically the headline data. It needs massively improving." (TOC)

"I don't find it as intuitive as it could be." (TOC)

"We had a few problems with the user password, getting it reset." (TOC)

"I would have liked to have seen a few more variables thrown in there. I would prefer to use Reportal if all the variables were in there (e.g. ticket type or 1st versus standard) but they're not." (TOC)

"I've used Reportal in the past but I found it extremely fiddly to get what we were after." (TOC)

"It's a bit clunky and slow - could be smoother." (TOC)

"I did start to use that originally, and I stopped using it - because it was annoying, because you could get some of the data, but you couldn't get it in the form you wanted it . . . so we ended up either doing what we could with what we had, or - where the requirement was sufficiently high - actually going to Passenger Focus and asking for the source data - which rather defeats the purpose of having a portal - and it shouldn't be that difficult. The database is not that large." (Other)

"We have used Reportal. The problem is it doesn't always work for us. It's a reliability issue." (Other)

"When I tried to use it I found it a bit esoteric, not very user friendly." (Other)

Some prefer to interrogate the database using SPSS or Quanvert, rather than Reportal.

"I prefer to use a separate package to look at it - I have reservations about Reportal . . . you can look at data through it but it's their understanding of what they have done, you always run the risk of people misusing it . . . I use Quanvert." (TOC)

"Continental do send out the SPSS files which contain all the variables, so I use the SPSS file quite a lot." (TOC)

Some were not aware that such database access was possible.

"I wasn't aware of that. I didn't think it was an option." (TOC)

"Do they do that?" (TOC)

Most seemed to be aware of the Verbatim facility, but not all had used it.

"Yes it's great, but it's time intensive. It's great if you've got the time to go through it." (TOC)

"You used to be able to get them written in a report. They were quite good fun." (TOC)

"I've had one look. For us easier access might be very useful." (Other)

"I know it exists." (Other)

"I haven't used that." (Other)

"I think I was aware of it . . . perhaps Passenger Focus could remind people that's there." (Other)

It was suggested that there should be more verbatims, and they should be coded.

"I think there should be more spaces for verbatim comments. I find verbatim comments useful . . . It's not included in the pack you get. You have to go and seek them." (TOC)

"If it could be grouped (*e.g. coded and analysed*) we might be able more easily to draw conclusions from that." (TOC)

Finally, one stakeholder's main criticism of NPS analysis was that the TOC-type sector breakdown (London SE v Long Distance v Regional) does not compare like with like.

"The categorisation, which TOC-type you are in. I think there must be a better way of doing it . . . The media pick up the easiest results they can see, but why should we be compared against _____⁵ ?" (TOC)

⁵ At this point the consultee named another specific TOC classified as the same TOC-type as their own, but one (it was argued) that had very different network characteristics.

2.11 The NPS Going Forward

We asked consultees how they would like to see the NPS methodology or reporting evolve in future to make it more useful or valuable, or more effective in driving up standards. Their replies suggested that many were basically satisfied with most aspects of the NPS, but would just like the sample size to be larger.

Certainly, very few suggested changes to the report format or content at this point.

"Don't really know. We are familiar with the layout of the reports. It's fairly easy to update our information on our trend charts." (TOC)

"I do think it's well published. It's well regarded." (TOC)

"Simpler reports (and then use the portal)." (TOC)

"Key driver analysis. They don't give it much prominence." (Other)

On methodology, several argued that no basic changes were necessary (or even that changes are actually inadvisable if you wish to preserve comparability of data).

"What I like most about NPS is that because it's fixed, it gives you a measure to go back over the last 8 or 9 years which is very important see how your performance has changed." (TOC)

"I have no fundamental problems with the NPS and the way it's designed . . . The NPS is a useful tool for us, and I am not sure I would want to see it changed in any dramatic way." (TOC)

"I'm not sure that I can think of a way of making it more valuable (other than increasing the base size and looking at the way some of the questions are asked). It achieves pretty much what it sets out to do . . . which I think is to provide us with a representation of . . . our customers' views on the quality and experience of using our service." (TOC)

"Methodologically there's nothing I would like to see changed." (TOC)

"There's nothing at the moment we would want to change." (TOC)

"I don't think so. I have had no adverse comments back from anyone in the business." (TOC)

"I don't think it could really. I think it's pretty good." (TOC)

"No, not really." (TOC)

"Can't think of any. No not really." (TOC)

"I think it's fully fit for purpose. There are some minor improvements you can make, but there always are in something like this." (Other)

Of all the changes that were suggested at this point, increased sample size was mentioned more frequently than anything else.

"I'd be reluctant to change the style of the survey but I'd like a bigger sample size." (TOC)

"It would be nice to have it more often, but I would rather have a bigger sample." (TOC)

"If budget wasn't a factor, then increasing sample sizes, to give us a greater depth of data, a greater depth of understanding - on a geographic level or by passenger type."(TOC)

"It's not granular enough . . . there's not enough people saying it for it to carry any weight." (TOC)

"At a national level it paints a very good picture, but it's more challenging when you drill right down." (TOC)

"In my case I'd want a bigger sample size." (TOC)

"The main thing for me is to get bigger sample sizes." (TOC)

"It is sample size - big enough so we can segment." (TOC)

"The sample size issue." (TOC)

"Address the ongoing criticisms about sample size." (Other)

Of the other detailed changes that were suggested, only a few were mentioned at this point by more than one or two individuals.

These were seasonality:

"It wouldn't do any harm to start the spring wave a little later." (TOC)

"More surveys, more survey periods. Maybe a continuous survey." (TOC)

"Making it more even through the year in the sampling." (Other)

And accessibility:

"They could probably spend a little more time telling people about it . . . this is what you can use it for." (TOC)

"Making it more interactive, making it more accessible so you can share it with your colleagues." (TOC)

"Getting access to the information, making it more readily available and easier to use." (Other)

And franchise targets:

"It needs to be there as part of franchise obligations." (TOC)

"For me it's a question of building it into franchise agreements, that type of thing, so it really does bite." (Other)

"Get it into more franchises!" (Other)

The other suggestions, each mentioned by only one or two individuals, covered a wide range.

"Sampling and weighting at small stations needs to be looked at." (TOC)

"Could improve it by focussing on (*medium size*) hub stations." (TOC)

"I would like to see more face-to-face questions." (TOC)

"I would like to see interviewing on train." (TOC)

"Could response rate be boosted?" (TOC)

"Does interviewing have to be at the station where your journey is commencing? Why can't customers coming in to the station also be surveyed?" (TOC)

"Definitely further improvement on the weightings in the evenings . . . Evenings and late evenings are always a bit light compared to the number of respondents in the morning and middle of the day." (Other)

"Publish it quicker." (Other)

One consultee remarked spontaneously that the review was a useful exercise.

"I think it's useful that the review that you're doing is being undertaken to see whether the NPS is covering all the things that the industry would like." (TOC)

2.12 Overall Assessment of the NPS

Finally, we asked consultees how they thought the NPS was regarded by stakeholders generally. There was very little negative comment (and what little there was tended to restate specific issues already mentioned).

The vast majority of the feedback was very favourable. Many of the TOCs expressed positive opinions.

"It's seen as independent research which is obviously in high regard. It is held as the standard." (TOC)

"It is one of the key drivers in terms of where we focus taking the business forward." (TOC)

"Everybody sees it as an important KPI across the industry and a useful tool to help improve the industry. I think most TOCs have a high regard for it." (TOC)

"I think people are very happy with it . . . The sample size isn't as big as we like which is why we go out and do other things. But it's a good base measure. We are happy with it. It delivers a base, a benchmark, against performance over time . . . and then we do other research to give us the data on other things we want to look at." (TOC)

"It's changed dramatically in the last five years. Five years ago they paid it no mind. Now they take it very seriously. We certainly do. Some seem to take it seriously but try to get out of it and try to find excuses." (TOC)

"I think the whole industry takes it seriously because it's the customer satisfaction barometer in the industry. All TOCs take it seriously but the TOCs with a franchise commitment would analyse it even more." (TOC)

Some other TOCs also expressed positive views but added that the NPS could be used to beat the industry over the head.

"I would like to think they regard it as we do: as a positive, useful tool for measuring customer satisfaction rates. I know certain elements of the media regard it as a twice-yearly opportunity to beat the industry over the head, and likewise certain politicians also have similar views . . . but from my perspective I would prefer to think that people view it in a positive light . . . the way it breaks down the individual elements of the on-board experience allows us to see where we are failing, and to be able to say to our team 'let's look at that', so it's a useful tool in that way." (TOC)

"They find it very useful to action plan service improvements. Hidden behind that, it's felt it's a tool to batter the industry over the head. I don't know if it is but perhaps it sometimes feels that way . . . I wouldn't say that's my personal opinion, I get on very well with Passenger Focus." (TOC)

Other (non-TOC) stakeholders also expressed positive views.

"I think they view it as valuable. I think it's a valued tool for all railway stakeholders." (Other)

"It drives a lot of management activity . . . Overall it is important that there is consistency in the measure of customer satisfaction. Also, that it is done independently (if things aren't done independently people question their freedom from bias)." (Other)

"Very highly I would say, yes certainly. It does seem to have a much higher profile than the BPSS, it seems to be more actively used." (Other)

Some stakeholders (both TOCs and others) expressed positive views but added that TOCs' opinions could be affected by their respective scores.

"I think it's taken seriously, because it's looked at seriously by our stakeholders, the media and the government . . . I think it's viewed with a certain degree of resignation to be honest, but that immediately turns to jubilation if the figures are good!" (TOC)

"I think we probably all use it in the same way, it seems like it's embedded in the industry, a necessary part, everybody probably slightly dreads it - you are either going up or down." (TOC)

"My personal opinion is that yes it's a useful informative piece of work. It's only natural that TOCs with scores that are near the top think it's wonderful, but if not they may take on other ideas about it." (Other)

Another stakeholder who expressed positive views also suggested that the NPS was in some cases being pushed to the limits of its current design.

"Very, very useful. Although I don't think they (*users*) understand the limitations of it when you want to look at a localised level. They don't like the 'health warnings'." (Other)

Taking into account all the comments received during the consultations, the majority view of the NPS can perhaps be summed up by saying that, while there were some criticisms of specific aspects of the NPS, most of those consulted regard the NPS as a unique and important benchmark measure of key elements of satisfaction with rail performance (at the national level, at the TOC level, and now also at building block level), and one that is used as valuable input to their decision making processes.

However, so many of them now wish to drill down further that, as one consultee said (and with only a modest degree of exaggeration):

"The sample size is too small, everybody thinks that." (TOC)

3. THE TECHNICAL REVIEW

3.1 NPS Changes Since 2006

RMA carried out a review of the NPS for Passenger Focus in 2005. The resulting report was published by Passenger Focus as "Findings of a Review of the National Passenger Survey" in February 2006.

Among the conclusions in the 2006 report were the following:

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

At the same time, the review suggested that certain aspects of the sampling and weighting design and other survey elements should be improved.

Several of our suggestions have since been implemented.

Among these, the most important related to the fieldwork shift pattern. This was identified in the Report as an area of weakness in the 2005 NPS sampling method, leading to oversampling of peak outbound (a.m.) journeys and undersampling of peak return (p.m.) journeys. In 2007, around 100 shifts were moved from the morning peak to the evening peak to try to address this imbalance.

We also had some concerns re: shift phasing, and one of the recommendations was that shifts should staggered so that they start hourly (instead of starting simultaneously at 3-hourly intervals). This was done.

These and other changes represent clear improvements to the NPS design that have been implemented since 2006.

We also drew attention to the fact that "TOC-type" sector distinctions were becoming very blurred: TOCs had been divided for the purpose of performance comparisons into three sectors (London/SE, Long Distance, Regional), but by 2005 most TOCs included a mix of route types, and therefore TOC-type analysis could not provide valid like-for-like comparisons. However, TOC-type analysis is still shown in NPS reports in 2010. Fortunately, the "building block" approach is evolving into a potentially superior solution to the problem.

A general comment we made in 2006 was that NPS reports at that time typically contained insufficient detail on the sampling process. We recommended that the standard of NPS technical documentation should be significantly enhanced, and also that efforts should be made to increase understanding and awareness of the survey design and its potential among users. This situation has also improved. For example, the Overview Report that is produced annually contains detailed information on sample design, response rates, key driver analysis and the questionnaires used.

The following sections review a number of technical design issues as they are now implemented in the NPS (i.e. in the waves conducted in 2009 & 2010).

3.2 Sampling & Weighting Rationale

3.2.1 Sampling

It is important when evaluating the NPS sampling methodology to understand that the objective of the NPS is to provide a representative sample of passenger journeys for each TOC. Another objective is that these TOC estimates can be amalgamated to provide a representative sample of all GB rail passenger journeys.

It must be noted that the resultant sample does not give a representative sample of rail passengers. For example, if passenger A commutes Monday to Friday from Sevenoaks to Charing Cross, then he potentially has 10 journeys to contribute to the sample, whereas passenger B may work part-time only one day a week and will only contribute 2 journeys to the sample. Clearly A's journeys are 5 times more likely than B's to be in the sample of journeys. Thus the resulting sample is biased towards more frequently travelling passengers, but it must be argued that this is correct, and is what is required for this study, as the objective is a sample of journeys, not passengers.

Passenger journey volumes are estimated for each TOC (see Section 3.3). The target sample sizes for each TOC are pre-set depending on assessments of the size and complexity of each TOC (see Section 3.4).

The design employed is a two-stage sample, where the first stage consists of individual stations, and the second stage consists of shifts within selected stations. Originally the sampling at the first stage was done for each TOC using a list of all the stations included in the TOC's area. This was modified if a franchise boundary changed, and the "virtual TOC" approach was adopted, but over the years this has evolved, and now most TOCs sampled employ the concept of "building blocks" ("BBs" - see Section 3.8). Some BBs are based on groups of stations, and some are based on routes.

For the station-based BBs, the number of journeys for each station originally calculated for the TOC was assigned to that station within its BB. For route-based BBs, some stations may appear in more than one BB. In these situations passenger volumes were split between BBs. Stations were then selected with probability proportional to size (PPS) using the derived passenger volume figures for each BB. Stratification based on four size strata (defined within each TOC so that each stratum contained 25% of its passengers) was used in this selection process. This means that the very large stations may be selected several times and very small stations will have a low probability of selection. This produces a sample for each TOC that will include all of its largest stations, and only a proportion of the stations in the smaller strata, but is precisely representative of each of its four station size strata in total.

In the RMA 2006 NPS Review Report, the then NPS sampling plan was generally found to be satisfactory, but some faults were found with the allocation of the shifts, particularly re: the pattern of outward/a.m. and return/p.m. journeys. We comment on these points in later sections. We find the current sample design to be generally satisfactory, and believe that it needs only minor changes (e.g. further balancing of time of day and direction of journey - see Sections 3.6 and 3.7).

3.2.2 Weighting

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

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

Rim weighting is an accepted method which is used when data exists separately for each weighting variable but data does not exist for the cells (cross-tabulations) between them.

The survey results for all TOCs are then simultaneously weighted for the above factors (and at the same time grossed up, according to the total number of journeys per annum for each TOC, so that the national totals contain the correct ratios of passengers per TOC).

The results on the following page show the differences between weighted and unweighted results for the satisfaction scales on all Waves from 13 to 22. Weighting consistently reduces the scores for overall satisfaction by about 1%, and all other measures by about 2%. The consistency of the difference suggests that the weighting, i.e. the process of matching the sample profile achieved to the profile targeted, does not greatly alter the results (which is a positive comment in survey design terms).

A review of the merits/demerits of the above three weighting variables is contained in the RMA 2006 Review. The recommendation there was that “the NPS weighting technology (rim weighting) is satisfactory, but that the actual weight matrix has some weaknesses, principally that journey purpose weighting is based on ticket type, which is not an accurate predictor of journey purpose”. Further, the report says “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”.

The 2008 report on weighting by Paul Harris Associates (PHA) for Passenger Focus also looked into this aspect of the weighting. Extensive weighting analyses were carried out, replacing journey purpose with actual ticket type. The main conclusion was that changing the weighting regime in this way did actually consistently decrease the overall satisfaction levels for the survey by an average two percentage points down. The PHA 2008 report also suggested that the change to ticket type weighting should be considered.

In any event, the differences between the unweighted data, the current (journey purpose) weighted data and the proposed (ticket-type) weighted data are all relatively small. However, the differences are enough that changing the journey purpose weighting to ticket-type weighting would conflict with the objective of maintaining consistency of comparison with back data, and it was decided to retain the current weighting method.

NPS Satisfaction Scores - Weighted v Unweighted Results

Source: NPS Waves 13 to 22

Negative values mean weighted scores were lower than unweighted by that amount.

	Total	Autumn '05 Wave 13	Spring '06 Wave 14	Autumn '06 Wave 15	Spring '07 Wave 16	Autumn '07 Wave 17	Spring '08 Wave 18	Autumn '08 Wave 19	Spring '09 Wave 20	Autumn '09 Wave 21	Spring '10 Wave 22
Number of journeys per annum (000's)	273569	28286	27009	25672	25868	28482	25664	26448	27474	26849	31617
Overall Satisfaction	-1	0	-1	0	-1	-1	0	0	-1	-1	-2
STATION - Ticket Buying Facilities	-2	-3	-3	-2	-2	-3	-2	-2	-2	-2	-2
STATION - Provision Of Information About Train Times/Platforms	-1	-1	0	-1	-1	-1	-1	0	-1	-1	-2
STATION - The Upkeep/Repair Of The Station Buildings/Platforms	-2	-1	-1	-1	-1	-2	-2	-2	-2	-2	-2
STATION - Cleanliness	-1	-1	0	-1	-1	-2	-4	-3	-2	-2	-4
STATION - The Facilities And Services	-3	-3	-2	-1	-1	-4	-4	-3	-3	-4	-4
STATION - The Attitudes And Helpfulness Of The Staff	-1	-2	-2	-1	-1	-1	-1	-1	-1	-2	-1
STATION - Connections With Other Forms Of Public Transport	0	0	1	1	1	1	1	0	0	0	-1
STATION - Connections With Other Forms Of Public Transport	-3	-4	-3	-3	-2	-3	-2	-3	-2	-3	-2
STATION - The Overall Environment	-2	-2	-1	-1	-1	-2	-2	-2	-2	-2	-2
STATION - Your Personal Security Whilst Using	-2	-2	-2	-1	-2	-3	-3	-2	-2	-3	-2
STATION - The Availability Of Staff At The Station	-2	-3	-2	-2	-2	-2	-3	-2	-1	-1	-2
STATION - How Request To Station Staff Was Handled	-1	-1	-1	-1	-1	0	0	-1	-1	-1	-2
STATION - The Frequency Of The Trains On That Route	-1	-1	-1	-1	-1	-2	-2	-1	-1	-2	-2
TRAIN - Punctuality/Reliability (i.e. The Train Arriving/Departing On Time)	0	0	0	1	0	0	1	0	-1	-1	-2
TRAIN - The Length Of Time The Journey Was Scheduled To Take (Speed)	-1	-1	-1	-1	-2	-1	-1	-1	-1	-2	-2
TRAIN - Connections With Other Train Services	0	0	0	0	0	0	0	0	0	0	-1
TRAIN - The Value For Money For The Price Of Your Ticket	-2	-1	-1	0	0	-2	-1	-1	-3	-4	-4
TRAIN - Cleanliness Of The Train	-2	-2	-2	-1	-2	-3	-2	-2	-3	-3	-4
TRAIN - Upkeep And Repair Of The Train	-2	-2	-2	-1	-1	-2	-1	-2	-2	-3	-3
TRAIN - The Provision Of Information During The Journey	-1	-2	-2	-1	-1	-2	-1	-1	-2	-1	-3
TRAIN - The Helpfulness And Attitude Of Staff On Train	-5	-5	-4	-4	-4	-6	-5	-4	-5	-6	-6
TRAIN - The Space For Luggage	-3	-3	-3	-3	-3	-2	-2	-2	-3	-3	-2
TRAIN - The Toilet Facilities	-4	-4	-4	-3	-3	-4	-4	-4	-5	-6	-7
TRAIN - Sufficient Room For All Passengers To Sit/Stand	-1	-1	-1	-1	-1	-1	-2	-1	-2	-3	-2
TRAIN - The Comfort Of The Seating Area	-2	-2	-2	-2	-2	-2	-2	-2	-2	-3	-3
TRAIN - The Ease Of Being Able To Get On And Off	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
TRAIN - Your Personal Security Whilst On Board	-3	-3	-3	-3	-3	-4	-4	-3	-3	-4	-3
TRAIN - The Cleanliness Of The Inside	-2	-2	-2	-2	-2	-3	-2	-2	-3	-3	-4
TRAIN - The Cleanliness Of The Outside	-1	-1	-1	-1	-1	-2	-1	-2	-2	-3	-3
TRAIN - The Availability Of Staff	-6	-7	-7	-6	-5	-7	-6	-5	-6	-7	-7
TRAIN - How Well Train Company Dealt With Delays	-3	-3	-2	-1	-2	-3	-3	-2	-3	-3	-5
Average difference	-2	-2	-1	-1	-1	-2	-2	-1	-2	-2	-2

3.3 Data Sources for Journey Volumes

The RMA 2006 Review Report refers to the CAPRI/LENNON database as having been the main source then used for calculating the numbers of journeys per station. Continental have supplied us with a detailed description of the method currently employed using the ORR database (see text on the following page). It describes a heuristic/iterative procedure very akin to that used by rim weighting. Continental noted that the ORR database does contain other sources of journey information in addition to ticket sales from LENNON, to cover sales made through other channels. For one or two TOCs the data does not match well and data the TOC supplies may be used as a substitute, but in general the iterative method produces estimates for each TOC/station combination that aggregate to both the TOC totals and the station totals.

Once the numbers have been produced they are sent to the TOCs for comment and in some cases TOCs may suggest revisions to the estimates. Discussion then takes place with the TOC and agreement on the “best” figure is reached.

It is our recommendation that this allocation/estimation process continues to be used as the basis of the survey sampling frame. Further we feel that the description on the following page should be included in the survey documentation (at least in an abridged form) given its importance in designing the sample.

Other data sources are needed to produce the final sample of shifts, namely to allocate shifts across day of week and time of day, and these are discussed later.

Methodology for Calculating Journey Volumes by TOC and Station

This text describes how ORR data is used to calculate passenger journey volumes for each TOC at each station in the national rail network (Source: Continental, 2010).

Step 1

Passenger journey data for each station is taken from the ORR database. This database uses ticket sales data from LENNON supplemented with journey data from a number of other sources that LENNON does not include, principally:

- journey data from TfL for London Underground stations that offer national rail services
- PTE journeys from sales that are made from sources other than national rail stations

The data used is number of entries plus number of interchanges. For example, the total annual passenger journeys estimated from London Victoria is 39,626,050 (35,127,971 entries and 4,498,079).

Step 2

This data is then aggregated for all stations across the rail network and compared to the total obtained by aggregating data for all TOCs as supplied by DfT. In 2010, the station aggregation total was 1,227,778,667, whereas the DfT TOC total was 1,240,218,685. An adjustment factor is calculated for each station so that the station totals add to the TOC totals – this initial adjustment factor is 1.010132134 and the adjusted total for London Victoria is 40,027,546.

Step 3

Data from the electronic timetable is used to count how many services each TOC runs from a station in the four weeks in February each year. This is then profiled, so that we estimate what percent of the services run from a station are by each TOC. At London Victoria, the % breakdown of services run from the station is as follows:

Southeastern	28.07%
Gatwick Express	11.88%
Southern	60.05%

Step 4

These profiles are then applied to the total passenger count for the station derived in step 1. Implicitly, the assumption is that the proportion of journeys by TOC from the station is the same as the proportion of number of services by TOC from the station. For London Victoria, this results in estimated passenger volumes as follows:

Southeastern	11,235,150
Gatwick Express	4,756,615
Southern	24,035,782

Step 5

The total estimated passenger journeys for each TOC is computed by adding up the estimate for each station at which the TOC calls. For Southeastern, this gives a total of 162,471,848 compared to the TOC total of 154,073,470. This produces a TOC scaling factor for Southeastern of 0.94830872. A similar process for Gatwick Express and Southern produces factors of 0.72579627 and 1.08620260 respectively.

Step 6

These factors are then applied, TOC by TOC, to the estimated passenger journeys for each station at which the TOC calls. This gives an updated estimated passenger journeys for the TOC for each station. So at London Victoria, the updated figures are as follows:

Southeastern	10,654,391 (11,235,150 times 0.94830872)
Gatwick Express	3,452,333 (4,756,615 times 0.72579627)
Southern	25,942,024 (24,035,782 times 1.08620260)

A revised estimate for London Victoria is then calculated by adding up these totals - 40,048,747 compared to the original station total of 40,027,546. A station scaling factor for London Victoria is now produced - 0.9994706.

Steps 5 and 6 are then repeated until the process converges in that station factors remain as they were from the previous iteration (TOC totals are preserved in the final run as these are regarded as sacrosanct).

At the end of this process we have a set of estimated passenger journeys for each TOC at each station that adds to the TOC totals and adds as closely as possible to the station totals.

3.4 Sample Sizes Per TOC

Sample sizes are important for two reasons, (i) they have to be big enough to reflect the size and complexity of the TOC network and (ii) they have to be big enough to give a required precision (confidence limits).

3.4.1 Sample Size, Passenger Volume & Network Complexity

The Detailed Technical Survey Overview for Waves 21/22 states on the question of sample sizes that “Each TOC has a target sample size. Initially this was set at 500 for each TOC”. It goes on to describe how some sample sizes were raised over time to 750, 1,000, 1,500, 2,000 and in one case to 2,750.

The table below gives the current targets for all TOCs.

TOC Passenger Volume, Complexity and Target Sample Size

FRANCHISED TOCS	annual volume	route structure	CURRENT TARGET	SUGGESTED TARGET	reason
SOUTH WEST TRAINS	190,064,794	complex	1750	1500	large and complex
SOUTHERN	162,014,190	complex	1500	1500	large and complex
SOUTHEASTERN	153,263,554	complex	1500	1500	large and complex
NATIONAL EXPRESS EAST ANGLIA	106,689,221	complex	2000	1500	large and complex
FIRST CAPITAL CONNECT	97,671,860	complex	1500	1500	large and complex
NORTHERN RAIL	94,517,769	complex	1000	1500	large and complex
FIRST GREAT WESTERN	83,870,434	complex	2750	1500	large and complex
SCOTRAIL	73,238,340	complex	1000	1500	large and complex
LONDON OVERGROUND	55,745,618	complex	750	1250	complex
LONDON MIDLAND	52,930,162	complex	1000	1250	complex
CROSSCOUNTRY	29,700,417	complex	1000	1250	complex
ARRIVA TRAINS WALES	26,419,732	complex	750	1000	complex but small
VIRGIN	23,171,843	complex	1000	1000	complex but small
EAST MIDLANDS TRAINS	22,316,655	complex	1000	1000	complex but small
TRANSPENNINE EXPRESS	22,294,549	complex	1000	1000	complex but small
EAST COAST	17,732,582	complex	1000	1000	complex but small
MERSEYRAIL	40,081,911	simple	500	750	simple
C2C	32,175,381	simple	1000	750	simple
CHILTERN RAILWAYS	17,768,185	simple	1000	750	simple
TOTAL			23000	23000	

Source: Continental Research

Although the current targets appear to follow a fairly logical pattern, with target sample sizes being to some extent correlated with annual volume and complexity, there are also some apparent anomalies, e.g. First Great Western and National Express East Anglia.

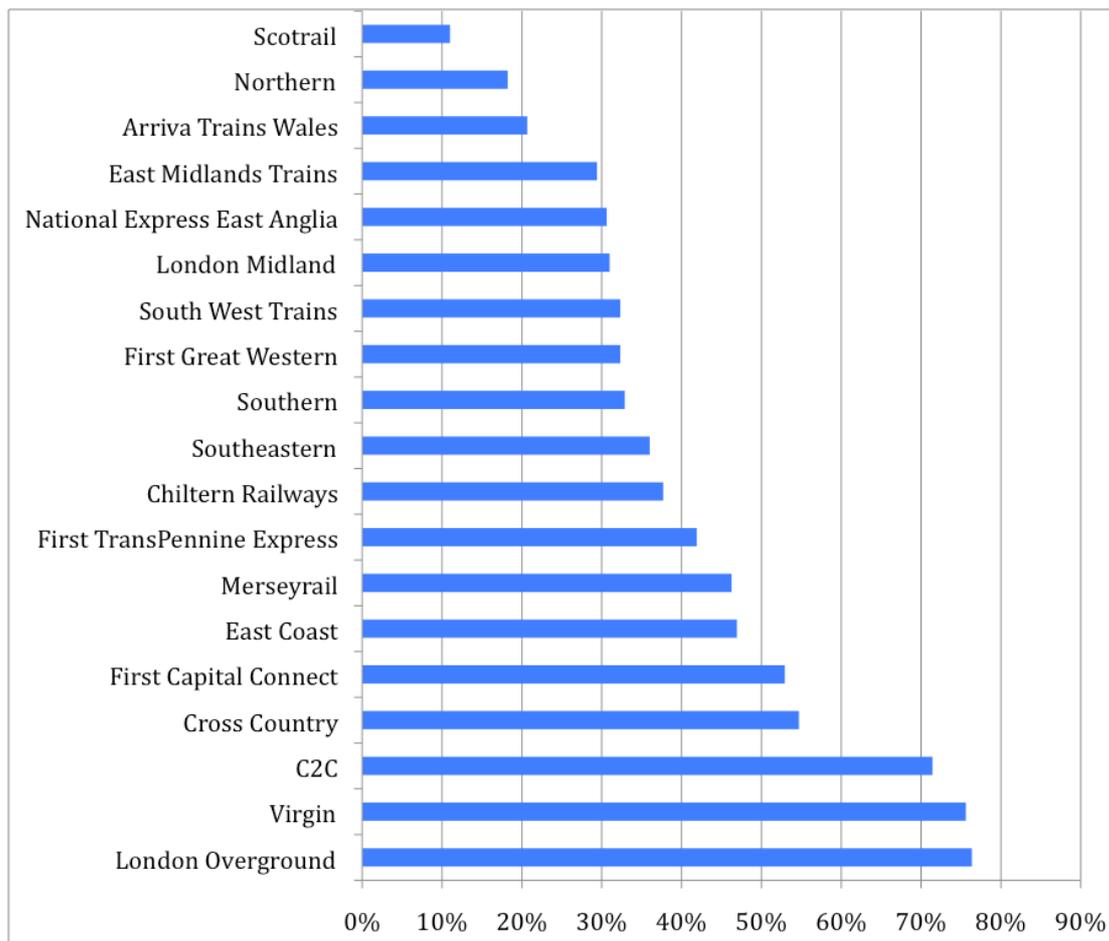
Continental Research, who supplied us with above table, suggest that the target sample sizes could be rationalised and equalised in the future (see 'suggested target' column), the better to reflect the size and complexity of each TOC's network, and we concur.

3.4.2 Sample Size and Station Coverage

However, as well as size and complexity⁶, we believe there is one further factor which it may be useful to consider in this context, and that is the extent to which the sample covers each TOC's list of stations.

If we calculate, for each TOC, the proportions of its stations actually surveyed (total number of stations surveyed per wave over the total number of stations served) it can be seen that there are extremely wide variations. Only 11% of Scotrail's stations were surveyed in Spring 2010, whereas 76% of London Overground was covered.

Proportion of Stations Surveyed by NPS in Spring 2010



Source: Data provided by Continental Research.

Adjusting the target sample sizes per TOC (within any given NPS total sample size) is likely to make only a small contribution to countering this effect, as it is the result of various factors: e.g. the chosen target sample size (which determines the total number of shifts available); the total number of stations the TOC serves (and over which the sample could be distributed); and the extent to which the TOCs passenger journey total is concentrated into a few very large stations (these will naturally attract a larger proportion of the total sample than will small stations).

⁶ Assuming size refers to journey volume and complexity refers to route complexity.

A TOC whose network includes a large proportion of very small stations - say 5 large stations and 45 very small ones - will probably have a much smaller proportion of all its stations covered than would a TOC with 50 roughly average sized stations (even if both TOCs have the same total number of passengers journeys and the same target number of total interviews). However, in both cases the TOCs will have roughly the same proportion of their passenger journeys surveyed. In both cases the sample will include a representative mix of the TOC's passenger journeys from large and small stations.

This is a logical outcome of the sample design, which was intended to provide a representative sample of each TOC as a whole (including a representative profile of each TOC by station size), and it does.

This issue has become much more visible because there has been an increasing level of interest in drilling down below TOC level and looking at individual routes and individual stations. TOCs are finding that this can be done for large stations, or large groupings of routes or stations (e.g. building blocks), but is not always possible at medium station level (as base sizes may be small) and is often impossible at very small station level (as many of them are not covered at all in any given survey Wave).

There are approximately 2522 stations in GB. The NPS typically samples all the largest stations, and decreasing proportions of the second, third and fourth size strata. This means that, although the NPS samples only about 695 stations in each wave (i.e. about 28% of all 2522 stations), those 695 stations handle about 82% of all GB passenger journey departures. In these terms, the NPS is an efficient design, and produces a representative sample of journeys at stations of all sizes.

The corollary of the above ratios is that about 1827 stations are not sampled in any given NPS wave. However, these 1827 stations (despite being 72% of all stations) between them account for only 18% of all GB passenger journeys. Another way to look at this is that the stations not covered, on average, handle one-twelfth as many journeys as do the ones sampled.

Two stakeholders suggested that we explore the possibility of increasing coverage of smaller stations by trimming sample sizes of very large stations, e.g. London Victoria could perhaps be "capped", and the interviews thus saved could be used to increase coverage of smaller stations that feed into it.

Unfortunately, on closer examination, this is not a very attractive solution. It would alter the selection probability for station strata (PPS, as described in Section 3.2), and would therefore require a further level of weighting (not necessarily beneficial, as it would tend to reduce effective sample size). Also, NPS users who are interested in the data for very large stations may find the reduced sample base (and further reduced effective sample size) to be less acceptable.

A further adverse factor to consider is that small stations generate fewer interviews per shift. Skewing the sample in favour of smaller stations would reduce the average number of interviews per shift, and thus reduce total NPS sample size and at the same time increase the cost per NPS interview.

A possible alternative approach would be, at small stations only, to increase shift numbers by 50% but cut their duration from 3hrs to 2hrs. This would improve small station coverage without changing the total number of small station interviews, and without materially changing the total NPS sample size, i.e. the average number of interviews at each small station covered would fall by about 33%, but the number of small stations covered would increase by about 50%.⁷

The cost per interview (at small stations only) would increase marginally due to the shorter shift length. No additional weighting would be required, provided the shift selection ratio is adjusted appropriately in that stratum.

Calculating the exact cost:benefit ratio of this approach would require a more detailed study than has been possible in the time available to us, but we believe this option should be explored.

This approach would clearly improve granularity, but the numbers of interviews at each small station covered would still be too small to allow them to be analysed individually, and it would still leave many small stations without any coverage in any given survey wave.

Unfortunately, the only way we believe it would be possible substantially to improve coverage of smaller stations would be to increase the overall NPS sample size greatly (e.g. x4 or x8), thus allowing for far more shifts per 100 stations. This would allow not only greatly increased coverage of smaller stations, but would also significantly improve coverage and granularity by day of week and time of day at all medium and large stations. If it is done without skewing the sample towards small stations, the average cost per interview would be unaffected, but the total cost would increase to what we imagine would be an unacceptable degree (e.g. x4 or x8).

⁷ Alternatively increase small station shift numbers by 100% and cut shift length from 3hrs to 1hr 30 minutes, in which case the average number of interviews at each small station covered would fall by about 50%, but the number of small stations covered would increase by about 100%.

3.5 Design Effects & Accuracy

The RMA 2006 Review report gives extensive calculations of design effects which describe the impact of weighting on effective sample sizes and compares unweighted sample sizes with their weighted equivalents. Overall it was found on Wave 11 that the weighted estimates for the total survey had to have their precision (confidence limits) based on a sample size of 14,439 rather than the unweighted sample size of 25,596, i.e. a substantial reduction (44%).

For Wave 22 the equivalent figures were 15,223 and 29,057, i.e. a similar degree of reduction (47%). This is due to the substantial weighting necessary to bring the TOC passenger numbers in line with their correct volume estimates. However, it must be pointed out that even after weighting the total sample results are based on a massive effective sample size of 15,223 and the impact of design effects on the confidence limits of total (i.e. national) survey estimates will be negligible.

At the TOC level, only the first stage of weighting applies to data for each TOC separately (i.e. the weighting required to combine the TOC data into national data does not affect the individual TOC results). As a result, at TOC level, the effective sample sizes are generally only about 20% smaller than their unweighted sample sizes: clearly a very satisfactory result.

To be pedantic, two potential sources of ‘error’ in survey results that are often confused are accuracy and precision. Accuracy refers to getting the right answer and precision refers to saying how precise our survey results are around that right answer. Thus the weighted survey results give an accurate picture (they are the right answer) whereas the unweighted answers are in a sense wrong and not accurate. However, it must be said that when people ask for accuracy they usually mean they want estimates of precision as measured by confidence intervals.

The defining factor for precision is the effective sample size on which the survey results are based and on the NPS we have (from the 2010 User Guidance Report) confidence intervals by TOC ranging from +/- 1.81% to +/- 3.84%. Most are in the range of about 2.2% to 3.0%, which would be deemed satisfactory for surveys of this type. Increasing precision to a noticeable degree requires a large increase in sample size. As illustrated by the following examples (for a result of 70% at 95% confidence) you have to quadruple your sample size to halve the error limits.

SAMPLE SIZE	CONFIDENCE LIMITS
2,000	+/- 2.0%
1,500	+/- 2.3%
1,250	+/- 2.5%
1,000	+/- 2.8%
750	+/- 3.3%
500	+/- 4.0%

3.6 Spread of Shifts Across Days & Hours

Once the station sampling has taken place, the shifts need to be allocated: a day of week is assigned at random to each shift, based on the day of week profile provided probability proportional to passenger volumes. Profiles are determined separately for city centre and other stations, and separately for weekdays and weekends and (according to the latest NPS Overview Report) the distributions currently used were based on data from NPS Wave 9 (tables below). Following the 2006 Review, which commented on the imbalance in outbound:return ratios, "approximately 100 shifts" were moved from the morning peak time to an evening peak time (although the details appear to be undocumented).

Profile of Passenger Journeys by Day & Time

City Centres	%	%	%
Time band	Weekday	Weekend	Total
06:00 – 10:00	8.02	0.33	8.35
10:01 – 13:00	19.48	15.88	35.36
13:01 – 16:00	22.01	5.91	27.91
16:01 – 19:00	25.32	0.37	25.69
19:01 – 22:00	2.52	0.16	2.68
Total	77.35	22.65	100.00

Other Stations			
Time band	Weekday	Weekend	Total
06:00 – 10:00	48.73	0.51	49.24
10:01 – 13:00	27.93	10.78	38.70
13:01 – 16:00	5.98	0.79	6.77
16:01 – 19:00	4.99	0.04	5.03
19:01 – 22:00	0.26	0.00	0.26
Total	87.88	12.12	100.00

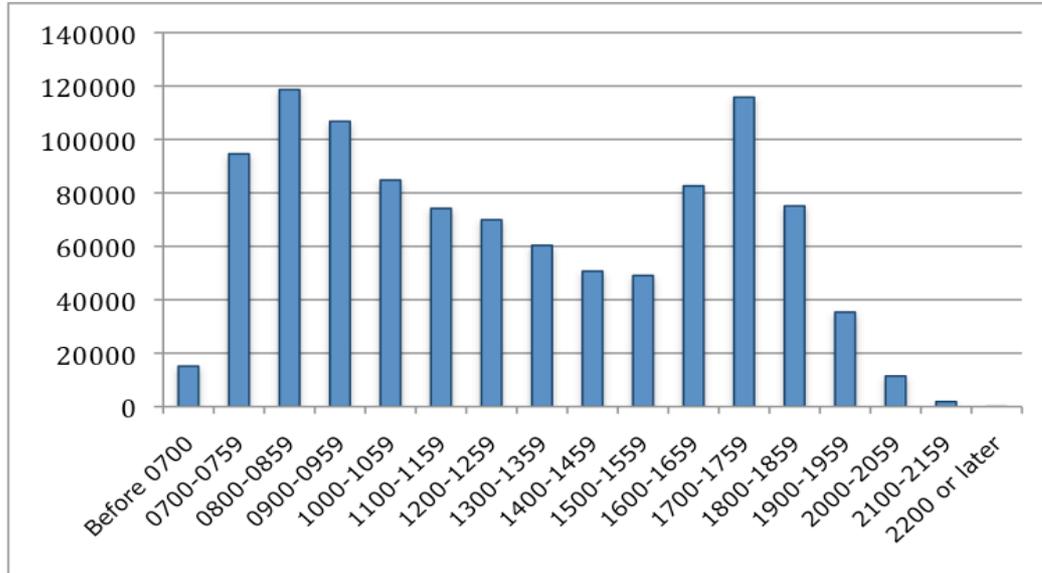
NB: Each of the two tables above is a matrix in which the weekday + weekend columns together add to 100%. Source: Wave 9 NPS data, Continental Research.

It will be noted that the data used came from NPS Wave 9 (now 7 years old). It is recommended that more up-to-date data be used in future even if it is thought that the pattern of distribution will have not changed much over the period, as regular updating (i.e. small incremental changes) will also avoid the need for less frequent but larger changes which may disturb trend lines.

More important, it would be sensible to seek external validation for these assumptions, rather than merely continuing to using past NPS results for setting targets for future NPS samples (which is how the errors in the outward:return ratio persisted from Wave 1 until Wave 17 - see Section 3.7).

It is already clear that the outward: return ratio still remains out of balance (see Section 3.7). It appears that the distribution by hour may also be skewed:

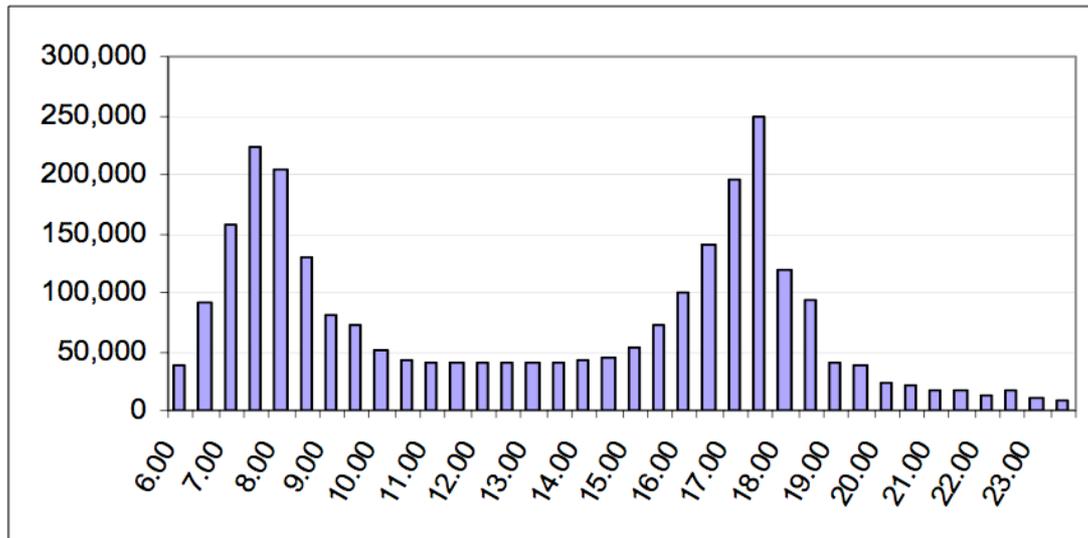
HOURLY PROFILE OF NPS RAIL JOURNEYS - WEEKDAYS



Data: All GB, Monday to Friday, 6am to 10pm.
Source: NPS weighted database, Wave 22 (Spring 2010), CR.

The NPS pattern (above), differs somewhat from that of NRTS (below): NPS appears too high between about 10am and 1pm, and too low around 5pm to 7pm.

HALF-HOURLY PROFILE OF GB RAIL JOURNEYS - WEEKDAYS



Data: All GB, Monday to Friday, 6am to midnight, fieldwork 2004 & 2005.
Source: NRTS (National Rail Travel Survey) Final Report, May 2008, DfT + Transport Scotland.

This anomaly must be investigated and corrected as necessary. It is almost certainly linked to the outward: return issue discussed in the next section.

3.7 Balance of Outward and Return Journeys

The RMA 2006 Review Report dealt at length with this issue. The conclusion reached was that the balance of outward and return journeys should be equalised. On Waves 1 to 13 it was found that the split was about 2 to 1 in favour of outward journeys, whereas it clearly should be around 50:50. The report recommended that this should be corrected, and at Wave 17 approx. 100 shifts were moved from the morning peak to the evening peak to address the imbalance.

Outward and Return Journey Ratios Waves 12 to 22

	W12	W13	W14	W15	W16	W17	W18	W19	W20	W21	W22
OUTWARD	62	64	63	64	64	52	53	54	54	54	54
RETURN	34	32	33	33	33	44	44	42	41	42	42
ONE WAY	3	3	3	3	3	3	3	3	4	3	3
DK/NA	1	1	1	0	1	1	0	1	1	1	1

Source: Continental Research.

Clearly the shift changes have improved the ratio, but it is still not fully balanced. We recommend that further measures be taken to bring the ratio closer to 48:48.⁸

The graph on the following page plots the total Spring 2010 weighted NPS passenger numbers by departure station against the total number of weighted NPS passengers arriving at that station (from NPS Q2b). If every station had an equal number of outward and return journeys the result would be a straight line.⁹

Given the fact that small and medium volume station samples are small, and sampled on different days and at different times (i.e. traffic patterns will vary between shifts), we would not expect the plotted line to be exactly straight.

However, the wide swings apparent even for stations with 10 million or more passengers p.a. (i.e. roughly between about 100 and 810 NPS interviews per wave) suggests that imbalances in outward v return journeys are affecting even the largest stations (and imbalances extend on both sides of the trend line). Given that these are weighted figures, these variations should not have arisen from variations in TOC sample targets, or from differences between stations shared by more or fewer TOCs.

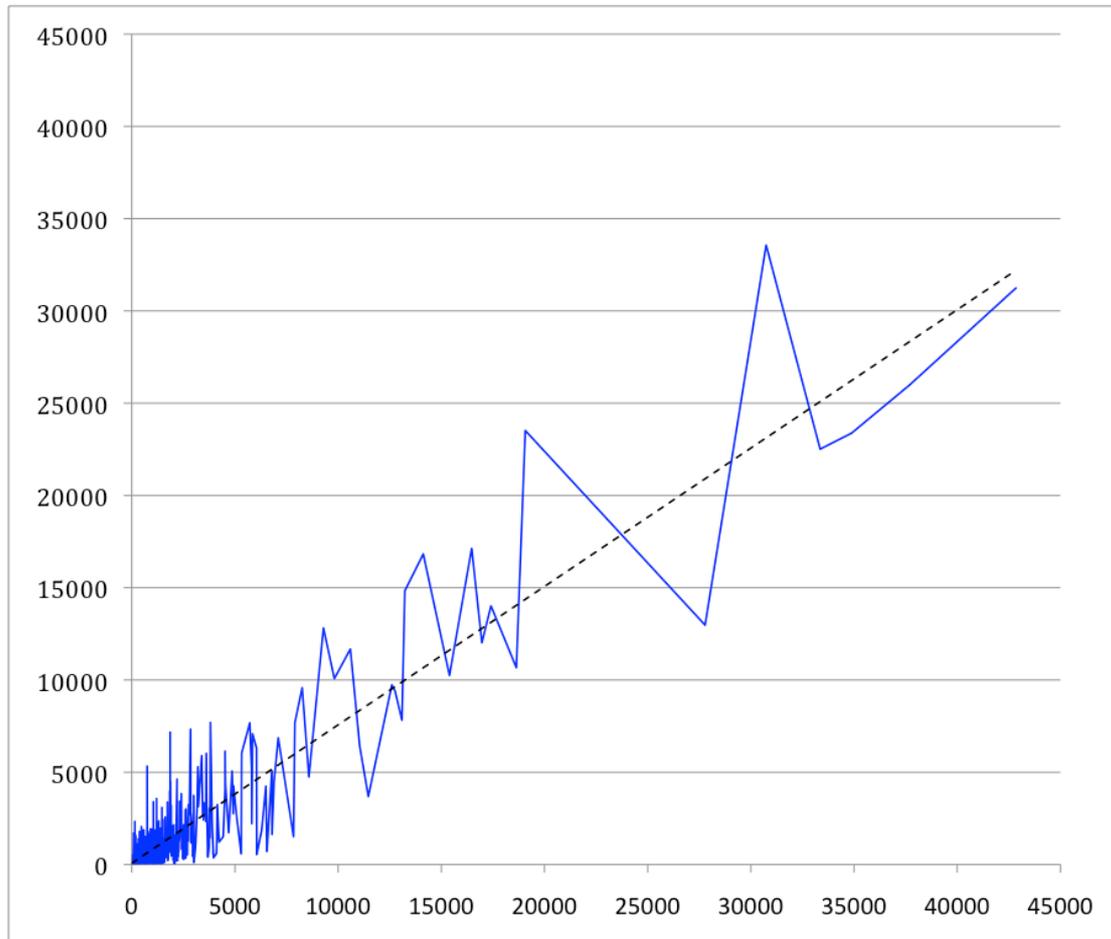
The review timetable has not allowed us to investigate this issue in more detail, but we recommend that further work is done to identify the source of these variations.

⁸ In practice it will be around 48:48 rather than 50:50, to allow for the small number of one way journeys (single legs in which the journey in the opposite direction, if any, may not be by train, or may be separated by an extended period of time) and for "don't know" (DK) and "not answered" (NA) responses.

⁹ If the NPS sample covered all GB departure stations the trend line would be at 45 degrees, but because the NPS covers only about 695 stations (which handle about 82% of departures), about 18% of the NPS departures will arrive at stations not included in the departure station sample, and hence cannot be plotted, therefore the trend line is flatter than 45 degrees.

This graph shows weighted NPS origin (departure) numbers plotted against weighted NPS destination numbers (Q2b), sorted by departure station passenger volume.

ORIGIN v DESTINATION VOLUMES FOR NPS STATIONS
(Origin = horizontal scale; Destination = vertical scale)



Source: NPS Wave 23 weighted departures v weighted arrivals, Spring 2010.

While ratios vary more widely at smaller stations as one would expect (bottom left) it is perhaps surprising that several of the largest stations in Britain have quite noticeable variations from the mean (top right).

The table on the following page uses the same weighted NPS departure and arrival data, but expresses them as percentages of weighted NPS departures plus arrivals, for all stations with passenger departures in excess of 5 million p.a.

This table is sorted by departure percentage, rather than station size, to illustrate the range of variance.

ORIGIN v DESTINATION RATIOS FOR LARGEST STATIONS

Stations over 5,000,000 pax p.a.	Actual Annual Departures	Weighted NPS Departures (Origin)	Weighted NPS Arrivals (Destination)
CAMBRIDGE	5,390,646	76%	24%
LEEDS	13,248,902	68%	32%
LONDON CANNON STREET	10,694,782	64%	36%
EAST CROYDON	17,998,581	64%	36%
WIMBLEDON	9,625,759	63%	37%
STRATFORD (LONDON)	9,471,103	63%	37%
EDINBURGH	8,094,232	60%	40%
CLAPHAM JUNCTION	27,108,758	60%	40%
LONDON LIVERPOOL STREET	29,988,206	60%	40%
LONDON VICTORIA	37,519,371	59%	41%
LONDON CHARING CROSS	18,580,180	59%	41%
LONDON WATERLOO	51,352,701	58%	42%
READING	11,188,510	58%	42%
VAUXHALL	7,713,710	57%	43%
GLASGOW CENTRAL	13,590,877	56%	44%
LEWISHAM	5,710,323	56%	44%
LONDON PADDINGTON	18,184,824	55%	45%
SURBITON	5,806,970	53%	47%
LONDON MARYLEBONE	6,904,904	51%	49%
LIVERPOOL CENTRAL	8,337,880	51%	49%
GATWICK AIRPORT	5,658,576	49%	51%
LONDON EUSTON	12,253,970	49%	51%
CARDIFF CENTRAL	6,504,288	49%	51%
WOKING	5,533,115	49%	51%
LONDON BRIDGE	32,008,339	48%	52%
GLASGOW QUEEN STREET	9,054,048	48%	52%
LONDON KINGS CROSS	12,597,886	47%	53%
LONDON FENCHURCH STREET	9,503,534	46%	54%
BIRMINGHAM NEW STREET	17,297,396	46%	54%
LONDON ST PANCRAS	14,681,769	45%	55%
BRIGHTON	9,439,813	43%	57%
BRISTOL TEMPLE MEADS	5,455,846	42%	58%
MANCHESTER PICCADILLY	12,034,931	42%	58%
LIVERPOOL LIME STREET	6,916,646	33%	67%
FINSBURY PARK	5,179,511	29%	71%
GUILDFORD	5,295,480	28%	72%

Source: NPS Wave 23 weighted departures v weighted arrivals Q2b, Spring 2010, CR.

3.8 Building Blocks

When TOC boundaries changed, e.g. when Silverlink was split into "Silverlink Metro" and "Silverlink County, and "one" was split into "one Great Eastern" and "one West Anglia", the NPS treated the transferred sections each as "virtual TOCs". This very useful concept was introduced to allow for like-for-like comparisons with historical NPS data even after some sections of a network had been transferred from one (real) TOC to another.

As long as Passenger Focus is aware of impending franchise area changes sufficiently in advance of fieldwork dates, the virtual TOC approach will allow the NPS to establish a prior track record for any geographic area in a way that can be combined to generate both pre- and post-franchise change results.

The "building block" (BB) approach takes the virtual TOC approach a stage further by dividing TOCs into segments. Effectively this provides an extra layer of sample stratification, allowing greater confidence when looking at the results for each BB when analysed separately.

Most TOCs use route-based definitions for the BB segments (a few TOCs define BBs based on station groupings).

We understand that a draft classification of building blocks has been suggested, based on a typology of seven categories :

- Short commute
- Long commute
- Interurban
- Rural
- Long distance
- High speed
- Airport

This is a sensible approach which, if carried out consistently for all TOCs, would enable NPS users to compare like with like. A TOC could, for example, compare results for its long distance routes with the aggregate of the long distance routes of all other TOCs. This would be far superior to the current "TOC-type" sector breakdown (London/SE, Long Distance, Regional) as very few of the TOCs fit exclusively into any one of these categories.

We recommend that the building block approach be maintained in the sampling and weighting scheme. If it employs a consistent typology across all TOCs (which we understand is not currently true for all TOCs) it will allow for more rational comparisons between TOC scores than is provided by the current (and invalid) "TOC-type" breakdown.

3.9 Future Weighting

Ticket Type / Journey Purpose Weighting

The 2006 Review concluded the NPS weighting technology (rim weighting) was satisfactory, but that the actual weighting matrix had some weaknesses, principally that journey-purpose weighting used in 2005 was based on ticket type which is not an accurate predictor of journey purpose. The debate surrounding journey purpose/ticket type weighting has continued and was the subject of the 2008 PHA report on weighting, which reached similar conclusions. However, the 2005 approach has continued to be used in the NPS design, in order to maintain the consistency of satisfaction score measurements. Thus the current weighting procedure remains: to weight by station size strata for each building block and by journey purpose (based on ticket type) and day of week for each TOC. As suggested in Section 3.2.2, we recommend that the journey purpose/ticket type discussion remains on the table. We believe that weighting of ticket type by ticket type (rather than weighting journey purpose by ticket type proxy) should be reconsidered as a solution, as ticket type volume data is one of the very few independent sources of detailed universe data. Even if the desire for consistency currently inhibits the change, it would be possible when setting up any new NPS-based target scores to set the targets based on new satisfaction scores (i.e. scores using the new ticket type weighting). It would be possible to produce both versions of the scores in parallel for a period if necessary.

Station Strata Weighting

This topic was discussed in Section 3.3. The weighting appears to be satisfactory.

Day of Week/Time of Day Weighting

This topic was discussed in Section 3.6. The main issue is to obtain confirmation that the matrix of days and (especially) times being used is accurate. If the matrix is then used to determine shift allocation, no target weighting will be necessary. However, other weighting may be considered (see following Section 3.10).

Route Weighting

We consider that the issue of any need for weighting by route can probably be dealt with satisfactorily by adopting a consistent approach to building block definition (see Section 3.8).

Demographic Weighting

In the absence of independent data on the demographic profile of passengers (which in any case varies widely by station, route, season, day, time etc.), the NPS itself provides the best source of demographic profile data, and so targeted demographic weighting is not possible (i.e. you would be weighting the NPS to match itself). However, other weighting may be considered (see following Section 3.10).

Weighting for PTE Analyses

This topic was not addressed explicitly in the 2006 report. The current User Guidance Report describes it as follows: "Boost shifts are grafted on to the NPS to augment the local journeys in the six PTE areas (Greater Manchester, Nexus, South Yorkshire, Strathclyde, West Midlands, and West Yorkshire). Some of these are not included in the main NPS database, as this would skew the profile too much towards local journeys".

Based on discussions with Continental, it appears that PTE geographies do not match building block geographies. For example, the Greater Manchester PTE includes parts but not all of the following building blocks:

- Arriva Trains Wales: North Wales
- Crosscountry: Birmingham - Manchester
- East Midlands Trains: Liverpool - Norwich
- Northern Rail: Manchester - Liverpool
- Transpennine Express: North
- Transpennine Express: Northwest
- Transpennine Express: South
- Virgin: London - Manchester

The data for GMPTE is obtained by filtering out from the above building blocks any journeys that both start and end within the GMPTE area.¹⁰ Although for example, the journeys for Arriva Trains Wales will be weighted to represent the profile of that TOC by station size, day of week and journey purpose, the journeys that happen to fall within the GMPTE area will not necessarily reflect the profile of such journeys.

For this reason, PTE area results have never been weighted, since use of existing weights based on building block profiles might be misleading and there is no other independent data to which to weight the PTE passenger journeys.

We understand that one possible approach that has been discussed is to take the weighted profile for each PTE aggregated over four or six waves and use this profile to weight the PTE data on this constant (and thus consistent) basis.

We recommend that approach is tried as it will enable one to see how stable the PTE profile is from wave to wave and perhaps obtain a benchmark by which to weight further waves. A similar approach could apply to other sub-TOC level analyses that may be considered.

¹⁰ Although this means that the sample will exclude very many journeys made by Greater Manchester residents, i.e. residents setting off from GMPTE to destinations outside the area, and residents returning to GMPTE from destinations outside the area. Including either or both groups would increase the sample base size available to GMPTE for analysis purposes, and would provide a more complete picture of travel in the area.

3.10 Response Rates and Response Bias

The NPS response rate has tended to decline gradually over the past five years:

RESPONSE RATES

NPS Wave	W13	W14	W15	W16	W17	W18	W19	W20	W21	W22
Response	37%	36%	33%	36%	36%	33%	34%	32%	32%	30%

Percentage of questionnaires issued that are used in the analysis. Source: Continental Research.¹¹

The downward trend is relatively slow, and the achieved response rate is still remarkably high for a self-completion postal survey. However, the robustness of the survey will not benefit from any further decline in the response rate.

We suggest that great care is taken that changes or additions to the questionnaire do not discourage response (e.g. do not lengthen the questionnaire; try to avoid complicated questions and low interest level questions that may make the questionnaire appear longer or more daunting to the average passenger).

In addition efforts should be made to (a) shorten the questionnaire if possible and (b) explore possible ways to increase the response rate.

Variations in simple response rates (i.e. questionnaires received over questionnaires distributed) can be measured relatively easily in relation to any one of the parameters that are recorded when the questionnaire is issued.¹² These are:

- individual departure station
- date
- day of week
- time of day band (i.e. by shift)
- gender
- observed age
- journey purpose

Biases in any of the above parameters, if they arise from differential response rates, can be corrected by weighting, but in order to measure bias reliably it will be necessary to carry out some extra analysis. Some work of this type has been done in the past, but we believe it was occasional and very limited in scale.

We recommend that formal, detailed response analysis is carried out and reported on regularly (i.e. on every wave). If simple response biases become apparent (and some already are), in order to be confident that corrective weighting is both necessary and justified, it would be sensible also to examine the associated participation response rates (i.e. what proportion of people accepted a questionnaire when it was offered to them).

¹¹ For Waves 21 and 22 calculations we have included questionnaires arriving after the cut-off date. Wave 21 figures includes also 1093 questionnaires discarded as over quota. The percentages would be slightly higher if incomplete or invalid questionnaires were also included.

¹² All of these items appear on the NPS Respondent Record Forms supplied to interviewers. Numbers recorded can therefore be compared with the NPS sample numbers achieved and the response rate effects on profile, in any, can be calculated for any of these parameters.

It is known, for example, that the response rate for older passengers is higher than the response rate for younger people. At the same time, we have been told that, based on informal observation of NPS questionnaire distribution, the participation rate for younger passengers may be higher than for older people.

Without formal quantification, we do not know whether or not these age-related participation rate biases and response rate biases cancel each other out.

Our 2006 Report suggested that further work should be carried out on analysing response rates and measuring non-response bias and we believe this is even more important now.¹³

In addition, we recommend that a formal, quantitative ad hoc study is carried out to estimate the biases, if any, that may be arising from variations in participation rates. We are not aware that any such NPS study has ever been done on this subject.

Certainly, both the participation rates and the response rates should be examined more closely before any demographic or other response-rate-based weighting is considered.

¹³ As well as having the potential to improve the NPS sample in terms of the parameters listed above, evidence exists that suggests it could also improve the profile data re: journey purpose, by examining interlocking response rates by age and journey purpose (see NPS Review 2006, Section 3.8).

4. CONCLUSIONS & RECOMMENDATIONS

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

4.1 Gold Standard Uses, Key Objectives, Limitations

Objectives and Uses

Technically speaking the NPS is a trip-based transactional survey - a survey of rail journeys. Currently the NPS is carried out twice a year, over about 10 weeks in January, February and March ("spring") and about 10 weeks in September, October and November ("autumn") and covers journeys commencing after 6am 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, both overall and for each of a list of parameters (about 30) that have been identified as important contributors to satisfaction.

The results are 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 to questionnaire length, would be likely to prejudice the achievement of the "gold standard" goals.

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 shown to be representative, (ii) any changes to the design which are important or essential (such as those that may be judged 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.

It is to be expected that the survey will produce reliable and representative results for major journey subgroups: e.g. peak v. off-peak, commuter v. business v. leisure, station size band, weekday v. weekend, a.m. v. p.m. (or outbound v. return), and for major passenger types: e.g. gender, age, working status, etc. for each TOC individually and at the national level.

It has never been regarded as an objective that the NPS should be able to produce representative results for sub-TOC level areas (e.g. a city, a route or a station).

Limitations

The above statements do not differ in any material respect from the way the key objectives and uses of the NPS were described in the 2006 Review report.

However, it must be noted that, since 2006, the use of the NPS by stakeholders has become wider and noticeably more intensive. In particular, TOCs and others have increasingly sought to cross-analyse the NPS data to obtain results at sub-TOC levels.

The recent incorporation of the "building block" framework into the NPS sample design has allowed the NPS data (for medium- and large-sized TOCs) to be presented and used with moderate confidence at a "building block" level: typically this probably equates to a unit of about 400 or 500 NPS interviews, such as a group of routes or stations comprising say one third or one quarter of a medium or large TOC.

This development appears to have been well received by the TOCs who have taken advantage of it.

However, it probably represents the technical limit of the ability of the NPS (at its present scale) to produce reasonably reliable local data.

It is clear that many TOCs and some others would like the NPS to be able to "dive down" much deeper, e.g. to individual routes or even individual stations, and a good number of NPS users have already tried to do so through extra analysis - but (apart from a few exceptional cases¹⁴) this is well beyond the capacity of today's NPS.

¹⁴ Examples could be (i) a very simple TOC that consisted of, say, just two routes (each of which could possibly be treated as a building block if its NPS sample size is adequate), and (ii) a very large station or large city (but that could well require detailed adjustments to the sampling and/or weighting).

4.2 Self-Completion Method & Response Analysis

The standard 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 specific 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 and is also very cost-effective.

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.

A common criticism 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. However, the NPS response rate over the last ten waves has been between 30% and 36% (in terms of questionnaires processed divided by those handed out), which is high for such a long self-completion questionnaire for completion of which no incentive is offered.

However, at 30%, Wave 22 had the lowest response of the last 10 waves. We therefore recommend that efforts are made to reverse the decline. Specifically, we recommend that the questionnaire is not allowed to grow longer or more complex. In fact we suggest that it should be made shorter if possible. We recommend also that:

- detailed response analysis is published for every wave (currently annual)
- a detailed participation and response study is conducted for the first time
- the possible impact of incentives on response rate is explored¹⁵
- completed questionnaires should not be discarded if over quota¹⁶

We believe that if more thorough participation and response analyses are carried out, they could suggest ways in which NPS response rate, survey design and weighting could all be improved. This could 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 increase response or provide more representative results with less weighting, then both the actual and the effective sample size of the NPS will increase without any increase in actual fieldwork.

See also Sections 3.10 and 4.8.5.

¹⁵ Incentives for respondents cannot be discounted or free rail tickets (due to legal and MRS issues). In fact incentives should be totally neutral with regard to the survey topic (i.e. cannot even be railway-related) and also with regard to passenger profile (e.g. should not be biased toward young, or old, etc.). To be cost-effective it must have a very low cost per respondent. We imagine this must mean it will have to be a draw: e.g. an opportunity to win an M&S or Boots voucher. There could also possibly be a contractual incentive for the agency, i.e. a response rate based KPI.

¹⁶ 1093 questionnaires were discarded from Wave 21 as they were for TOC samples that were already well over target quota.

4.3 Distribution of Questionnaires

The standard NPS method of distributing questionnaires (with prepaid return envelopes) at stations prior to a train's departure, allows interviewers to distribute them to people travelling on a number of trains without the interviewer having to leave the station, hence there is less sample clustering than with on board distribution¹⁷. This is an efficient use of fieldwork time, especially at busier stations.

The method employed also has the merit of allowing respondents to complete the questionnaire at a time of their choosing when the journey is over, at work or at home, which is an advantage over e.g. on board collection, especially when dealing with crowded journeys and with long questionnaires.

We believe that the present NPS approach is satisfactory.

¹⁷ Some NPS questionnaires are distributed on board where sampling issues, cost-efficiency and other considerations are atypical, i.e. for Island Line (part of South West Trains) and non-franchised TOCs. This is acceptable.

4.4 Sampling Plan

The process employed by Continental Research for allocating the number of shifts to specific stations appears to be robust. We have no recommendations for specific changes to this aspect of the NPS design, nor to the method for allocation of shifts to days of the week.

One element of the approach to shift patterns that has improved since 2006 is that three-hour shifts are launched hourly throughout the day, rather than in simultaneous three-hour waves, as this is likely to generate a smoother hourly coverage across the hours during the main part of the day. However, this means that the first one or two hours each day (from 6am, as shift numbers build up) and the last one or two hours (up to 10pm, as shift numbers wind down) may have slightly lower numbers of shifts (and interviews) than will the other hours (8am to 8pm). Provided this is recognised, and has only a limited impact, we do not regard it as a problem, and do not consider it necessary to weight this parameter.

However, we are concerned that the distribution of shifts through the day as a whole appears still to be based principally on previous NPS results (i.e. Wave 9). Also, the weekday pattern of departures by hour through the day appears to differ significantly from available independent data (see Section 3.6).

In addition, although it has improved since 2006, the ratio of outbound to return journeys is still not fully balanced (see Section 3.7).

Clearly this area needs further refinement. We recommend that (a) NRTS and any other available external evidence should be examined re: journey volumes by time of day and day of week, in particular, the hourly volumes on weekdays, and (b) the ratio of outbound to return journeys should definitely be equalised overall.

The fact that the current outbound:return imbalance apparently has a surprisingly large (and directionally varying) incidence even at large stations is worrying and should also be further investigated.¹⁸

Finally, the cost-benefit of using more but shorter shifts at small stations should be evaluated (see Section 3.4.2).

¹⁸ We have been unable to investigate this issue in detail in the time available.

4.5 Weighting

4.5.1 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.

4.5.2 Journey Purpose

The present method of weighting journey purpose using ticket type data as a proxy should be reviewed once again. There is clearly a mismatch between these two variables. Although there are obvious implications for consistency and comparability, and for any TOC targets that may be affected, these problems could be overcome if the change is planned as a long term goal (see Section 3.9).

This remains the major issue re: potential changes to the weighting system. We have suggested elsewhere in this report that one should weight a survey to get 'accurate' results rather than inaccurate results. This is a function of any weighing scheme, where known defects in sample profile are corrected to get unbiased results that reflect the universe. If a 'better' weighting variable is known to be available, one should use it.

The 2008 PHA Report found that changes in data would occur when weighting by ticket type, and analysis of the effect was determined for Waves 9 to 16:

Differences if Weighting By Ticket Type

Source: NPS Waves 9 to 16, PHA analysis of CR data

	Total W9	Total W10	Total W11	Total W12	Total W13	Total W14	Total W15	Total W16
Sample size	29341	31289	25596	28074	28286	26388	25142	25334
OVERALL SATISFACTION	-2	-2	-3	-2	-2	-2	-2	-3
STATION - TICKET BUYING FACILITIES	-1	-2	-2	-1	-2	-3	-2	-2
STATION - PROVISION OF INFORMATION ABOUT TRAIN TIMES/ PLATFORMS	-1	-2	-2	-2	-1	-2	-1	-2
STATION - THE UPKEEP/ REPAIR OF THE STATION BUILDINGS/ PLATFORMS	-1	-1	-1	-1	-1	-2	-1	-1
STATION - CLEANLINESS	-1	-1	-1	-1	-1	-1	-1	-1
STATION - FACILITIES AND SERVICES	-1	-2	-1	-2	-2	-2	-2	-2
STATION - THE ATTITUDES AND HELPFULNESS OF THE STAFF	-2	-2	-1	-1	-2	-2	-2	-2
STATION - CONNECTIONS WITH OTHER FORMS OF PUBLIC TRANSPORT	-1	-1	-1	-1	-1	-2	-2	-2
STATION - FACILITIES FOR CAR PARKING	0	0	0	0	0	-1	0	0
STATION - THE OVERALL ENVIRONMENT	-1	-1	-1	-1	-1	-2	-1	-1
STATION - YOUR PERSONAL SECURITY WHILST USING	-1	-1	-1	-1	-1	-1	-1	-1
STATION - THE AVAILABILITY OF STAFF AT THE STATION	0	-1	-1	-1	-1	-1	-1	-2
STATION - HOW REQUEST TO STATION STAFF WAS HANDLED	-2	-2	-2	-1	-2	-2	-2	-3
TRAIN - THE FREQUENCY OF THE TRAINS ON THAT ROUTE	-1	-2	-2	-2	-2	-2	-2	-2
TRAIN - PUNCTUALITY/ RELIABILITY (I.E. THE TRAIN ARRIVING/ DEPARTING ON TIME)	-2	-3	-3	-2	-3	-3	-2	-3
TRAIN - THE LENGTH OF TIME THE JOURNEY WAS SCHEDULED TO TAKE (SPEED)	-1	-2	-2	-1	-2	-1	-1	-1
TRAIN - CONNECTIONS WITH OTHER TRAIN SERVICES	-1	-2	-2	-2	-2	-2	-2	-2
TRAIN - THE VALUE FOR MONEY FOR THE PRICE OF YOUR TICKET	-3	-3	-3	-3	-4	-3	-4	-4
TRAIN - CLEANLINESS OF THE TRAIN	0	-1	-1	-1	-2	-2	-1	-1
TRAIN - UP KEEP AND REPAIR OF THE TRAIN	-1	-1	-1	-1	-2	-2	-1	-2
TRAIN - THE PROVISION OF INFORMATION DURING THE JOURNEY	-2	-2	-2	-2	-2	-2	-1	-2
TRAIN - THE HELPFULNESS AND ATTITUDE OF STAFF ON TRAIN	-1	-1	-1	-1	-2	-2	-1	-1
TRAIN - THE SPACE FOR LUGGAGE	-1	-1	-1	-1	-2	-2	-1	-2
TRAIN - THE TOILET FACILITIES	-2	-1	-2	-1	-3	-2	-1	-2
TRAIN - SUFFICIENT ROOM FOR ALL THE PASSENGERS TO SIT/STAND	-2	-3	-4	-3	-4	-5	-4	-4
TRAIN - THE COMFORT OF THE SEATING AREA	-1	-1	-2	-2	-2	-2	-2	-3
TRAIN - THE EASE OF BEING ABLE TO GET ON AND OFF	-1	-2	-2	-1	-2	-2	-2	-2
TRAIN - YOUR PERSONAL SECURITY WHILST ON BOARD	0	-1	-1	0	-1	-1	-1	-1
TRAIN - THE CLEANLINESS OF THE INSIDE	0	-1	-1	-1	-2	-2	-1	-1
TRAIN - THE CLEANLINESS OF THE OUTSIDE	-1	-1	-2	-1	-2	-2	-1	-2
TRAIN - THE AVAILABILITY OF STAFF	0	0	-1	-1	-1	-1	-1	-1
TRAIN - HOW WELL TRAIN COMPANY DEALT WITH DELAYS	-2	-3	-3	-2	-3	-4	-3	-3

Most of the results are brought down by about 2 percentage points or so but there are variations by wave and measure. A general conclusion was reached at that time that the change should not be implemented.

One can see why. A discontinuity in the published figures would be difficult to explain and might result in a loss of confidence in previous results. Also, we understand that some franchise agreements contain NPS targets based on the current survey methodology and these would need modifying. A major reason for making the change is that the 2008 report data clearly showed that although the figures were different they both almost always had the same trend. Other major market research studies in Britain have faced this dilemma before and have implemented a change in survey methodology.

4.5.3 Day of Week /Time of Day

The weekday data from NRTS may help 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) may continue to be necessary and appropriate.

4.5.4 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.

The design factor effect on the national NPS results of the wide variations in TOC weights necessary to produce correct national figures is not significant (given the very large total national sample base).

The design factor effect of TOC volume weighting on individual TOC results is negligible, as it is applied to TOC totals, rather than within the TOC samples.¹⁹ It is of more importance to ensure that design effects within TOC samples are low.

4.5.5 Other Weighting Issues

It is theoretically possible to weight NPS data by stations or by routes. It is uncertain how much benefit could be gained by weighting individual small stations (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 are greater doubts about the practicality of weighting by route.

In practice the NPS is already weighted by station stratum within TOC, and is now also stratified and weighted by building block, which we regard as a satisfactory method given the current objectives of the NPS. The NPS is intended to produce representative data at the TOC level and the national level, and its design reflects that. The recent development of building blocks has created the possibility that the NPS can provide data at building block level, i.e. just one level below TOC level (for TOCs with a sufficiently large sample size).

¹⁹ Although possibly there may be minor effects due to building blocks or where TOCs share stations.

However, many users wish to analyse the NPS data well below building block level (e.g. route or station level) and indeed many have tried to do so through NPS database analysis. 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 certainly require a substantial increase in the number of local stations sampled per NPS wave, and a substantial increase in total NPS sample size, which would greatly increase cost.

4.6 Technical Documentation

As we stated in the report on the 2006 Review:

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 technical documentation of the NPS has improved markedly since 2006. The main source is the Overview Report, which now provides very much more technical detail than was publicly available in 2005, on such matters as sampling, weighting and response rates.

However, the Overview Report is produced only annually, and we believe that some details should be published with every wave, for example response rates (which have been declining and need to be monitored closely), and the questionnaire (some details of which have altered with every recent wave).

Also, some technical matters are still not sufficiently transparent, e.g. the way in which the shift pattern by hour of day, based on NPS Wave 9, was changed in Wave 17. The Overview Report merely states "approximately 100 shifts moved from the original morning peak time generated by the above procedure to an evening peak time" but no further details were provided.

4.7 NPS Reporting and Analysis

4.7.1 Regular NPS Reports

During the consultations, when discussing NPS reports, it was not always clear what document stakeholders were talking about (e.g. was it the Technical Overview, the Consultees Report, the TOC Report, or the one that is sometimes referred to as the "glossy" one?). This is certainly not a major issue, but it could be helpful if the "glossy" report had a definite name (e.g. NPS Headline Report)?

Stakeholder comments on the report layouts and contents were generally fairly positive or at least neutral overall. Our recommendation would be that the formats should not be changed much (and not very often - as the fact that they were in a stable format meant that people could find their way around them more easily).

At the same time there were some criticisms. Some said the reports should not contain any opinions or commentary, while others thought they should include a lot more "insight" or suggested action. Others felt that commentaries tended to concentrate on the negative results (i.e. criticising the TOCs) more than on the positive results.

One logical solution, which could perhaps be helpful, would be to make it clear that only one report (e.g. the "glossy" one) includes any opinions (which should be clearly identified as such), while all the others are impartial, factual "research documents" containing no opinions or selective comments.

Some said the reports were too detailed, others said that they were not detailed enough (but see paragraph at the top of the page). This probably reflects the fact that some users regularly need to explore the detail, while others are mainly interested in the headlines (or do want the details but can get them through their own internal reports, or Reportal, or through Quanvert or SPSS access to the database).

Key drivers analysis is a very useful NPS special analysis that is now available to all users. However, some stakeholders were unaware of this - or didn't know where to find it. It would perhaps be helpful if every NPS report contained a paragraph explaining what tabulations are available and in which NPS report they can be found.

The only other specific item of printed cross-analysis we wish to mention is the "TOC-type" analysis. No doubt many NPS users have grown used to it and some (but certainly not all of them) would perhaps be sad to see it go, but frankly it is now inappropriate and misleading.

It should be replaced by a more precise typology. Building block groups could provide a far superior substitute (provided exactly the same typology definitions are applied to all TOCs, which does not seem to be totally the case at present).

Another criticism was that reports do not appear until long after fieldwork is over, and seem not to be published until the next wave of fieldwork is about to start.

4.7.2 Occasional NPS Reports and Presentations

The level of knowledge regarding the methodology and scope of the NPS still varies widely. Some stakeholders were very well informed, others less so. This is unsurprising because the stakeholders range from market researchers who spend a high proportion of their time working on NPS data (or data from their own surveys), to others who are not researchers and refer to it only periodically.

Several stakeholders commented positively on the value of the occasional Passenger Focus reports on specific subjects (some based on NPS data, some from ad hoc research). The NPS contains such a vast amount of information that there must be very many more opportunities to explore the database in this way.

Users that had received NPS presentations from Passenger Focus or Continental Research commented that they had been useful.

Our only observation here is that increasing the level of knowledge of the NPS, through presentations, ad hoc reports, Passenger Link Manager contacts, the user's guide etc., seemed to raise the stakeholders' knowledge of the value and potential of the NPS - and also their understanding of its limitations, which is helpful too, as a significant proportion of the criticism received related to issues beyond the scope of the objectives or the design of the NPS (such as specific station data, specific route data, follow-up qualitative questions, extra questions re: TOC marketing issues, etc.).

4.7.3 Extra NPS Analysis

Far more users appear to be employing extra analysis than was the case in 2005/6.

One exception is that some TOCs do little NPS analysis because their own parallel survey has a much bigger sample size in their area, so they concentrate on that.

The ability of Passenger Focus personnel to provide extra NPS analysis rapidly on request was commented on spontaneously and very positively.

The Reportal facility received very mixed reviews. Some thought it was good and useful. Rather more complained about it, e.g. that cross-analysis options were more limited than they wished, or that it was fiddly, clunky and not user-friendly²⁰. Some had never used Reportal and a few had not heard of it.

The Verbatim facility also received mixed reactions. Some wanted more of it. Some thought it was too time-consuming to wade through it and wanted it all coded and analysed for them. Some had not heard of it.

Some stakeholders were not aware that it was possible to get a copy of the database, but the few who do, and analyse it themselves using Quanvert or SPSS, seemed very content with that (although they very probably want the NPS to have a much large sample size).

²⁰ Some enhancements were made in mid-2010, and some comments may relate to prior experiences.

4.8 Changes to the Questionnaire

NOTE: References in this section to NPS questions relate to Wave 23 (Autumn 2010) unless otherwise stated.

4.8.1 General Principles

Adding to the length or complexity of a questionnaire can reduce its response rate. Any changes to a questionnaire content, order, wording, interest level or appearance can potentially affect its response rate and/or can alter the actual responses received.

For a tracker survey such as the NPS a very conservative approach to all the above such changes or additions is essential.

For a survey for which the response rate is at its lowest for five years, there will be even greater concern regarding any such issues that may affect the response rate.

We recommend that efforts are made to increase the response rate, from the current 30%, to 40% (which we understand it has achieved in the past). This will require (a) definitely not allowing the questionnaire to become any longer or more complex, and (b) in our view, actually reducing its length by 5% to 10%.

4.8.2 Questions to Cut

There are very few regular questions that are not valued by most users. The most obvious candidate for removal is one of the three train cleanliness questions (overall, inside, and outside, Q22a line 1, plus Q23). The overall cleanliness question cannot be removed immediately as (we understand) it is included in one or more current contractual targets.

The end date of those contracts should be chosen as the date of the question's removal, which should be announced in advance now. Future contracts can use the Q23 inside/outside questions (which at that point should be moved to replace the deleted "overall" question in the Q22 grid, to make the questionnaire layout simpler).

If any extra groups of questions are added from time to time, they should be short. The one page section on security in Wave 22 (Q57 to Q64) coincided with the lowest NPS response rate yet. The section on security in Wave 23 is twice as long, (2 pages, Q40 to Q50, roughly 20% of all the questions) which is arguably much too long. In our view, sections of extra questions that long should not be accepted for the NPS.

The need to ask a long series of extra questions, especially on subjects that relate to a small proportion of passengers (e.g. those who have recently experienced a security problem on the railway, or who are disabled) can be accommodated by follow-up surveys (see section 4.8.6).

4.8.3 Questions to Change

Most of the regular questions on the NPS do not need to change. However, there are some that may need amendment. All the questions that relate to facilities that are not universally present can cause problems and need to be reconsidered.

The question re: on board toilets, for example, allows respondents only to rate the toilet as "very good, fairly good, neither good nor poor, fairly poor, very poor, or did not use/no opinion". It is hard to see why there is not an option allowing them to say that there was no toilet on the train (unless you wish to measure satisfaction with its absence - in which case the question should make that clear). This issue should be reviewed.

Similar issues arise re: other facilities on board (e.g. space for bicycles, staff other than the driver) or at the station (e.g. toilets, shops, cafés, catering, staff, ticket gates).

With all these, and other questions relating to events that do not always occur (e.g. did you ask for help, did you experience a delay) there is a further problem, which is that the base size responding will fall, and the questions need to be presented and interpreted in that light.

For example, only a proportion of NPS respondents may legitimately answer Q28: "How well do you think the train company dealt with the delay?". In the case of individual TOCs, the base size may be very low, but the scores are presented alongside all other "train factors II" (including toilet facilities) as if they are on the same basis, which clearly they are not.

Either these questions should be changed, or alternatively they should be presented in reports a different way, e.g. as a separate group which includes all the non-universal items and clearly shows the relevant base size and/or percentage applying.

Other questions which may possibly need examination are:

Q5: Is it understood that "outward" = from home (not from 'town' or 'terminus')?

Q12: Is it clear which information is provided and where?

Q16: Is it valuable to know familiarity with station?

Q34: Is it clear which ticket office this refers to?

Q35: What do they reply if it is never checked?

Q60: Why ask if they use internet "regularly"? Isn't access more important?

4.8.4 Questions to Add

Most NPS users think the NPS questionnaire is comprehensive. There were some individual suggestions for adding extra questions, but not many that had widespread support. Probably the one that is most justified is to add a question on rating the "train overall, ignoring the station" (to parallel the one on "station overall, ignoring the train"). This would be welcomed by TOCs (especially those who run very few stations).

One suggestion was to add post code of residence (this would very useful for catchment area studies and other types of traffic analysis: only the first 3 characters would be needed, and these could in fact be recorded on the NPS recruitment form instead of the NPS questionnaire). Another suggestion was to add train headcode.

4.8.5 Incentives

Increasing the response rate increases the probability that the results are representative. It could reduce the need for extra weighting. Importantly, it would also reduce the cost per interview of the survey²¹.

We recommend that the idea of response incentives is thoroughly explored. Something could be printed on the questionnaire along the lines of: "If we receive your completed questionnaire by (closing date) you could win a £100 voucher for...".

4.8.6 Follow-up Surveys

Long extra sections, especially those of minority interest, should be avoided in order to protect NPS response rates. An alternative way to accommodate these needs effectively would be to have a facility for follow-up surveys. We recommend that a suitable question is added at the end of the questionnaire, e.g.:

Would you be willing to take part in further surveys by Passenger Focus on subjects affecting railway passengers . . .

By telephone? (if yes please enter your phone number)

By post? (if yes please enter your postal address)

On-line? (if yes please enter your email address)

A small addition to the NPS questionnaire roughly along these lines could eliminate the need for many (or most?) of the extra questions that have been added to the NPS. Passenger Focus could accumulate a large database of willing respondents that would be very valuable when surveys of minority groups were planned. A further advantage of this approach is that the cost of all follow-up surveys could be precisely identified.

4.8.7 Semantic Scales

The RMA 2006 Review was asked to evaluate question scales and did so at length. In brief: Semantic satisfaction scales are probably the most commonly used approach to measuring customer satisfaction, and that is the format used on the NPS satisfaction questions. The NPS employs a very widely used 5-point version i.e.:

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

²¹ We estimate that an increase in the response rate from the current 30% to 40% would increase the NPS sample base from about 27,000 per wave to about 36,000, and at a marginal cost (our rough estimate is that this 33% increase in sample size would increase costs by less than 10%)

Semantic evaluation scales are another common method, and this format is used on the NPS rating questions. The NPS employs a widely used 5-point version i.e.:

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

Technically, the NPS fulfils a requirement for trip-based transactional research. The satisfaction scales and evaluative scales employed on the NPS are appropriate for this purpose. Both scale types are very widely used, and both are generally regarded as effective.

The most important issue from the point of view of measuring service delivery is the trend data. Consistency of approach is essential when tracking change. Clearly, the NPS should continue to use its existing scales.

However, a few may need an extra "tick box" (see Section 4.8.3).

4.8.8 Question Testing

We understand that pilot testing on some NPS questions was carried out in the past using qualitative or small scale quantitative studies. This is helpful, but is unlikely to provide statistically reliable evidence of the effect of changes to questions.

If a change is planned to the wording or precode list of a regular NPS question (such as those suggested at 4.8.3), it is possible also to check the impact of the change by carrying out a split run. Half the next NPS Wave questionnaires could use the old question, and half the new, providing a split run matched sample. The results could be compared in detail statistically, and uncertainties regarding interpretation etc. could be resolved.

4.9 Other Topics

4.9.1 Validation Work

The NPS is an important, large scale and complex survey. We recommend that it would be sensible to set aside regularly a proportion of the annual cost, say 3%, to spend specifically on technical validation work.

This could include for example the quantitative monitoring of NPS participation rates and response rates that we believe now urgently needs to be done (see Section 4.2).

We believe this relatively small allocation of funds is essential to ensure the continued reliability of the NPS data and confidence in the survey's robustness.

4.9.2 Special Reports

The NPS database contains a vast amount of information. We believe that it would be possible to get more value from the existing data by carrying out extra analysis on the database in order to produce a series of published reports on particular subjects of interest to the rail industry. We believe there could be many such topics, and suggest that a budget is set aside for this purpose.

We do not in any way wish to exclude the possibility of carrying out ad hoc research (such as the recent work on TVMs) which we believe should also be budgetted to continue. However, the cost of extra analysis is much less than for ad hoc fieldwork, and we wish to draw attention to the enormous scope that exists for special reports using the existing database.

4.9.3 Boosted TOC Samples

The one thing that most NPS users seemed to agree on was that they would like a larger sample size, in order to look at individual routes or stations, or to boost coverage of small stations. As we have indicated, this would probably require an NPS sample size several times larger than it is at present and we seriously doubt that this could be considered given present government priorities.

However, it could be the case that some TOCs would be prepared to contribute funds to be used specifically for boosting the sample size in their area above the level currently set by Passenger Focus. We think some would be very unlikely to do so, but believe that some would give it serious consideration.

This could result in sample sizes being much larger for TOC "A" than TOC "B" and weighting would be adjusted so that the national NPS results were correct. This would in turn increase design effects on the national sample, but as it would also increase the already large national total sample size, and would have no adverse impact on the effective sample size for any individual TOC, this is not a concern.

A note of caution is necessary here: this process should not allow TOCs to add to (or alter) the NPS questionnaire in any way. There is clear evidence that doing so would adversely affect the response rate and/or the results.

4.10 Summary of Recommendations

Some of the key recommendations we have made in this report are:

4.10.1 The NPS response rate must be increased

The use of response incentives should be explored (see Sections 3.10, 4.2, 4.8.5).

4.10.2 Profile by hour of day must be checked

The profile of departures by hour of day (on weekdays) differs from external data. This must be checked and if necessary corrected (Sections 3.6, 4.4, 4.5.3).

4.10.3 The outward: return ratio should be balanced

The ratio of outward journeys to return journeys is closer to reality than it was in 2006, but it is apparent that it is not yet fully balanced. This issue should be checked and adjusted as necessary (Sections 3.7, 4.4).

4.10.4 TOC-type cross-analysis should be dropped

TOC-type cross-analysis is invalid and should be dropped. It should be replaced by a consistent approach to building blocks (Sections 3.8, 4.7.1).

4.10.5 Follow-up survey facility

A follow-up survey facility should be developed. This could be a valuable asset, particularly if Passenger Focus wishes to research minority groups (Section 4.8.6).

4.10.6 NPS questionnaire length must be reduced

In order to assist the response rate, the NPS questionnaire should definitely not get any longer. In fact we think it should be shorter. Extra questions on e.g. security could use the follow-up facility above, not the NPS (Sections 4.8.1, 4.8.2, 4.8.6).

4.10.7 NPS Technical Documentation

This has improved greatly since 2006, but more detail is still needed (Section 4.6).

4.10.8 TOC target sample sizes

These should be reviewed and harmonised (Section 3.4).

4.10.9 Journey Purpose v Ticket Type Weighting

This may seem somewhat arcane, and may be difficult to achieve, but it is hard to avoid the conclusion that it should be kept in mind, perhaps only as a long term goal, that the current weighting could be made more accurate (Sections 3.2.2, 3.9, 4.5).

###

APPENDICES

Appendix A: The Project Team

Appendix B: The Review Method

Appendix C: The Consultees

Appendix D: The Discussion Guide

Appendix E: The Documents Reviewed

Appendix A - The Project Team

The Project Team comprised:

Peter Bartram, FMRS

Paul Harris, CStat, FMRS

Richard Roberts-Miller, FMRS

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

Peter Bartram, FMRS - PB Consulting

A senior marketing research consultant, and former Chairman of the Market Research Society, 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, 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 of Applied Research & Communications, and other senior positions with NOP, City Research Group, American Express, and Thomson Newspapers. A former Chairman of the Market Research Society, visiting lecturer at City University Cass Business School, and convenor of many seminars and educational workshops. Author of many papers and textbook contributions on research and its applications.

Contact address:

1 Ashurst Road, Tadworth, Surrey KT20 5ET

Tel: 01737-812983

Mobile: 07880-797160

peter.bartram1@btinternet.com

Paul Harris, CStat, FMRS

Paul Harris runs his own statistical consultancy and is a consultant to a limited number of market research agencies. He had been the Chief Statistician of NOP Market Research Ltd for twenty six years. He was formerly with the Market Research Department of the Electricity Council and before that with the Central Electricity Generating Board, working on power station statistics.

His interests include sampling methodology and the application of multivariate statistical methods to market research data as well as the use of computers in statistical analysis. He is widely experienced in the design and analysis of segmentation studies and has much experience in trade-off and pricing studies.

Among his published works is a Market Research Society Silver Medal paper on the effects of clustering on random sample surveys. He recently was awarded an MRS Conference Technical prize for a joint author paper on Data Fusion. This new important topic deals with merging data from two separate surveys to form one database.

He lectures widely on various topics for the MRS and other bodies and has contributed chapters to two published handbooks of market research. He is a Chartered Statistician of the Royal Statistical Society (the Society's highest professional award) and a Fellow of the Market Research Society. In 1996 he was awarded a Market Research Society Gold Medal for exceptional contributions to market research and the MRS over many years.

Paul Harris Associates
4 Pinehurst
Sevenoaks
Kent
TN14 5AQ

Tel. 01732 763297
E-mail paulharris@aol.com

Richard Roberts-Miller, FMRS



Richard Roberts-Miller established RMA in 1989, as a management, marketing and research consultancy specialising in travel, transport and tourism. Since then, RMA has carried out a wide range of research and consultancy projects for national tourist boards, tour operators, airline, cruise line and railway companies, hotel groups, cottage and villa companies, as well as many projects in other sectors such as finance and education. RMA projects most analogous to the current proposal were reviews of the NPS, the BPSS and the BMTS (carried out for Passenger Focus) and a review of the UKTS carried out for its sponsors (ETB, NITB, STB, WTB & DCMS). Roberts-Miller's background includes working on other major continuous surveys (BNTS in travel, HBI/HBS in tourism, and the NRS in media, plus the single-source TGI). He has extensive experience of customer satisfaction monitoring (including developing the Thomson "CSQ" system which has been widely imitated and is almost the de facto standard for tour operators). RMA projects have included:

- quantitative and qualitative surveys
- consumer research surveys
- business to business research surveys
- international research surveys
- customer satisfaction surveys
- research design consultancy
- technical evaluation of surveys
- branding, advertising and brochure research
- marketing development studies
- product launches and re-launches
- management studies
- takeover and merger projects
- investment project evaluation studies

Richard 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 market research 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 on their Tourism Advisory Council, and was presented with the "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), Thomson Organisation (group research manager), Thomson Travel (research and planning manager), Thomson Holidays (marketing controller), Thomson Travel (board member), Thomson Vacations Inc. (President & CEO), Thomson Travel Inc. (North American President & CEO) and International Thomson (US) Inc. (Executive Vice President).

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
Stakeholder Review
Technical Review
Conclusions and Recommendations

The project was co-ordinated by Richard Roberts-Miller, acting as project manager. Work was shared between the project team members as described below.

Study of Documentation

At the beginning of the project we were provided by Passenger Focus with a range of documentation relating to the NPS, and further documentation was provided by Continental Research. This material was reviewed by members of the Project Team. This in turn led us to request further information from PF, CR and others, and to carry out extra analysis of NPS data. This work was not reported separately but contributed to the evaluation of the Consultation Exercise (Section 2), to the work done in the Technical Review (Section 3) and to our Conclusions & Recommendations (Section 4). For examples of documentation reviewed see Appendix E.

Stakeholder Review

In October 2010 Passenger Focus provided us with a list of 31 NPS stakeholder organisations (21 TOCs and 10 other organisations) together with a contact name (or names) at each. Each organisation was advised in writing by Passenger Focus that they would be approached by RMA for the purposes of arranging a consultation on the subject of the NPS. We then sought to make appointments for face-to-face or telephone consultations (at a time of the consultee's choosing) with representatives of each of the organisations.

We were able to obtain feedback from representatives of 30 of the 31 organisations (one declined to participate as he felt his views would already be represented adequately by other stakeholders). All consultees were assured that their comments would remain anonymous (it was explained that some of their remarks might be quoted in our report for illustrative purposes, but only on an anonymous and unattributable basis).²²

Most consultations were with individuals but some were with groups of two to five people. In most cases the consultation was with the named individual(s), but in a few cases the organisations suggested that different (or additional) individuals participate. All consultees agreed that their sessions could be recorded (for our analysis purposes only), and all agreed to be named in this report as having participated.

²² This approach was not chosen because the topic was thought to be contentious (it was not expected to be), but because our experience is that this approach is more likely to encourage a frank and easy exchange of views.

In total, 14 consultations were conducted face-to-face (in the consultees' offices) and 15 were conducted by telephone. All 29 were conducted by Richard Roberts-Miller, using the discussion guide shown in Appendix D, between October 11th and November 12th, 2010. In one more case a convenient appointment could not be arranged in the time allotted, but the consultee provided written feedback which was included in the material reviewed.

In total, these formal consultations generated feedback from 44 individuals at 30 stakeholder organisations (21 TOCs and 9 others).

In addition, a further 24 stakeholder organisations (4 TOCs and 20 others) were invited to submit written feedback regarding the NPS (although relatively few did so).

All 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 the consultees, plus a list of the 24 other stakeholder organisations invited to provide feedback, please see Appendix C.

The Technical Review

The main part of the analysis and review of the sampling and weighting was carried out by Paul Harris, whose input formed the basis of Section 3 of this document.

Conclusions and Recommendations

All three members of the project team evaluated the draft report on the Stakeholder Review and the draft report on the Technical Review and agreed the final versions of the two reports, and then produced an agreed series of Conclusions and Recommendations.

The end products of the review process are included in this document as follows:

Stakeholder Review - see Section 2

Technical Review - see Section 3

Conclusions and Recommendations - see Section 4

APPENDIX C - The Consultees

We consulted the following 44 individuals at these 30 organisations, who provided feedback at face-to-face meetings (F) or in telephone discussions (T), or in one case in written form (W).

TOCs

Arriva Trains Wales, Vanessa Schotes, Head of Marketing (T)
Arriva Trains Wales, Chris Williams, Yield & Data Analyst (T)
Association of Train Operating Companies (ATOC), Adrian Chapman, Market Research Manager (F)
Association of Train Operating Companies (ATOC), Billy Denyer, Head of Business Analysis (F)
Association of Train Operating Companies (ATOC), Edward Welsh, Director of Corporate Affairs (F)
c2c, Kim Gorman, Customer Service Delivery Manager (W)
Chiltern Railways, Emma Palmer, Marketing Development Manager (F)
crossCountry (XC Trains Ltd), Richard Gibson, Head of Communications (T)
East Coast, Jim Muir, Insight Manager (T)
East Midlands Trains, Stuart Haywood, Business Excellence Manager (T)
First Capital Connect, David Burns, Senior Commercial Manager (T)
First Great Western, Jeremy Clarke, Senior Insights Manager (T)
First ScotRail Ltd, Kenny McAlpine, Research & Pricing Manager (T)
First Transpennine Express, Louise Ebbs, Strategic Planning Manager (T)
Grand Central, David Crocker, Sales & Marketing Manager (T)
London Midland, Bob Haywood-Lister, Revenue Analysis Manager (T)
London Overground Rail Operations Ltd (LOROL), Mark Eaton, Concession Director (F)
London Overground Rail Operations Ltd (LOROL), Gerry Kirk, Head of Customer Service Delivery (F)
Merseyrail, David Print, Head of Strategic Development & Concession Management (T)
Northern Rail, Peter Myers, Head of Service Quality (T)
NXEA, Clive Morris, Head of Customer Service - South East (F)
South West Trains, Ian Johnston, Customer Services Director (F)
Southeastern, Sarah Boundy, Head of Communications & Publicity (F)
Southern, Emma Tomes, Head of Customer Experience (F)
Virgin Trains, Andrew Crump, Insight Manager (F)

OTHER STAKEHOLDERS

British Transport Police, Nic Pole, Quality of Service Research Manager (F)
CDL, Paul Cooper, Head of Transport Planning (F)
Department for Transport (DfT), David Hibbs, Rail Passenger Policy Manager & Passenger Focus sponsor (F)
Department for Transport (DfT), Rachel Kelley, Franchise Policy and Programme Manager, Rail Specification Team (F)
Department for Transport (DfT), Margaret Shaw, Rail Statistician & DfT Representative on NPS Stakeholder Board (F)
Department for Transport (DfT), Peter West, Franchise Manager, Rail Contracts Directorate (F)
GMPTE, Sally Holgate, Research & Intelligence Officer (T)
GMPTE, Tom Sansom, Consultation & Research Officer* (T)
Network Rail, Pete Allen, Senior Market Research Specialist (F)
Network Rail, Peter Collins, Operations Development Manager (Stations) (F)
Network Rail, Bill Davidson, Strategic Planning Manager (F)
Network Rail, Mike Goggin, Head of Stations & Customer Service (F)
Network Rail, Andrew Regan, Research Specialist (F)
Office of Rail Regulation (ORR), Annette Eggington, Head of Competition & Consumer, Economics Directorate (F)
Office of Rail Regulation (ORR), Nigel Fisher, Head of Information & Analysis, Rail Planning & Performance Directorate** (F)
Transport for London (TfL), Carol Smales, Principal Transport Economist - London Rail (F)
Transport for London (TfL), Alan Smart, Senior Planner (F)
Transport Scotland, Allan Anderson, Rail Passenger Interest Manager (T)
Welsh Assembly Government, Andrew Franklin, Rail Franchise Manager (T)

* Also acting as *pteg* Liaison with Passenger Focus on the NPS Review Project.

** Also ORR Representative on the NPS Stakeholders' Group.

In addition, written feedback was invited from the following 24 other stakeholder organisations:

TOCs

First Hull Trains
Heathrow Connect
Heathrow Express
Wrexham and Shropshire

Other Stakeholders

ARUP
Centro
LEK Consulting
Line by Line Ltd
London Travelwatch
London Underground Ltd
Merseytravel
Mott Macdonald
Napier University
National Audit Office

Nexus

pteg

Rail Safety & Standards Board
South West Observatory
South Yorkshire PTE
Steer Davies Gleave
Strathclyde Passenger Transport
University of Nottingham
University of West of England
West Yorkshire PTE

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:

NPS Stakeholder Review 2010 - Discussion Guide

f2f / tel

Name of consultee(s): _____

Job title(s): _____

Organization: _____

Date: _____ Interviewer: _____

Explain: Any comments you make will be strictly confidential. We may wish to quote some remarks people make but this will be for illustrative purposes only: any remarks we report will be completely anonymous and unattributable.

May I record our conversation? This is for my own note-taking purposes only.

YES – agree NO – do not agree

May I list you in our report as having participated in the consultation process.

YES – agree NO – do not agree

NPS USAGE

How does your organisation use the results of the NPS?

Which of the NPS results do you think are the most valuable?

How useful are these results to you/your org?

Do you think the NPS helps to drive up standards of performance for passengers?

Are there any of the NPS results you do not look at/use at all?

Is the NPS missing any questions that you would like to see included in future?

Do you look at any other rail passenger surveys - apart from the NPS?

What would you like to use NPS for that you currently cannot?

NPS METHODOLOGY

How familiar are you with NPS methodology and design?

Do you have any comments on the survey method or design?

CHECK ON:

Accuracy/reliability/credibility of results generally

Self-completion method

Questionnaire content and design

Sample design

Building block approach

Weighting method

Fieldwork frequency, timing, seasonality, duration

Do you think the NPS methodology needs to change in future?

NPS OUTPUT

Are you happy with the way the NPS results are reported?

CHECK ON:

Report format

Report frequency

Report turnaround time

Report coverage (scope of topics reported)

Level of detail of results reported/cross-analysis categories

What other analyses would you like to see?

PF is considering presenting more building block analysis in future – any comments?

How would you like to see the NPS reporting evolve in future?

Do you use Reportal/Verbatim database? How much/what for?

OTHER COMMENTS

How do rail industry stakeholders regard the NPS generally?

How do you think the NPS could be made more valuable to the industry?

How do you think the NPS could do more to raise standards for the passengers?

Do you have any other comments on the NPS?

THANK YOU FOR YOUR CO-OPERATION

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 Autumn 2009 (Wave 21 "glossy" report), PF, 2010.
NPS Spring 2010 (Wave 22 "glossy" report), PF, 2010.
NPS Autumn 2009 (Wave 21), TOC Reports (various), PF, 2010.
NPS Spring 2010 (Wave 22), TOC Reports (various), PF, 2010.
NPS Consultees Report - Spring 2010, Wave 22, PF/CR, 2010.
NPS Detailed Technical Overview Report, Waves 21 & 22, PF/CR, 2010.
NPS User Guidance Report Updated October 2010, CR, 2010.

Examples of NPS questionnaires and fieldwork documentation:

NPS Questionnaire Wave 22, CR, 2010.
NPS Questionnaire Wave 23, CR, 2010.
NPS Questionnaire Showcard X - Journey Purpose Definition, Wave 23, CR, 2010.
NPS Questionnaire Distribution Instructions, Wave 23, CR, 2010.
NPS Respondent Record Form, Wave 23, CR, 2010.
NPS Interviewer Attendance Sheet, Wave 23, CR, 2010.
NPS Interviewer Shift Schedule, Wave 23, CR, 2010.
NPS Shift Alteration Form, Wave 23, CR, 2010.

Other related documents:

National Passenger Survey - Report on a Research Review, RMA, November 2005.
Published as: Findings of a Review of the National Passenger Survey, PF/RMA, February 2006.
National Passenger Survey - Report on Weighting, PHA, April 2008.
Published as: National Passenger Survey - Report on Weighting, PF/PHA, August 2008.
Statement of compliance with the Pre-Release Access to Official Statistics Order 2008, PF, 2009.
ATOC NPS Summary report, Wave 22, ATOC, 2010.
PF Statistics Governance Group Paper SGG.10, PF, 2009.
Examining Links Between Train Operation and Customer Satisfaction, CDL, 2009.

In addition, analysis was carried out on the following data:

NPS Wave 22 (Spring 2010) database, CR, 2010.

Key:

ATOC = Association of Train Operating Companies
CDL = Collinson Dutton Ltd
CR = Continental Research
NPS = National Passenger Survey
PHA = Paul Harris Associates
RMA = Roberts-Miller Associates