Rail Passenger Assistance: Benchmarking Train Operators' Performance
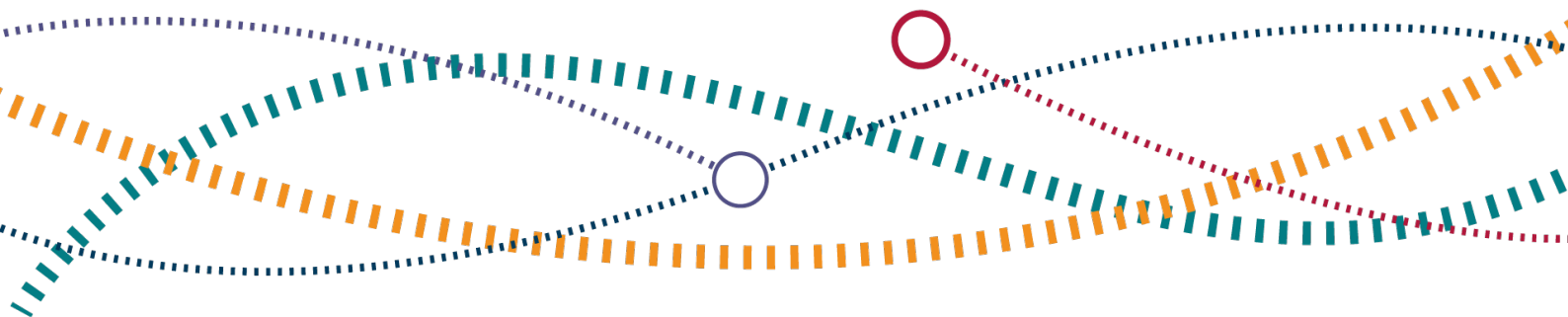
Revised framework and summary of consultation responses

11 July 2025

Contents

Executive summary

ORR is introducing a new framework for benchmarking the delivery of passenger assistance by train and station operators. This will strengthen our ability to hold operators to account for poor performance in this area and drive improvements in service provision.

This document summarises responses to a consultation on our initial benchmarking proposals which took place earlier this year and, having carefully considered the feedback we received, sets out our revised approach.

We will use quantitative data to rank operators' performance on their delivery of assistance, with metrics on reliability, passenger satisfaction, and staff training. Data will be drawn from our ongoing passenger survey and, as proposed by a number of consultation respondents, we will continue to explore other potential sources of data that could add additional insight.

We will then qualitatively assess the poorest performing operators on their capability to improve, including considering how they monitor their delivery of assistance, identify risks that affect performance and develop appropriate solutions, deliver staff training and engage with passengers and industry colleagues.

We had proposed to place operators into one of three tiers for delivery of assistance (upper/middle/lower). These tiers were quite narrow, reflecting the relatively small variation in performance between operators. We have concluded that it is not reasonable to place operators into such tiers, given the level of statistical variation in the underlying data. We consider that a relative performance ranking is a fairer approach that will still highlight both weaker and stronger performance and incentivise improvements.

We had proposed assessing all operators on their capability to improve. Taking into account feedback from consultation respondents, we have decided to focus our efforts where improvements are needed the most. This will reduce the regulatory burden on better performing operators.

Later this year we will publish our first benchmarking report. As well as the outcomes of our quantitative and qualitative assessments, our report will seek to share good practice from across the industry to help drive improvements in assistance delivery.

1. Introduction

1.1 We want to empower confident travel by all, including disabled passengers, whether their journeys are made independently or with assistance.

1.2 Our Annual Consumer Report sets out our recent work to improve the delivery of assistance to rail passengers, the complaints and redress processes for disabled passengers, and industry performance in key areas that contribute to the passenger experience.

1.3 To help us drive improvements, we sought feedback on a new framework to benchmark the performance of train and station operators' delivery of assistance. Our Rail Passenger Assistance: Benchmarking Train Operators' Performance consultation ran from 11 December 2024 to 14 February 2025.

1.4 This document provides a summary of responses to the consultation on key issues, our response to the feedback we received (Section 2), a description of the revised framework we will put into effect later this year (Section 3), and our Next Steps (Section 4). A more detailed question-by-question overview of responses is provided at Annex A. Revised qualitative performance descriptors are set out in Annex B.

1.5 An analysis of each proposal put to us is provided in the accompanying Equality Impact Assessment.

Background

1.6 In accordance with their licence obligations, every train and station operator must, as part of an Accessible Travel Policy approved by ORR, commit to providing assistance to rail passengers that need it. Failures in this service can leave passengers feeling angry, frustrated and helpless. Our passenger research suggests that satisfaction with assistance where it is delivered is high, but that service reliability is not good enough. In 2024-25, only 78% of respondents to our passenger survey reported receiving all the assistance they had booked; 11% of respondents to the survey reported that they received none of the assistance they had booked.

1.7 As part of our role in protecting rail passengers, we proposed to introduce a new benchmarking framework to:

- strengthen our ability to hold operators to account for poor performance in this delivering assistance; and

- drive improvements in service provision.

1.8 Our proposed framework for benchmarking each operator's performance focused specifically on two areas: assistance delivery and organisational capability to improve. We planned to commence the framework later in 2025 when we would publish an annual assessment that ranked the performance of each operator across a range of categories.

1.9 Our consultation document set out our proposals in detail and included questions on each aspect of the framework and how we would measure performance.

Responses to the consultation

1.10 We received 33 responses to our consultation.

1.11 We would like to thank all respondents for their helpful and constructive comments. Every response has been read and considered carefully. This document provides a summary of the key points raised by respondents, and how these were taken into account in the revised framework.

1.12 In addition to the written responses received, we have had ongoing dialogue with the Rail Delivery Group (RDG), various operators and passenger groups, including those representing disabled passengers. We have twice engaged with disabled passenger groups through our Accessible Travel Stakeholder Forum.

Public Sector Equality Duty

1.13 Under the Equality Act 2010, in carrying out our activities the ORR has a duty to eliminate discrimination, advance equality of opportunity and foster good relations between persons who share a relevant protected characteristic and persons who do not share it. The relevant [this excludes marriage and civil partnership] protected characteristics are age, disability, gender reassignment, pregnancy and maternity, race, religion or belief, sex and sexual orientation. An Equality Impact Assessment of the introduction of this benchmarking framework is provided at **Annex C** in accordance with our Public Sector Equality Duty under the Equality Act 2010.

Alternative formats

1.14   ORR has actively considered the needs of blind and partially sighted people who are accessing this document in PDF format. The text is available in full on the ORR website, and can be freely downloaded. Individuals and organisations can use free Adobe Reader accessibility features or screen readers to read the contents of this document.

1.15   Easy Read and Large Print versions of this document are available on our website. If you need conversion into any other formats, please contact us at: ATP@orr.gov.uk.

2. Summary of consultation responses

2.1   This section gives a brief overview of responses to the consultation on key issues and describes how we have taken them into account in our revised framework.

Overview

2.2   Respondents were in favour of ORR publishing a report that benchmarks operators' performance in delivering assistance, highlighting how this would make it easier to identify issues and areas for improvement as well as celebrate success. Respondents also supported a broad focus on two key areas of assessment: delivery of assistance, and the capability to improve. A range of suggestions and concerns were raised around the specific metrics to be used and the sources of data and information to inform the assessment.

2.3   We will proceed with plans to benchmark operators' performance in delivering assistance, against the two categories of Delivery and Capability to improve. We have made a number of changes to our approach in response to feedback received through the consultation, which are described below. We expect our approach to continue to evolve over time where new sources of data become available.

A. Delivery Assessment Area

2.4   We proposed three sub-categories to capture the extent to which an operator is meeting the basic requirements of assistance delivery:

- **Reliability** - are passengers receiving the assistance that they have requested?

- **Satisfaction** - are they happy with the assistance that they get?

- **Training** - are staff receiving the appropriate training?

Reliability and Satisfaction Source Data

2.5   We proposed that, at least initially, we would use data from our ongoing survey of booked assistance users to assess reliability of and satisfaction with assistance. A number of respondents raised concerns about the representativeness and size of

the survey's samples at the operator level. Having done further analysis, we remain satisfied that the survey is of sufficient size (with around 9000 responses) and accuracy (with only small margins of error of +/-2 percentage points when measuring the performance of most operators) to give a representative picture of passengers' experiences of booked assistance. Nonetheless, we are making changes to the survey that are intended to improve the response rate, increase sample sizes and provide even more confidence and accuracy in our performance benchmarking.

2.6   At the 22 stations included in Network Rail's licence, we will collect data that will enable us to attribute results to the operator that delivers assistance (Network Rail provides assistance at only 10 of these stations). We will benchmark operators' performance based on this updated allocation of survey responses to better reflect how well assistance is being delivered by both Network Rail and train operators.

Ranking Operators

2.7   We proposed that our assessment should place operators in one of three performance tiers. Respondents on behalf of passengers and disabled people were keen for operators to be set challenging performance targets, while industry respondents were concerned our proposed tier thresholds were too challenging.

2.8   We have looked carefully at our approach again, focussing on how confidently we could place operators into tiers. There was a 5 percentage point difference between our proposed performance tiers (for example, fewer than 5% of passengers reporting that they received none of their booked assistance for top tier, 6-10% for middle tier, above 11% for lower tier), reflecting the relatively small degree of variance in performance between operators and our expectation that operators should be aiming to drive improvements. As noted above, the margin of error is typically +/-2 percentage points for each operator. We have concluded that it is not reasonable to place operators into such narrow tiers with this level of statistical variation in the underlying data. We considered widening the tiers. However, that could only reasonably be achieved by lowering the thresholds which we do not consider to be consistent with our expectations that operators should be aiming to eliminate assistance failures.

2.9   Our benchmarking report will instead set out a league table ranking relative performance for delivery of assistance, calculated from performance on reliability, satisfaction and training. Our revised approach is explained in more detail in Section 3.

Additional Reliability and Satisfaction Data Sources

2.10   Respondents suggested various additional sources of performance data. We expect to introduce a new measure of confidence to travel from 2026, which Transport Focus has advised may provide additional insight into passengers' experiences and indicate greater variance in operator performance.

2.11   We are also exploring with RDG how we might get a better picture of the experience of passengers that request assistance on demand, known as Turn up and Go (TUAG), including from passenger feedback provided through the Passenger Assist app.

Assessment of Staff Training

2.12   To enable benchmarking of the training provided to staff, we proposed a qualitative assessment of training materials and content, alongside a regular measurement of the percentage of eligible staff who have received either induction training or refresher training within the past two years. Respondents agreed training is an important area for ORR to focus on but differed in their suggestions for how and to what extent this might be best done via the benchmarking framework.

2.13   On the one hand, a qualitative assessment of training materials and delivery was considered important; equally, some respondents preferred a more quantitative approach focusing on outcome measures of good training such as passengers' experiences of assistance.

2.14   We have decided to retain both aspects. Our 'Delivery' Assessment area will now include a measurement of staff knowledge and proficiency, using an existing Passenger Assist survey question which means it can be included in the 2025 assessment. Alongside this, we will add a quantitative measure of training delivery from 2026. The qualitative assessment of training will now be included in our revised approach to the 'Capability to Improve' assessment area (see below).

B. Capability to Improve Assessment Area

2.15   We proposed an assessment of the documentary evidence of each operator's processes across three sub-categories that focus on organisational capability to manage and improve performance:

- Monitoring & Reporting: does the operator monitor delivery of assistance effectively?

- Risks & Mitigations: has it identified the risks that affect its performance and developed appropriate solutions?

- Engagement: does it effectively engage with passengers and industry through this process?

Assessment Content

2.16   We received a number of suggestions for further sub-categories in this area. These included complaints handling, innovation, interchange and collaboration. We also received several suggestions to change the name of the area to 'Continuous Improvement', to avoid implications of universal poor performance, as well as detailed suggestions for changes to the assessment criteria.

Regulatory Burden

2.17   More broadly, industry respondents were concerned about the additional reporting burden imposed by the information requests we would make to provide the evidence for our assessment in this area. We received a suggestion to instead focus on best practice as an improvement tool. We recognise the importance of ensuring that any regulatory burden is proportionate.

2.18   We will therefore carry out a qualitative assessment on the Capability to Improve only for those operators we have most concerns about, as identified by our assessment of Delivery. We have decided to retain the focus of this assessment on the three sub-categories proposed in the consultation (together with Training, which we proposed to include within Delivery). We have updated the assessment criteria in the light of respondents' comments. This focus on operators that need to improve means we retain the 'Capability to Improve' title. We explain how this will work, including any next steps following the initial qualitative assessment, in Section 3.

Best Practice

2.19   We also recognise the value of sharing and implementing good practice in driving service improvements, and so we intend to use the benchmarking report to highlight best practice across the industry across the areas covered by the Capability to Improve sub-categories.

3. Revised benchmarking framework

Introduction

3.1   Our revised benchmarking framework will consist of an annual assessment of performance in two areas:

    A. **The delivery of assistance** by every passenger train operator that is responsible for providing assistance at stations, as well as Network Rail. This will be a quantitative assessment, based on data we collect on the quality and reliability of assistance they provide at stations on the national rail network; and

    B. **The capability to improve.** This will be a qualitative assessment, focusing on those operators that require the most improvement in the delivery of assistance. In undertaking our assessment of the delivery of assistance, we will identify a small number of operators we are most concerned about. We expect this number to be no more than five, but the exact number will depend on a number of factors, including relative performance and an assessment of the potential impact on passengers.

3.2   We will publish our assessment every year and use the findings to inform our engagement with operators. This will focus on the worst performers, as expanded upon below.

3.3   Notwithstanding our annual assessment, we will continue to monitor industry performance in these areas throughout the year. Where we have serious concerns about this performance we will continue to take this up with operators, at any time. This is in addition to our broader monitoring and compliance of ATP commitments and accessible station requirements.

Delivery of Assistance

Revised Delivery Metrics

3.4   Our revised benchmarking framework will set out operators' performance against the following metrics in 2025, based on the results of our Passenger Assist user survey:

    A. Reliability – the percentage of passengers reporting they received none of the assistance they booked.

B. Satisfaction - the percentage of passengers reporting that they were 'satisfied' or 'very satisfied' with the assistance they received.

C. Staff knowledge and proficiency – the percentage of passengers reporting they were themselves 'satisfied' or 'very satisfied' with the knowledge and proficiency of the staff members providing their assistance.

3.5 We will use a three-year weighted score weighted 1/2/3 in favour of the most recent years. This both increases sample sizes and recognises performance over time.

3.6 Using these metrics, we will calculate a combined 'Delivery' score. We expect to weight this overall score in favour of reliability as follows: 50% Reliability, 25% Satisfaction, 25% Staff knowledge and proficiency.

3.7 Our published benchmarking report will rank operators in order based on this combined delivery score.

Operators in scope

3.8 In 2025, we will include in our benchmarking assessment any operator for which we have more than 50 responses per year to our survey:

Avanti West Coast, Chiltern, East Midlands Railway, Greater Anglia, Govia Thameslink Railway, GWR, LNER, Network Rail, Northern, Scotrail, SWR, Southeastern, Transpenine Express, Transport for Wales, and West Midlands Trains.

3.9 We receive lower response rates from operators that provide a lower number of booked assists at stations. As a result, in the 2025 report, we expect to exclude Arriva Rail London (London Overground), c2c, London Underground, and GTS Rail (the Elizabeth Line). Cross Country, Grand Central, Hull Trains, and Lumo do not operate stations and so are not included in our survey. They are also excluded from the assessment.

3.10 We are taking steps to increase the sample of Passenger Assist users we survey, and to explore alternative sources of data such as feedback from the Passenger Assist app, so that in future years so we can include as many operators as possible in our quantitative assessments.

Future evolution of the framework

3.11   From 2025, we are incorporating an additional question into the Passenger Assist survey to assess passengers' confidence to travel with an operator following their experience of the specific assistance in question. From 2026, we intend to report on the percentage of passengers that report they are 'more confident' or 'much more confident' about doing a similar journey with the same operator in future.

3.12   From 2026, we also intend to report on the percentage of eligible staff who have received either induction training or refresher training within the past two years.

3.13   Operators have been reporting to us against new passenger assistance outcome categories since April 2025. We will consider whether and how to build this data into the benchmarking assessment once we are content with data quality.

3.14   We are also working with RDG to explore how we might capture the reliability of Turn Up and Go users of Passenger Assist.

3.15   For satisfaction, we are exploring with RDG how we might use feedback provided by users of the Passenger Assist mobile app. We will consider whether and how to build this data into the benchmarking assessment once we are content with data quality.

Capability to Improve

Qualitative assessment

3.16   We will carry out a qualitative assessment of the poorest performing operators, identified through our assessment of delivery as those that most require improvement.

3.17   This assessment will consider four areas crucial to the organisational ability to manage and improve performance:

A.   **Monitoring & Reporting**: does the operator monitor delivery of assistance effectively?

B.   **Risks & Mitigations**: has the operator identified the risks that affect its performance and developed appropriate solutions?

C.   **Engagement**: does the operator effectively engage with passengers, staff and industry through this process?

D. **Training**: does the operator have evidence of achieving the training outcomes set out in our ATP Guidance?

3.18   Our assessment will benchmark operators against the descriptors set out in **Annex B**.

3.19   We will first send an information request to the targeted operators, with the expectation that they will be able to provide documentary evidence of their existing processes and how these lead to improvements. This process may then become an iterative one as the initial evidence leads to further questions and conversations and the provision of additional evidence.

3.20   The findings of this assessment will then be used to determine if we need to have further engagement with any operator to drive improvements.

Best practice

3.21   We recognise the value of identifying and sharing good practice in driving improvements. Alongside our information request of the weakest performers, we will also ask all operators for examples of good practice in the same areas, and include the best practices in our annual benchmarking report.

4. Next steps

4.1   We are grateful to respondents for their helpful and considered replies to our consultation. These have helped us in the development of our benchmarking framework.

4.2   Alongside the publication of this document, we are sharing our draft information request with the operators we have identified as requiring most improvement, as the first step in our qualitative assessment of their capability to improve.

4.3   We intend to publish the first benchmarking report later this year.

Annex A: Summary of consultation responses and ORR responses

Overview of proposed framework

Question 1: Do you agree with our plan to publish a report that benchmarks operators' performance levels in providing assistance to rail passengers? Explain your answer.

A.1   Respondents were largely in favour of ORR publishing a report that benchmarks operators' performance, highlighting how this would make it easier to identify issues and areas of improvement.

A.2   A number of respondents had concerns about the methodology being used to inform ORR's reporting.

A.3   Other respondents, particularly those representing industry, were concerned about the risk of undermining confidence to travel if the report presented a negative picture, highlighting that consideration would need to be given to how to best present the findings to passengers. On the other hand, some respondents felt that passengers would find comparisons between operators useful – helping them to set realistic expectations or choose between operators where applicable.

A.4   Some respondents felt that ORR should account for differences in the operating characteristics and customer needs between different operators, and subsequently the different and varied risks faced. It was suggested by some that operators should be benchmarked against similar operators.

ORR response

A.5   After considering the feedback received, we have developed a revised framework. The key changes are described in Section 2 of this document and the revised framework in Section 3. Having discussed it with members of our Accessible Stakeholder Forum, we agree with them that the benefits of driving improvements by highlighting poor performance outweigh the low risk of undermining confidence to travel.

A.6   We do not intend to benchmark operators only with similar operators; we measure assistance at stations, the largest of which (where the most assists take place) serve a number of different operators and markets, and so it does not make sense to group operators in the way that can be done with broader satisfaction surveys.

Question 2: Are the two headline performance categories (delivery and capability to improve) clear and meaningful? Is there anything else that should be included?

A.7   There was general support from respondents for the two proposed headline categories. Some industry respondents queried whether the 'Delivery' category was specific enough to assistance, and whether it was more concerned with subjective passenger experience than objective measures of assistance delivery. They proposed alternative categories such as 'customer experience' or 'Passenger Assist delivery'. Several industry respondents proposed we take account of effort, staffing levels or staff availability in our assessment. One respondent suggested we add a measure of confidence in the assistance service.

A.8   Some operators were also concerned that 'Capability to Improve' implied all operators are underperforming. A change to 'Continuous Improvement' was proposed. One respondent suggested adding a 'Complaints Handling' sub-category. A small number of respondents suggested an additional 'Disruption' category.

A.9   One respondent suggested the Training sub-category be moved from 'Delivery' to 'Capability to Improve', while another suggested the better performing operators should not need to provide evidence of their capability to improve.

ORR response

A.10   Our revised benchmarking framework retains two headline categories, but the 'Training' sub-category is now split, with quantitative measures of training completion and outcomes within 'Delivery' and a qualitative assessment of training in 'Capability to Improve', which we agree sits better alongside the other qualitative sub-categories. These will be focused on the worst performers.

A.11   We consider a confidence to travel metric has the potential to provide added insight and have added a question about confidence to travel to our Passenger Assist survey. We will consider how to include this data into the benchmarking framework from 2026. We considered the other suggestions for additions either out of scope of our regulatory role or already included in our proposed sub-categories.

Question 3: Are the performance level descriptors clear and meaningful? How should we name these performance levels (if at all)? Is three the right number of levels?

A.12   Respondents generally supported a 3-tier structure, though some organisations representing passengers and disabled people suggested adding a 'non-compliant' level, either as the third tier or an additional fourth tier. Conversely, industry respondents considered that compliance should be kept separate from benchmarking performance. One respondent suggested percentage scores would be more suitable than what they considered were 'artificial' tiers.

A.13   Many respondents expressed an appetite for more clarity and proposed a number of detailed changes to the performance level descriptors and level names to achieve this.

ORR response

A.14   As discussed further below, taking into account responses on other matters, we have made amendments to our approach. For the Delivery category, we will rank operators rather than place them into levels or tiers. For the Capability to Improve category, we will undertake an overall assessment of each sub-category for a small number of operators that our assessment of Delivery indicates we should be most concerned about.

Delivery (Reliability)

Question 4: Is the proposed Reliability metric clear and sufficiently reflective of reliability performance to allow comparison?

A.15   There was a mixed response about this metric, particularly concerning our proposed usage of our existing survey of booked assistance. However, there was appreciation for use of reliability as a key performance metric, noting the impact this can have on passengers who use assistance.

A.16   We received a range of responses in relation to our proposal to use the percentage of survey respondents who reported receiving none of the assistance they had booked as the data for this metric. While disabled passenger groups highlighted that receiving only part of an assistance booking could be as impactful as receiving none, another group felt a focus on just those reporting having received none was suitable – citing the fact that different elements of a booking may have different value to passengers.

A.17   We received feedback, especially from industry, arguing that the existing survey is not sufficiently representative of those using assistance services. They highlighted that a single demographic made up a disproportionately large proportion of survey responses, suggesting that the experiences of other passengers was therefore not being accounted for in the comparison between operators. Industry respondents also highlighted, either as a cause or effect, that the majority of survey respondents booked by phone, with those booking via other methods, especially the Passenger Assist app, under-represented.

A.18   A number of respondents noted the exclusion of TUAG assistance under this metric, as our existing survey is for booked assistance only. Two passenger groups felt this was problematic, citing TUAG assistance as a core ATP requirement. Industry responses highlighted that for some operators TUAG makes up the majority of the assistance delivered, so exclusion of this from their assessment would not fairly represent their performance.

A.19   Industry respondents raised concerns about the data on reliability provided by our survey not accounting for instances where assistance could not be provided fully due to factors outside their control. Some respondents highlighted the potential for new data that could be used, as an output of ongoing work to expand the data available using the Passenger Assist system.

A.20   A small number of respondents shared concerns about the amount of time between the assistance being received and the survey being undertaken, questioning whether this could undermine the data's reliability.

ORR response

A.21   We have considered the responses provided in response to Question 4. Our current Passenger Assist survey has a sample of over 9000 responses. We also boost the samples of the operators that provide fewer assists, by always assigning a response to their station in cases where a passengers travels between that station and another operator's station. Nonetheless, we acknowledge the limitations presented by using our existing survey of booked assistance, and have explored ways to improve the methodology behind this.

A.22   We share respondents' interest in TUAG assistance being incorporated as part of this metric. We have engaged with the Rail Delivery Group (RDG) to understand the potential for TUAG passengers being able to authorise their details being passed on for participation in our existing survey of assistance. However, we understand this will continue to not be possible in the short term as, for example,

operators would need to more routinely ask TUAG passengers provide their contact details. We will continue to engage with RDG on this matter.

A.23   At present, passengers who have only ever booked using the app will be excluded from our survey as they are not asked for consent to participate; whereas if a passenger has booked through other means they will have been asked for permission (which would then also apply to any future app bookings). To increase the number of respondents who have booked using the app, those who have only ever booked by app need to be given the opportunity to provide consent to be contacted. We have asked RDG to consider when this functionality, together with suitable processes to provide information to our survey provider, could be incorporated into the app

A.24   We have noted respondents' concerns about the time delay between passengers receiving their assistance and being invited to participate our survey. With RDG we are exploring possible opportunities to reduce this delay.

A.25   With the aim of increasing the number of respondents, we have shortened the survey by removing questions that were providing less useful insight. We hope this produces increases in sample sizes that enable more operators to be included in the quantitative benchmarking exercise. We note respondents' concerns about the survey's representativeness and, as suggested by some respondents, expect efforts to improve the number of responses to our survey from those who booked by app to improve this. However, having examined our survey results to determine whether passenger experience varies between different demographic groups (including age and gender, two dimensions raised by multiple industry respondents), we found there is little variance.

A.26   We have considered the suggestion to include passengers who received only some of their booked assistance alongside those who received none, and decided to continue with the current proposal. While we acknowledge the potential impact that receiving only some of your booked assistance could have, the severity of this will vary enormously from case to case and so provide a less clear measure of experience; we consider it provides no additional value as a tool for benchmarking compared with 'none'.

Question 5: Are the proposed Reliability performance level thresholds set appropriately?

A.27   There was a variety of responses to this question. Some respondents considered the thresholds had been set appropriately; others, mostly representing operators,

considered the upper threshold too challenging; meanwhile, some respondents representing passengers and disabled people proposed more stringent upper tier targets.

ORR response

A.28   Having reconsidered whether we can reasonably assess operators against suitably ambitious performance levels or tiers, and as set out in more detail in Section 3, our revised framework replaces performance levels or tiers with a ranking based on percentage scores for individual quantitative metrics, including Reliability. We will not be setting performance level thresholds, but we will still hold operators to account for poor performance. Operators should be aiming to eliminate assistance failures.

Question 6: Is the minimum sample size threshold for inclusion in this area of the assessment set appropriately?

A.29   A key concern shared by respondents, especially from those representing industry, was that individual operators' sample sizes were too low to be representative of actual performance – and therefore the proposed minimum sample size threshold itself, of 50 per year, was too low.

A.30   Other responses pointed towards the fact that all operators are proposed to be subject to the same minimum sample size threshold regardless of how many booked assists they deliver. An example was given of a smaller operator who would be excluded due to a sample size under 50, for whom a greater percentage of total assists would be represented in the survey than larger operators with sample sizes over 50. It was therefore proposed that each operator have its own minimum sample size threshold relative to its own booked assistance volumes.

A.31   The proposed minimum sample size threshold of 50 would result in some operators being excluded from the quantitative benchmarking analysis. Some respondents felt that this was not acceptable as part of the framework, while others emphasised the potential benefits associated with their inclusion.

ORR response

A.32   For those operators already included in the assessment we already consider the sample size to be sufficient for enabling assessment of operators' reliability of assistance. Using three years' worth of data, applying a 95% confidence level where satisfaction performance is at 90% or reliability performance is at 10% not received, results in confidence intervals ranging from 0.7% to 2.5% for all but one operator.

A.33 Nonetheless, as set out in our response to question 4, we share respondents' desire to increase individual operators' sample sizes, both to further increase the survey's representativeness and to reduce margins of error in our reporting.

A.34 We note some respondents' concerns about the exclusion of train operating companies that carry significant numbers of passengers. This benchmarking framework will only apply to those whose station staff are responsible for delivering assistance, which necessarily excludes train operators that do not operate stations.

Delivery (Satisfaction)

Question 7: Is the proposed Satisfaction metric clear and sufficiently reflective of reliability performance to allow comparison?

A.35 There was support for the inclusion of this metric across all respondent types, with it considered to add to our picture of operator's delivery of assistance.

A.36 A number of respondents, particularly those representing industry, repeated concerns about use of our existing survey of booked assistance, citing the same reasons as those covered under Question 4.

A.37 Some respondents expressed concern about the inclusion of a satisfaction metric, because of the subjectivity surrounding responses. They felt that a survey respondent's response could be too heavily influenced by other factors, both personal and circumstantial, and that it was therefore not an accurate representation of quality of service delivery.

A.38 Respondents made alternative suggestions for how satisfaction could be represented in the framework. It was felt that disaggregating the satisfaction metric and having different scores for different aspects of the service would be more valuable to operators in enabling them to identify improvements.

A.39 Multiple respondents, including from industry, proposed using customer satisfaction feedback provided through the app which is not limited to those who booked using the app). Users are prompted to rate the assistance (although it does not specify if this is about the assistance received, booking process or any other aspect) on a one to five star scale. This could provide a substantial sample size, with it being suggested that feedback is provided for approximately 1 in 25 assistance bookings.

A.40   It was suggested in consultation feedback that a new metric be added in place to complement the satisfaction (and reliability) metrics, relating to confidence. They cited that lack of confidence in the service had potential to discourage passengers from using the service and recommending it to others.

ORR response

A.41   We acknowledged concerns about use of the existing survey of booked assistance in our response to Question 4, and the actions identified will equally apply in relation to Satisfaction.

A.42   We remain confident that the inclusion of a Satisfaction metric is worthwhile in benchmarking passengers' experience of the Passenger Assist service. However, we see value in inclusion of a new Confidence metric, alongside Reliability and Satisfaction, in order to reflect the impact of a specific assistance experience on a passenger's future use or recommendation of the service. To provide this data, a new question has been added our existing booked assistance survey, which would allow for this metric to be added in the 2026 iteration of the benchmarking.

A.43   The current satisfaction metric will not be further disaggregated as part of the benchmarking framework as we want this to remain streamlined. Operators receive data outputs from our survey of booked assistance on a quarterly basis, and this provides them with a useful opportunity to interrogate their own data and use this to identify improvements.

A.44   We are exploring with RDG how we might use feedback provided by users of the Passenger Assist mobile app. We will consider if and how to build this data into the benchmarking assessment once we are content with data quality.

Question 8: Are the Satisfaction performance level thresholds set appropriately?

A.45   Similar to Reliability, responses varied between supporting the proposed thresholds, and those (mostly industry respondents) considering them too challenging. No respondents considered the thresholds too low.

ORR response

A.46   As with Reliability, having reconsidered whether we can reasonably place operators above or below meaningful performance level thresholds, and as set out in more detail in Section 3, our revised framework replaces performance levels or tiers with percentage scores for individual quantitative metrics, including Satisfaction, and an overall Delivery score. We will not be setting performance

level thresholds but will continue to hold operators to account for poor performance.

Question 9: Is the minimum sample size threshold for inclusion in this area of the assessment set appropriately?

A.47   The majority of responses to this question directly referred to their previous response to Question 6, which also asked about a minimum sample size threshold of 50 for operators.

A.48   A disabled passenger group cited their own research that found older passengers tend to report higher levels of satisfaction with bus and rail journeys. They therefore expressed concern that satisfaction levels may not accurately reflect the views of users more generally, including younger passengers.

ORR response

A.49   As noted in our response to Question 6, we share the desire to increase individual operators' sample sizes and have outlined our approach to increasing sample size. However, we remain confident that this proposed minimum sample size threshold provides representative samples.

A.50   We note the concern about the demographic representativeness of respondents and the specific impact this could have on a Satisfaction metric. We note there are no industry data on the age profile of users of Passenger Assist. Within our survey results, there is a slight difference between the satisfaction of older (over 65) and younger users, but we do not consider this significant enough to impact on our ability to compare the relative performance of operators.

A.51   In order to gain better insight into a possibly wider cross-section of users, we are exploring with RDG how we might use feedback provided by users of the Passenger Assist mobile app. We will consider if and how to build this data into the benchmarking assessment once we are content with data quality.

Delivery (training)

Question 10: Is the proposed Training metric clear (taking into account the relevant detail provided in the ATP guidance on mandatory training outcomes)?

A.52   Most respondents supported the inclusion of training elements within the benchmarking framework, but responses on behalf of the industry proposed their removal in favour of a separate reporting process.

A.53   The proposed training metric was not felt to be particularly useful by itself; alternatives proposed included measuring training effectiveness through complaints, or other measurable outcomes; or instead focusing more on a qualitative assessment of training programmes, for example via audits.

A.54   Two respondents proposed reviewing the training requirements within our ATP Guidance.

ORR response

A.55   We agree that Training is an important element of the revised benchmarking framework.

A.56   Within 'Delivery', we are introducing a new staff knowledge and proficiency metric as an indicator of the effectiveness of staff training to sit alongside the measure of training completion. This new metric will measure the percentage of passengers in our Passenger Assist user survey that report themselves 'satisfied' or 'very satisfied' with the knowledge and proficiency of the staff members providing their assistance.

A.57   A qualitative assessment of training provision will now sit within the 'Capability to Improve' assessment.

A.58   We note the comments on the ATP Guidance; we continue to keep the ATP Guidance under review but making changes to requirements is outside the scope of the consultation on the benchmarking framework.

Question 11: Are the Training performance level thresholds set appropriately?

A.59   A small number of responses proposed minor changes to either the qualitative or quantitative elements of the performance thresholds: adding staff feedback and training by people with lived experience of disability as a performance measure; and explicitly linking the thresholds to licence requirements or passenger expectations

ORR response

A.60   We have revised the Training qualitative performance descriptor for use as a benchmark within 'Capability to Improve' for the worst performing operators, taking account of respondents' comments.

A.61   The framework will not include quantitative performance thresholds for training, but the training metrics for staff knowledge and proficiency and staff training

completion will form part of the overall Delivery score for each operator. We will continue to hold operators to account for poor performance.

Capability to improve

Question 12: Taken together, do these criteria capture the key factors that determine an operator's capability to effectively manage and improve its provision of assistance? Is there anything missing, and is there anything listed that is not relevant?

A.62   We received a number of suggestions for further sub-categories in this area, in addition to Monitoring and Reporting, Risks and Mitigations and Engagement. These included complaints handling, innovation, interchange and collaboration. We also received several suggestions to change the name of the area to 'Continuous Improvement', to avoid implications of universal poor performance.

ORR response

A.63   We have added the Training sub-category as an additional criterion to this area. Using respondents' feedback, we then revised the performance descriptors of all the sub-categories, including Training. We will use these as benchmarks when we qualitatively assess the capability to improve of the worst performing operators.

Question 13: Can these criteria be critically evaluated, based on the assessment of a proportionate evidence gathering, submission and assessment process?

A.64   Many respondents interpreted this question as applying to the whole framework rather than the Capability to Improve criteria, suggesting our overall approach was disproportionately burdensome and repeating their concerns about relying on our Passenger Assist survey for the quantitative evidence base.

A.65   Industry respondents were keen to better understand how our proposed approach to qualitative benchmarking of organisational capability to improve would interact with our ongoing monitoring of Accessible Travel Policy commitments, and compliance measures such as the improvement plan agreed with Northern. They were concerned about any additional reporting burden imposed by the information requests we would make to provide the evidence for our assessment in this area.

A.66   They also queried whether it was reasonable to benchmark operators on a subjective assessment of qualitative measures, and whether external factors would be taken into account.

A.67   We received one suggestion to instead focus on best practice as an improvement tool.

ORR response

A.68   Recognising the importance of minimising regulatory burdens to where we can have most impact, our revised framework focuses our qualitative assessment on the Capability to Improve of the worst deliverers of assistance, as explained in Section 2, and remains limited to the sub-categories proposed in the consultation, updated in the light of respondents' comments. We explain how this would work, including any next steps following the initial qualitative assessment, in Section 3.

A.69   Recognising the importance of sharing and adopting best practice in driving improvements, we also intend to use the benchmarking report to highlight this across the industry, focusing on the Capability to Improve sub-categories.

Equality Impact Assessment

Question 14: Please indicate how the assistance benchmarking framework might be used to even further advance equality of opportunity and reduce discrimination for groups protected under the Equality Act.

A.70   Respondents reinforced the need for a full Equality Impact Assessment of the final framework to capture its impact on advancing equality of opportunity.

A.71   A number of respondents used this question to reinforce suggestions made in response to other questions as measures that would further advance equality of opportunity and reduce discrimination. Several also called for additional accessibility initiatives such as support for neurodiverse passengers and provision of level boarding facilities.

A.72   One respondent suggested more focus on the reliability of assistance and on opportunities to expand its availability would be key to meeting ORR's Equality Act duties.

ORR response

A.73   We thank respondents for their suggestions. We have published a regulatory and equality impact assessment of the revised benchmarking framework at Annex C.

Annex B: Capability to Improve Performance descriptors

Monitoring and Reporting

B.1   When we undertake an assessment of an operator's **Capability to Improve** we will ask them to provide evidence of how and to what extent:

(i)   The operator conducts its own data quality assurance, including assurance of data generated at the frontline, for accuracy and completeness, using comparator data sources where appropriate.

(ii)   The operator takes appropriate steps to address any gaps in data, or data quality issues.

(iii)   The operator has a clear, consistent process for frontline staff to capture accurate, timely and comprehensive data on assistance outcomes – for both booked and Turn Up and Go (TUAG) assistance.

(iv)   The operator considers a range of other sources to capture information, feedback and insight on issues that impact reliability, including complaints.

(v)   The operator uses this data to generate performance reports, regularly reviewed by management.

(vi)   As a result, the operator has a good, evidence-based understanding of the usage profile for assistance, including types of assistance (booked and TUAG), and demand by station.

(vii)   The operator completes accurate and timely Core Data submissions on assistance (booked and TUAG) and complaints for ORR, each rail period.

B.2   We will also ask all operators for **Best Practice** examples of performance that goes above and beyond these baseline expectations.

Risk and Mitigation

B.3   When we undertake an assessment of an operator's **Capability to Improve** we will ask them to provide evidence of how and to what extent they:

       (i)   A solid, evidence-based understanding of the main risks to delivery of assistance, including service disruption.

       (ii)   Appropriate mitigations are developed on the basis of this risk analysis.

       (iii)   Mitigations are promptly implemented (this may be on a trial basis).

       (iv)   The roll-out of mitigations is monitored, and their impact kept under review, with appropriate follow-up action where necessary.

       (v)   Responsibility for assistance is documented in job roles and up-to-date operational handbooks.

B.4   We will ask all operators for **Best Practice** examples of the following:

       (i)   Where there are failures, managers proactively investigate what happened, and why.

       (ii)   Risk analysis and mitigations are, where appropriate, specific to stations, staff delivery model, and other relevant facets of the operator's services.

       (iii)   The operator looks for opportunities to improve – adopting or trialling technical innovations or good practice developed in other sectors.

       (iv)   Challenging internal performance targets are set and monitored.

Engagement

B.5   When we undertake an assessment of an operator's **Capability to Improve** we will ask them to provide evidence of:

       (i)   Regular, structured engagement with a forum of disabled passengers (including those with non-visible disabilities), with performance reports shared and discussed with this forum.

       (ii)   Engagement with interfacing operators to investigate the causes of assistance failures.

    (iii)   Engagement with staff to elicit feedback that drives service improvements

    (iv)   Provide appropriate input to cross-industry groups and forums, to share mitigations and help develop cross-industry solutions.

    (v)   Incorporate learning from this engagement into the operator's own mitigations.

B.6   We will ask all operators for **Best Practice** examples of the following:

    (i)   How the passenger forum feedback has contributed to improvements in delivery of assistance.

    (ii)   Regular engagement with interfacing operators, to discuss risks and mitigations, with tangible outputs.

    (iii)   Use of other forms of passenger feedback and complaints to drive service improvements.

    (iv)   The operator to be at the forefront of innovation, trialling or implementing technology, common learning and shared best practice.

Training

B.7   When we undertake an assessment of an operator's **Capability to Improve** we will ask them to provide evidence of how and to what extent:

    (i)   On an ongoing basis, the operator assures itself that all frontline staff are meeting the relevant training outcomes set out in Appendix D of the ATP Guidance, including where staff may need refresher training.

    (ii)   As part of their training programme, staff are provided with the opportunity to hear from passengers with lived experience of disability and have the opportunity to discuss what they have heard.

    (iii)   Training content remains up to date, fit-for-purpose, has input from people with lived experience of disability, and incorporates feedback from staff that have undertaken the training.

B.8   We will ask all operators for **Best Practice** examples of how they have clearly gone above and beyond the baseline regulatory requirements. This may take several forms, but might include:

(i)   Opportunities for staff to further develop their understanding and skills.

(ii)   Additional training in the needs of passengers with specific disabilities.

(iii)   Training being delivered in a face-to-face format from individuals with lived experience.

Annex C: Equality & Regulatory Impact Assessment

C.1   As a public body, ORR is required by the Equality Act 2010 (EA2010) to comply with the Public Sector Equality Duty (PSED).

C.2   ORR must, in every policy decision that could affect individuals with protected characteristics - either positively or negatively - have due regard to the need to eliminate discrimination, harassment and victimisation, advance equality of opportunity, and foster good relations between all groups.

C.3   This annex combines regulatory and equality assessments of our new approach to benchmarking operators' passenger assistance delivery. It considers the impact both on rail passengers and industry. The intended beneficiaries of our new approach are rail passengers who receive assistance to travel on the rail network. This will include older and disabled people who need assistance to access the rail network. Pregnant women may also benefit from the proposal. These specific groups are the focus of our passenger assessment.

Equality Impact on Passengers

C.4   The benchmarking framework is designed to increase the reliability of passenger assistance, reduce cases of failed assistance and improve the experience of people that use the service, including older and disabled people. It will therefore advance equality of opportunity to travel and access education, training, employment, services and leisure and reduce discrimination for these groups.

C.5   We want to incorporate additional sources of data into the framework over time, for instance on the experiences of passengers who Turn Up and Go, which would further extend the positive impact..

C.6   As set out in Sections 2 and 3 of this document, we considered a range of alternative and supplementary proposals put forward by respondents to our consultation and, as a result, made a number of changes to our framework. Where we have not included suggestions in our final framework or in our plans to incorporate additional data in the future, we considered they would not be effective as a benchmarking tool and therefore would not advance equality of opportunity or reduce discrimination for older and disabled people.

Delivery of Assistance

C.7   Compared with our consultation proposals, we consider that the changes we made to how we benchmark operators for their delivery of passenger assistance will be more effective in driving performance improvements and therefore increase the positive impact described above.

Capability to Improve

C.8   With regard to operators' capability to improve, we considered either assessing all operators (as proposed in our consultation) or assessing only the worst performers (as per the revised framework). Both options would provide a positive impact on reducing discrimination and advancing equality of opportunity for older and disabled people. If no constraint on resources or regulatory burden is assumed, an assessment of all operators every year may lead to more widespread improvements that lead to a larger positive impact, when compared with only assessing the worst performers. However, within such constraints, we consider a deeper dive on the worst performers to deliver a more significant positive impact than a lighter touch assessment across all operators, which might risk a minimal impact on the better performers.

Regulatory Impact on Industry

C.9   Train and station operators are already obliged as part of a published Accessible Travel Policy to commit to the provision of assistance to passengers that require it. The new benchmarking framework adds no new requirements on operators; rather, it is designed to incentivise more reliable delivery of existing obligations.

Delivery of Assistance

C.10   The quantitative assessments we will undertake of train operators will utilise existing data sources: our survey of users of passenger assistance and core data we collect on staff training. In future we will consider using further sources of industry information that are already being developed. We therefore consider this part of the framework to have no or minimal regulatory impact on industry.

Capability to Improve

C.11   Our qualitative assessment of capability to improve will impose a new regulatory burden on a small number of the worst performing operators: each year, completion of an information request, engagement with ORR and undertaking any further intervention as may be required following the conclusion of the assessment.

C.12   Compared with our consultation proposal to undertake a capability to improve assessment of all operators, our revised framework places a smaller regulatory burden on better performing operators and a potentially somewhat greater regulatory burden on a small number of the worst performing operators.