

---

## Enhancing Service Quality Assessment in Rail Transport Through the Application of the Critical Incident Technique

---

Submitted 25/06/25, 1st revision 18/07/25, 2nd revision 11/08/25, accepted 08/09/25

Paulina Gajewska<sup>1</sup>, Irena Szewczyk<sup>2</sup>, Artur Budzyński<sup>3</sup>,  
Agnieszka Cholewa-Wójcik<sup>4</sup>

### **Abstract:**

**Purpose:** The purpose of this study is to evaluate the quality of rail transport services in Poland using the Critical Incident Technique (CIT). The research focuses on conductor–passenger interactions and aims to identify the most critical service encounters, both positive and negative, that shape passenger satisfaction.

**Design/Methodology/Approach:** The study was based on semi-structured interviews with passengers travelling on the Bielsko-Biala–Wrocław route operated by PKP Intercity. Ten critical incidents were collected (five positive and five negative). A thematic categorisation was applied, followed by the development of a methodological framework for assessing interrater reliability (including Fleiss' Kappa, Kendall's W, Intraclass Correlation Coefficient and Weighted Kappa) and an incident-weighting model derived from service quality theory.

**Findings:** The analysis revealed a dual image of train conductors: in some cases, they were empathetic and professional, while in others, they were passive and disengaged. Positive incidents highlighted helpfulness, information provision, and courtesy, while negative cases concerned lack of engagement, weak communication, and inadequate service recovery. Methodologically, the study shows that a multidimensional reliability framework offers a more robust evaluation than a single-coefficient approach. The proposed weighting model captures the asymmetric importance of incidents related to safety, interpersonal behaviour, information, and service recovery.

**Practical Implications:** The results underscore the need for targeted training of railway staff, particularly in areas such as communication, empathy, and problem-solving. The enhanced CIT framework can be applied by transport operators to identify priorities for service improvements and to allocate resources more effectively by considering both the frequency and importance of incidents.

**Originality/Value:** The study contributes to both theory and practice by advancing the CIT

---

<sup>1</sup>dr inż., Assist. Prof. University of Bielsko-Biala, Department of Management and Transport, Poland, e-mail: [pgajewska@ubb.edu.pl](mailto:pgajewska@ubb.edu.pl);

<sup>2</sup>dr inż., Assist. Prof. University of Bielsko-Biala, Department of Management and Transport, Poland, e-mail: [iszewczyk@ubb.edu.pl](mailto:iszewczyk@ubb.edu.pl);

<sup>3</sup>dr inż., Assistant, Department of Packaging and Logistics Processes, Poland, e-mail: [abudzyns@uek.krakow.pl](mailto:abudzyns@uek.krakow.pl);

<sup>4</sup>dr hab. inż., Prof., Department of Packaging and Logistics Processes, Poland, e-mail: [cholewaa@uek.krakow.pl](mailto:cholewaa@uek.krakow.pl);

*methodology. It introduces a novel combination of reliability measures, proposes an incident-weighting model, and suggests a dynamic consensus-building process among evaluators. These methodological innovations make the approach replicable in other contexts where human interactions, including those in other transport sectors, significantly influence service quality.*

**Keywords:** *Railway conductors, passenger experience, service recovery, incident weighting, interrater reliability.*

**JEL codes:** *R41, C83, D12.*

**Paper type:** *Research article.*

**Acknowledgment:** *The publication presents the results of the Project financed from the subsidy granted to the Krakow University of Economics.*

## 1. Introduction

The *Critical Incident Technique* (CIT) represents a significant methodological advancement in understanding the dynamics of service quality within complex organizational environments (Kostamo *et al.*, 2019). As a qualitative research methodology, CIT enables researchers to capture and analyse specific episodes of human behaviour that result in either exceptionally positive or negative service outcomes, providing unprecedented insights into the mechanisms underlying customer satisfaction and dissatisfaction (Butterfield *et al.*, 2005).

The contemporary service landscape presents unique challenges for quality assessment, particularly in sectors where human interaction forms the cornerstone of service delivery (Craig *et al.*, 2013). Unlike tangible products that can undergo standardized quality inspections before customer contact, services are characterized by their inherent variability and the simultaneity of production and consumption. This fundamental characteristic creates what researchers term the "*service quality gap*" - the disconnect between customer expectations and their actual service experiences (Moore *et al.*, 2015).

The evolution of CIT as a diagnostic tool has been particularly pronounced in transportation research, where service encounters often involve multiple touchpoints and varied stakeholder interactions (Stewart and Smith, 2014). In rail transport specifically, the quality of service delivery is influenced by numerous factors, including infrastructure reliability, staff competency, operational efficiency, and the complex interplay between organizational policies and frontline employee behaviour. Traditional quantitative approaches, while valuable for measuring aggregate satisfaction levels, often fail to capture the nuanced, contextual factors that drive critical service moments (Deaton and Cartwright, 2018).

CIT's theoretical foundation rests on the premise that customer perceptions are fundamentally shaped by discrete, memorable incidents rather than generalized service impressions (Flanagan, 1954). This approach acknowledges that service quality is not merely an objective measure but a subjective construct formed through specific experiential episodes. The methodology's strength lies in its ability to identify and categorize these pivotal moments, thereby revealing the underlying patterns and triggers that influence customer behaviour and loyalty decisions (Chell and Pittaway, 1998).

The application of CIT in service quality research offers several distinct methodological advantages (Holloway and Beatty, 2008):

First, it provides rich, contextual data that illuminates the 'why' behind customer satisfaction ratings, moving beyond surface-level metrics to uncover deeper motivational and emotional drivers.

Second, CIT captures the temporal dimension of service experiences, allowing researchers to understand how specific incidents unfold and their subsequent impact on customer perceptions.

Third, the methodology facilitates the identification of service recovery opportunities by highlighting both successful and unsuccessful service encounters.

Recent developments in CIT applications have demonstrated their particular relevance for evaluating complex service systems such as public transportation networks (Davis, 2006).

The methodology has proven especially valuable in contexts where service delivery involves multiple organizational levels and where both planned service elements and unexpected contingencies shape customer experiences. In rail transport, for instance, CIT has been instrumental in identifying critical factors that influence passenger loyalty, from routine operational procedures to crisis management responses (Callan, 1998).

The integration of CIT with modern analytical frameworks has expanded its utility in contemporary service research (Weber, 1990). Digital technologies now enable real-time capture of critical incidents through mobile applications and online platforms, while advanced analytical techniques allow for more sophisticated pattern recognition and predictive modeling. This technological evolution has transformed CIT from a retrospective analytical tool into a dynamic instrument for continuous service improvement.

Despite its proven utility, the application of CIT in evaluating rail transport services remains underexplored, particularly in the context of regional transportation networks where service quality perceptions may differ significantly from

metropolitan systems (Rodrigues *et al.*, 2015). Existing research has predominantly focused on large-scale transportation hubs or urban transit systems, leaving a gap in understanding how service quality dynamics operate in medium-scale regional connections.

Furthermore, previous CIT applications in transportation have often emphasized operational metrics rather than the interpersonal dimensions of service delivery (Rothman *et al.*, 2006). This study addresses this limitation by focusing specifically on conductor-passenger interactions as critical determinants of overall service quality perception. The role of frontline service personnel as "*service ambassadors*" has been well-established in hospitality and retail contexts, yet remains inadequately explored in rail transport research.

The significance of this research extends beyond academic inquiry to practical service management applications. Rail transport operators face increasing pressure to enhance service quality while managing operational constraints and cost considerations. CIT provides a framework for identifying high-impact service improvement opportunities, enabling organizations to prioritize investments in areas that most directly influence customer satisfaction and loyalty (Sallis *et al.*, 1999).

This research contributes to the growing body of literature examining service quality in transportation contexts by applying CIT methodology to evaluate PKP Intercity services on a specific regional route. The study's focus on the Bielsko-Biała – Wrocław connection provides insights into service quality dynamics in a medium-scale regional transport context, addressing the identified research gap while offering practical implications for service enhancement strategies.

The methodological approach employed in this study demonstrates the versatility of CIT as an analytical framework while establishing a replicable model for similar investigations in comparable transport contexts. By systematically categorizing critical incidents and analysing their underlying patterns, this research provides both theoretical insights into service quality formation and practical guidance for transport service improvement initiatives.

## 2. Literature Review

Current social science literature lacks comprehensive insights into critical incidents during service interactions. The *Critical Incident Technique* (CIT) offers a valuable framework that enables respondents to report atypical, non-routine events. Despite their recognized importance, such incidents have largely been underrepresented in traditional service procedure analyses or evaluations of service attributes (Callan, 1998).

Grace (2007) utilized CIT to collect detailed and genuine narratives, showcasing that this method is especially effective for exploring behavioural intentions.

Additionally, CIT is valuable in uncovering essential factors, dimensions, and outcomes connected to customer experiences that may entail discomfort or embarrassment in different service contexts.

The method's flexibility allows it to be adapted to various research settings (Flanagan, 1954). CIT produces rich qualitative data that can support scale development or provide deeper insights into service dynamics (Holloway and Beatty, 2008).

CIT contributes to theoretical development as part of the broader spectrum of qualitative methodologies (Yeoman *et al.*, 2001a). Its wide applicability makes it a powerful tool for diagnosing and resolving organizational issues. When used internally, CIT helps uncover practical challenges, improves internal processes, enhances collaboration, and fosters organizational learning (Davis, 2006).

Nevertheless, some methodological concerns have been identified in earlier studies. These include the potential impact of social desirability bias, difficulties with respondents' memory, questions regarding the reliability and validity of findings (Chell, 2004), and ambiguity related to coding and categorization (Weber, 1990). Like many qualitative techniques, researchers may interpret and classify incidents differently (Stewart and Smith, 2014). However, these challenges do not undermine the potential of CIT, mainly if addressed carefully.

CIT is often defined as a qualitative interview method designed to explore significant events—whether incidents, processes, or issues—identified by the respondent. It examines how these situations are managed and the outcomes they produce from the respondent's perspective, incorporating cognitive, emotional, and behavioural dimensions (Callan, 1998). Typically, data is gathered through interviews in which individuals are invited to recall a specific and memorable incident and describe it in detail. These accounts are then grouped by theme and systematically categorized using content analysis (Stauss and Weinlich, 1997).

### **3. Methodology**

#### **3.1 Critical Incident Technique**

Various methods can be applied to assess service quality, including SERVQUAL, SERVPERF, the *Customer Satisfaction Index* (CSI), and the *Critical Incident Technique* (CIT). This study selected the CIT method because it captures detailed customer experiences that evoke intense satisfaction or dissatisfaction.

The *Critical Incident Technique* is a qualitative research method used to collect and analyse specific incidents that customers perceive as particularly positive or negative. For an event to be classified as a critical incident, it must meet the following criteria:

- it involves direct interaction between a customer and a service employee,
- the customer evaluates the event, positively or negatively,
- it represents a distinct episode, separate from general impressions,
- it is vivid enough for the respondent to describe in detail and for the researcher to visualize.

John Flanagan introduced the CIT methodology in his seminal 1954 article, establishing a comprehensive framework that provides guidelines for the systematic application of the technique. According to Flanagan (Flanagan, 1954), the CIT process encompasses five distinct stages, each requiring careful consideration and methodological rigor (Kreps, 2017):

1. **General aims** - key decisions are made about the exact purpose of the CIT study. Researchers must clearly define what they want to discover through the investigation. Background data should be collected through comprehensive needs analysis, audience analysis, and literature reviews to assess contextual information about the program or service being studied (Flanagan, 1954).

2. **Plans and specifications** - decisions are made about the sources for critical incidents. Respondents should be carefully selected based on their knowledge and experience with the programs being evaluated. They should represent key stakeholders of the programs being assessed to ensure that critical incidents represent major relevant perspectives. Additionally, decisions must be made about the most effective approaches to recruit and sample respondents (Kreps, 2017).

3. **Collecting the data** - methodological decisions are made regarding data collection approaches, including whether to utilize direct or recorded observations, structured or semi-structured interviews, or standardized questionnaires. Critical incidents can be gathered through different methods - verbally, in writing, or by analysing preserved media. Data collection may take place face-to-face or through mediated channels such as phone or online platforms. Decisions must also be made about personnel training for data collectors and the development of standardized instruments to ensure uniform data collection procedures (Flanagan, 1954).

4. **Analysing the data** - systematic decisions are made about analytical approaches for interpreting collected critical incident data. Content analysis is frequently employed to identify recurring themes and patterns within the dataset. Critical incidents are typically recorded through written transcriptions or audio/video recordings, coded with unique identification numbers, and systematically reviewed and categorized into thematic groups through iterative processes. Intercooder reliability testing is essential to establish the intersubjective reliability of the coding process (Kreps, 2017).

5. **Interpreting and reporting** - critical incident themes are quantified and analysed to establish the frequency of positive and negative incidents, providing an

overall assessment of satisfaction with the programs being studied. The most frequently identified critical incident topics suggest both the strongest positive program features that should be reinforced and the most significant negative themes requiring correction.

When reporting findings, detailed descriptions of data collection procedures and thematic analysis are provided, along with implications for program enhancement drawn from theme frequency analysis and representative critical incidents (Flanagan, 1954).

This systematic five-stage approach ensures methodological rigor while maintaining the flexibility necessary for adaptation to various research contexts and organizational settings (Herndon and Kreps, 2001).

The CIT process for this study includes several key stages adapted from Flanagan's framework:

- **Data collection through interviews** - direct interviewing of service users using a semi-structured approach. Interviewers encouraged respondents to recall and describe detailed service encounters based on predefined guiding questions.
- **Initial categorization** - organizing incidents based on standard features identified in the interview data. Subsequently, independent experts were asked to classify these incidents into categories. An agreement of at least 80% among the experts was considered successful. Discrepancies were resolved through discussion or reclassification.
- **Final categorization** - a two-stage review process resulted in the final categorization of incidents into clearly defined groups.
- **Statistical summary** - the proportion of positive and negative incidents was calculated and compared within and across categories.

The methodological approach demonstrates that CIT can generate not only insightful, rich narrative descriptions of experiences with essential programs, but can also provide relevant evidence and direction for improving these programs, making it particularly valuable for service quality assessment in transportation contexts (Herndon and Kreps, 2001).

### **3.2 Research Procedure and Findings**

The empirical study utilizing the *Critical Incident Technique* was conducted in November 2024. This research involved six individuals regularly using PKP Intercity services on the Bielsko-Biała – Wrocław railway route. The participants were invited to engage in semi-structured interviews featuring six open-ended questions to elicit detailed descriptions of both positive and negative service encounters. Respondents were asked to answer the following questions:

- *How often do you use PKP Intercity services?*
- *Please provide a detailed account of an incident that occurred while using PKP Intercity services on the Bielsko-Biała – Wrocław route.*
- *When did this event take place?*
- *Were there any special circumstances surrounding this incident?*
- *How did the conductor team respond?*
- *How often do you have to wait for a delayed train?*

All interviews were recorded, transcribed, and analysed qualitatively using the CIT methodology. A total of 10 critical incidents were identified and validated—5 positive and 5 negative—resulting in an even distribution of experiences. Given the study's limited scope, there were six participants, all passengers. Every interview was determined to be methodologically valid.

The analysis revealed 10 incidents, evenly split between positive and negative. Based on the gathered data, it is impossible to definitively conclude whether the services are generally satisfactory or unsatisfactory. The findings suggest the presence of positive and negative patterns in the behaviour of on-board staff.

## **4. Research Results and Discussion**

### **4.1 Analysis of the Critical Incident Division**

The collected critical incidents were grouped thematically into three main categories that reflect key aspects of passenger service quality in rail transport. Each group contained both positive and negative incidents, illustrating the dual nature of passengers' service quality perceptions.

The analysis of empirical material allowed for the identification of three main groups of incidents:

*Staff response to service issues (Group 1).*

*Staff responsiveness to passenger needs and behaviour (Group 2).*

*General conduct and cultural behaviour of the conductor team (Group 3).*

Based on the detailed analysis of collected narratives, a thematic categorization of critical incidents was developed to highlight the main dimensions of passenger-conductor interactions. The resulting classification captures the most recurrent and meaningful behavioural patterns observed in the empirical material and forms the basis for the following visualization.

Table 1 illustrates the classification of critical events into distinct categories and groups, based on observations made at railway crossings along the Bielsko-Biała – Wrocław route.

To illustrate the key dimensions of perceived service quality, the most relevant observations from the thematic grouping of critical incidents are outlined below:

*Group 1. Staff response to service issues:*

The first category focused on how conductors responded to problems related to unavailable services or delays.

*Positive incidents:*

When the mobile catering service was unavailable, the conductor explained the reasons for this situation, even though it had been indicated when purchasing the ticket that drinks and food could be bought.

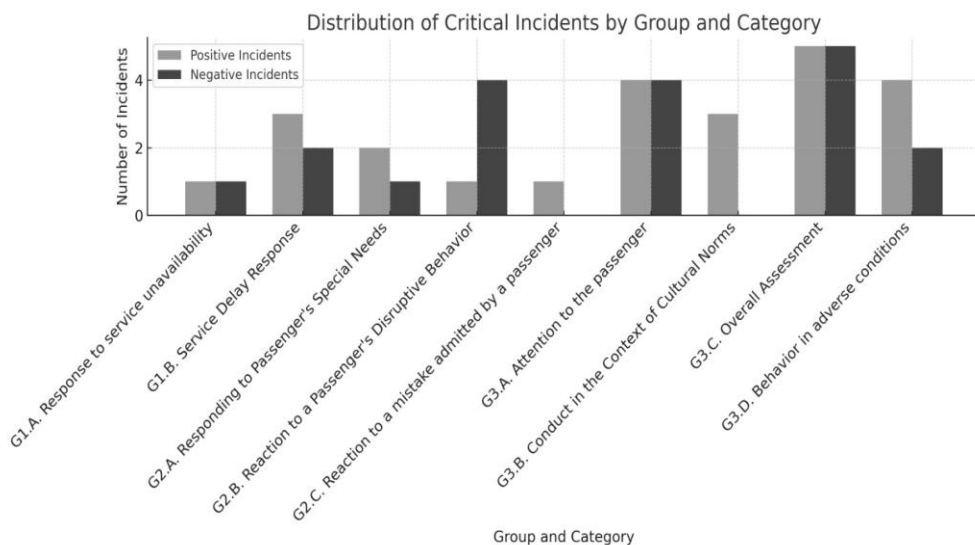
In the event of train delays, the conductor informs passengers about the expected delay time and its cause. Staff effectively communicated both the duration and cause of delays, helping to manage passenger expectations.

*Negative incidents:*

The conductor did not react to the lack of service that was supposed to be provided, nor did he inform about the length of the delay and its cause. In some cases, conductors failed to provide any information about unavailable services or reasons for delays, leaving passengers feeling uninformed and frustrated.

This category comprised seven incidents (four positive and three negative), accounting for 70% of the collected material.

**Figure 1.** Distribution of Critical Incidents by group and category (visualization)



**Source:** Authors' calculations.

**Table 1.** Division of critical events into groups and categories on the Bielsko-Biala – Wrocław route

| Groups and categories                                                                       | Positive events |      | Negative events |      | Sum    |      |
|---------------------------------------------------------------------------------------------|-----------------|------|-----------------|------|--------|------|
|                                                                                             | number          | %    | number          | %    | number | %    |
| <b>Gr. 1. The reaction of the conductor team to any errors on the part of PKP Intercity</b> |                 |      |                 |      |        |      |
| A. Response to service unavailability                                                       | 1               | 10%  | 1               | 10%  | 2      | 20%  |
| B. Service delay response                                                                   | 3               | 30%  | 2               | 20%  | 5      | 50%  |
| <b>Group 1 - total</b>                                                                      | 4               | 40%  | 3               | 30%  | 7      | 70%  |
| <b>Gr. 2. The reaction of the conductor team to the needs and requests of passengers</b>    |                 |      |                 |      |        |      |
| A. Responding to passengers' special needs                                                  | 2               | 20%  | 1               | 10%  | 3      | 30%  |
| B. Reaction to a passenger's disruptive behaviour                                           | 1               | 10%  | 4               | 40%  | 5      | 50%  |
| C. Reaction to a mistake admitted by a passenger                                            | 1               | 10%  |                 |      | 1      | 10%  |
| <b>Group 2 - total</b>                                                                      | 4               | 40%  | 5               | 50%  | 9      | 90%  |
| <b>Gr. 3. Reaction to spontaneous action of staff</b>                                       |                 |      |                 |      |        |      |
| A. Attention to the passenger                                                               | 4               | 40%  | 4               | 40%  | 8      | 80%  |
| B. Conduct of the conductor team in the context of cultural norms                           | 3               | 30%  |                 |      | 3      | 30%  |
| C. Overall assessment                                                                       | 5               | 50%  | 5               | 50%  | 10     | 100% |
| D. Behaviour in adverse conditions                                                          | 4               | 40%  | 2               | 20%  | 6      | 60%  |
| <b>Group 3 - total</b>                                                                      | 10              | 100% | 11              | 110% | 27     | 270% |

*Source:* Authors' calculations.

#### *Group 2. Responsiveness to passengers' needs and behaviour:*

The second category addressed how the train staff handled passenger requests, special needs, or behavioural issues.

##### *Positive incidents:*

A conductor assisted a passenger by checking details about connecting trains at the destination station. Staff appropriately intervened when a passenger behaved disruptively, maintaining order and comfort for others. A polite and respectful correction was issued when a passenger mistakenly took the wrong seat.

##### *Negative incidents:*

Some conductors ignored passenger requests, such as changing seats, showing a lack of initiative. No reaction to the annoying behaviour of passengers – mothers with screaming children running around the entire carriage. In several cases, staff did not

intervene when children were disruptive, negatively affecting nearby passengers' travel experience.

This group was responsible for nine incidents (four positive and five negative), accounting for 90% of all identified incidents.

*Group 3. General conduct and cultural norms:*

The final group evaluated the train staff's overall demeanour, including attentiveness, adherence to cultural norms, and behaviour under challenging conditions.

*Positive observations:*

Passengers frequently noted the attentiveness and courteous behavior of conductors. In challenging circumstances, conductors demonstrated patience, professionalism, and a customer-oriented approach. Conductors upheld cultural expectations of politeness and respect, which positively influenced perceptions of the company. Conductors devote a relatively large amount of time to passengers, demonstrating great culture, patience, and finding ways out of situations in unfavourable conditions.

*Negative observations:*

In contrast, some passengers reported that conductors limited their interaction to ticket inspection, failing to engage meaningfully during the journey. A passive approach to service delivery was particularly problematic during unexpected or disruptive situations. The conductor did not pay much attention to the passenger; interaction occurred only when checking the ticket, and there was a notable lack of interest when unusual situations arose.

This group included the largest share of incidents—twenty-seven (ten positive, eleven negative)—indicating its central role in shaping overall service impressions.

Based on the conducted research, it is possible to precisely analyse which groups of rail services provided on the Bielsko-Biała – Wrocław route meet customer expectations (positive events) and which do not meet expectations (adverse events).

The analysis reveals a dual image of train conductors. On one hand, they are viewed as empathetic professionals who provide support, resolve issues, and ensure smooth passenger experiences. On the other hand, some conductors seem disinterested and passive, engaging with passengers only when necessary.

This dichotomy has significant implications for service perception. The conductor's behaviour directly influences customer satisfaction and PKP Intercity's brand image. Positive interactions were characterized by attentiveness, clear communication, and courtesy, while negative ones stemmed from disengagement, poor conflict resolution, and an inability to meet basic expectations.

Nowadays, the profession of a conductor is not very popular and is associated with frequent and long trips. A conductor must be resistant to stress because every day he talks to hundreds of passengers, each of whom behaves differently – some are problem-free, while others find problems in everything. A conductor is responsible for the image of the entire company – how they behave towards passengers determines how passengers will perceive PKP as a whole.

These findings highlight the importance of targeted training programs that enhance soft skills, especially in customer service and conflict resolution. Investing in these competencies will likely improve service delivery consistency and positively impact passengers' overall perceptions.

The conducted research shows that the *Critical Incident Technique* (CIT) provides information on the quality of services offered on the Bielsko-Biała – Wrocław route. It enables us to understand the image of the conductor, identify who serves as the company's showcase, determine how it should be created, what satisfies passengers, and what situations and behaviours should be avoided. This knowledge enables the improvement of the company's image, which will result in a more positive perception of PKP among potential users of the company's services.

#### **4.2 Theoretical Framework for Enhancing CIT Reliability in Transportation Research**

The analysis of PKP Intercity conductor behaviours revealed a dual nature of service delivery, which not only substantiates the usefulness of the *Critical Incident Technique* (CIT) in service quality research but also exposes its methodological limitations when applied to complex transportation contexts.

Building upon this duality, the present study introduces a theoretical enhancement of the CIT methodology that addresses concerns over reliability, the weighting of service incidents, and the dynamic consensus of evaluators. These enhancements stem from empirical research in applied psychology, healthcare, and service marketing, aiming to enhance the analytical robustness of CIT in public transport studies.

Recent scholarship has identified serious problems with interrater reliability in CIT applications involving qualitative incident classification. (Koch *et al.*, 2009), in a study conducted under real-world conditions, observed that traditional CIT procedures yielded interrater Kappa coefficients between  $\kappa = 0.09$  and  $\kappa = 0.14$  ( $p > 0.05$ ), well below acceptable levels for scientific application.

According to interpretive criteria established by (Gorbaniuk, 2016), values below  $\kappa = 0.40$  are considered insufficient, while values of  $\kappa = 0.60$ - $0.80$  indicate good agreement, and  $\kappa > 0.80$  represents excellent agreement.

Similar concerns were noted in rating concordance, with Kendall's W not exceeding 0.32 ( $p < 0.01$ ), reflecting substantial subjectivity in incident evaluation. These findings align with earlier warnings from Anderson and Wilson (Anderson and Wilson, 1997) regarding CIT's vulnerability to categorization bias in multistakeholder service environments.

Recognizing these challenges, this study proposes a multidimensional approach to reliability assessment for the *Critical Incident Technique* in transportation service research. Instead of limiting analysis to a single coefficient, as has been common in many previous analyses, we propose the application of a complementary set of four psychometric indicators:

- **Fleiss' Kappa ( $\kappa$ )** - for categorical classification agreement of incidents
- **Kendall's W coefficient** - for ranking agreement of incident importance
- **Intraclass Correlation (ICC)** - for absolute agreement of numerical ratings
- **Cohen's Weighted Kappa ( $\kappa_w$ )** - for agreement considering rating gradations.

The following Table 2 presents a detailed compilation of the proposed reliability measures along with their mathematical formulas and application areas in CIT research.

**Table 2.** *Proposed multidimensional reliability assessment framework for CIT*

| Measure                | Formula                                                     | Application                                                  | Source                    |
|------------------------|-------------------------------------------------------------|--------------------------------------------------------------|---------------------------|
| Fleiss' Kappa          | $\kappa = (P_o - P_e) / (1 - P_e)$                          | Multi-rater agreement in categorical incident classification | (Fleiss, 1971)            |
| Kendall's W            | $W = 12S[k^2(n^3 - n)]$                                     | Concordance of incident importance rankings among raters     | Kendall (Fleiss, 1971)    |
| Intraclass Correlation | $ICC = (MSR - MSE) / MSK$                                   | Absolute agreement in incident intensity assessment          | (Shrout and Fleiss, 1979) |
| Weighted Kappa         | $\kappa_w = 1 - \sum w_{ij} P_{ij} / \sum w_{ij} P_{ij(e)}$ | Agreement considering the severity of disagreements          | (Cohen, 1968)             |

**Notes (legend):**  $P_o$  = observed agreement between raters,  $P_e$  = expected agreement under chance assumption,  $S$  = sum of squared deviations of ranks from the mean,  $k$  = number of raters,  $n$  = number of rated objects,  $MSR$  = mean square between rows,  $MSE$  = mean square error,  $w_{ij}$  = weight assigned to cell (i,j),  $P_{ij(e)}$  = safety (ws),

- information provision (wi),
- interpersonal interaction (wp),
- service recovery after failure (wr).

**Source:** Authors' work.

The proposed formulas for incident weighting are summarized below in Table 3, which outlines the key service dimensions, corresponding calculation methods, and their theoretical underpinnings.

**Table 3.** Incident weighting framework for transportation service evaluation

| Service dimension     | Weighting formula | Interpretation                                                    | Reference                               |
|-----------------------|-------------------|-------------------------------------------------------------------|-----------------------------------------|
| Safety incidents      | $ws = s * f$      | Incident weight = severity × frequency of occurrence              | (Parasuraman, Zeithaml and Berry, 1985) |
| Information provision | $wi = c * t$      | Information weight = communication clarity × delivery timeliness  | (Zeithaml, Parasuraman and Berry, 1990) |
| Interpersonal quality | $wp = r * e$      | Personal weight = staff responsiveness × empathy level            | (Grönroos, 1984)                        |
| Service recovery      | $wr = sp * ef$    | Recovery weight = problem response speed × solution effectiveness | (Smith, Bolton and Wagner, 1999)        |

**Notes (legend):**  $s$  = incident severity (0-1 scale),  $f$  = frequency of occurrence (0-1 scale),  $c$  = communication clarity (0-1 scale),  $t$  = delivery timeliness (0-1 scale),  $r$  = staff responsiveness (0-1 scale),  $e$  = empathy level (0-1 scale),  $sp$  = problem response speed (0-1 scale),  $ef$  = solution effectiveness (0-1 scale), Symbol "×" denotes mathematical multiplication

**Source:** Authors' work.

Each variable in these formulas is normalized on a 0-1 scale to facilitate comparative assessment across different dimensions. Through the incorporation of this model, researchers can prioritize interventions not only based on frequency but also incident severity, aligning methodological design with actual passenger sensitivity to various aspects of transportation services.

### 4.3 Managerial Implications for Quality Management in Rail Transport

The findings derived from the application of the Critical Incident Technique (CIT) provide meaningful insights for the management of service quality in rail transport. First, the study demonstrates that CIT can effectively complement established quality management systems, such as SERVQUAL, ISO 9001 or the PDCA cycle, by capturing unique aspects of passenger experiences that are often overlooked in traditional survey-based approaches. In this respect, CIT enables managers to identify the critical “moments of truth” that shape overall perceptions of service quality.

Equally important are the implications for frontline staff management. The dual image of conductors, perceived either as engaged professionals or passive ticket inspectors, underscores the strategic need for systematic training in soft skills such as empathy, communication and conflict resolution.

Linking these competences with performance appraisal systems and incentive mechanisms would allow organizations to strengthen the role of staff as ambassadors of service quality.

The study also highlights the role of CIT in supporting the development of a customer-oriented organizational culture. By systematically analysing incidents of both positive and negative service encounters, organizations can reinforce behaviours that enhance customer satisfaction while addressing practices that undermine the perceived quality of services. Such an approach is consistent with the principles of Total Quality Management, where every employee contributes to the achievement of quality excellence.

Furthermore, the proposed approach to incident weighting, based on severity and frequency, offers managers a practical tool for prioritizing quality improvement initiatives. Concentrating resources on areas that exert the strongest influence on customer satisfaction, such as communication during delays or effective service recovery, enables more efficient allocation of organizational efforts.

Overall, the integration of CIT into the practice of quality management allows for more precise monitoring of service delivery, facilitates systematic staff development, and supports the strategic prioritization of improvement actions. In doing so, it contributes not only to the enhancement of passenger satisfaction but also to the strengthening of the competitive position of rail operators in the contemporary service economy.

## **5. Conclusions, Proposals, Recommendations**

This study confirmed the usefulness of the Critical Incident Technique (CIT) for evaluating rail transport services and demonstrated its ability to capture both the strengths and weaknesses of service delivery.

The findings revealed a dual image of train conductors: while some are perceived as engaged professionals showing empathy, initiative, and effective problem-solving, others appear passive, limiting their role to ticket inspection. This duality has a direct impact on passengers' perceptions of PKP Intercity's service quality and underlines the conductor's role as a key ambassador of the company.

Empirical results showed that positive service experiences were primarily associated with clear communication during delays, assistance with travel connections, and appropriate responses to disruptions, whereas negative experiences stemmed from a lack of engagement, insufficient conflict management, and limited communication regarding service shortcomings. These observations emphasize the need to strengthen frontline staff competences, particularly in communication and customer interaction, as a central element of quality management in rail transport.

Beyond the empirical contribution, the study advances CIT methodology by proposing an integrated framework that enhances its reliability and applicability in complex service environments. The multidimensional approach to interrater reliability, the introduction of weighted incident classification grounded in service quality theory, and the adoption of dynamic consensus-building strategies together represent a significant step forward in adapting CIT for transportation research.

These methodological innovations allow for more robust assessment of service quality by accounting for the frequency, severity, and contextual importance of service encounters.

The integration of these elements demonstrates the potential of CIT not only as a diagnostic research tool but also as a practical instrument for continuous improvement in service organizations. The findings provide valuable managerial implications, particularly for the development of staff training programs, the prioritization of quality improvement initiatives, and the design of customer experience strategies in line with principles of Total Quality Management and continuous improvement.

In conclusion, this research contributes both theoretically and practically by demonstrating the utility of CIT in capturing critical dimensions of service quality in regional rail transport and by offering an enhanced methodological framework for its application. Future studies should focus on testing the framework across different transportation modes and organizational contexts, which would further strengthen its validity and support its broader use in service quality management.

## **References:**

- Anderson, L., Wilson, S. 1997. Critical incident technique. In: D.L. Whetzel, G.R. Wheaton (Eds.), *Applied measurement methods in industrial psychology*. Davies-Black Publishing, pp. 89-112.
- Butterfield, L.D., Borgen, W.A., Amundson, N.E., Maglio, A.S.T. 2005. Fifty years of the critical incident technique: 1954-2004 and beyond. *Qualitative Research*, 5(4), pp. 475-497. <https://doi.org/10.1177/1468794105056924>.
- Callan, R.J. 1998. The critical incident technique in hospitality research: an illustration from the UK lodge sector. *Tourism Management*, 19(1), pp. 93-98. [https://doi.org/10.1016/S0261-5177\(97\)00078-2](https://doi.org/10.1016/S0261-5177(97)00078-2).
- Chell, E. 2004. Critical Incident Technique. In: *Essential Guide to Qualitative Methods in Organizational Research*. SAGE Publications Ltd. pp. 45-60. <https://doi.org/10.4135/9781446280119.n5>.
- Chell, E., Pittaway, L. 1998. A study of entrepreneurship in the restaurant and café industry: exploratory work using the critical incident technique as a methodology. *International Journal of Hospitality Management*, 17(1), pp. 23-32. [https://doi.org/10.1016/S0278-4319\(98\)00006-1](https://doi.org/10.1016/S0278-4319(98)00006-1).
- Cohen, J. 1968. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4), pp. 213-220. <https://doi.org/10.1037/h0026256>.

- Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., Petticrew, M. 2013. Developing and evaluating complex interventions: The new Medical Research Council guidance. *International Journal of Nursing Studies*, 50(5), pp. 587-592. <https://doi.org/10.1016/j.ijnurstu.2012.09.010>.
- Davis, P.J. 2006. Critical incident technique: a learning intervention for organizational problem solving. *Development and Learning in Organizations: An International Journal*, 20(2), pp.13-16. <https://doi.org/10.1108/14777280610645877>.
- Deaton, A., Cartwright, N. 2018. Reflections on Randomized Control Trials. *Social Science & Medicine*, 210, pp. 86-90. <https://doi.org/10.1016/j.socscimed.2018.04.046>.
- Flanagan, J.C. 1954. The critical incident technique. *Psychological Bulletin*, 51(4), pp. 327-358. <https://doi.org/10.1037/h0061470>.
- Fleiss, J.L. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), pp. 378-382. <https://doi.org/10.1037/h0031619>.
- Gorbaniuk, O. 2016. Wykorzystywanie procedury sędziów kompetentnych w naukach społecznych i możliwości jej oceny psychometrycznej za pomocą narzędzi dostępnych w Statistica. StatSoft Polska.
- Grace, D. 2007. How Embarrassing! An Exploratory Study of Critical Incidents Including Affective Reactions. *Journal of Service Research*, 9(3), pp. 271-284. <https://doi.org/10.1177/109467050700900305>.
- Grönroos, C. 1984. A Service Quality Model and its Marketing Implications. *European Journal of Marketing*, 18(4), pp. 36-44. <https://doi.org/10.1108/EUM00000000004784>.
- Herndon, S.L., Kreps, G.L., eds. 2001. *Qualitative Research: applications in organizational communication*. 2ed ed. Communication and social organization. Creskill, NJ, Hampton Press.
- Holloway, B.B., Beatty, S.E. 2008. Satisfiers and Dissatisfiers in the Online Environment: A Critical Incident Assessment. *Journal of Service Research*, 10(4), pp. 347-364. <https://doi.org/10.1177/1094670508314266>.
- Kendall, M.G. 1948. Rank correlation methods. Griffin.
- Koch, A., Strobel, A., Guler, K., Karl, W. 2009. Quality of the Critical Incident Technique in practice: Interrater reliability and users' acceptance under real conditions. *Psychology Science Quarterly*, 51.
- Kostamo, K., Jallinoja, P., Vesala, K.M., Araújo-Soares, V., Sniehotta, F.F., Hankonen, N. 2019. Using the critical incident technique for qualitative process evaluation of interventions: The example of the "Let's Move It" trial. *Social Science & Medicine*, 232, pp. 389-397. <https://doi.org/10.1016/j.socscimed.2019.05.014>.
- Kreps, G.L. 2017. Critical Incident Technique. In: J. Matthes, C.S. Davis, R.F. Potter, eds. *The International Encyclopedia of Communication Research Methods*, 1st edn. Wiley, pp. 1-3. <https://doi.org/10.1002/9781118901731.iecrm0052>.
- Lin, W.B. 2006. Correlation Between Personality Characteristics, Situations of Service Failure, Customer Relation Strength and Remedial Recovery Strategy. *Services Marketing Quarterly*, 28(1), pp. 55-88. [https://doi.org/10.1300/J396v28n01\\_04](https://doi.org/10.1300/J396v28n01_04).
- Moore, G.F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., Moore, L., O'Cathain, A., Tinati, T., Wight, D., Baird, J. 2015. Process evaluation of complex interventions: Medical Research Council guidance. *BMJ*, 350(mar19 6), pp. h1258-h1258. <https://doi.org/10.1136/bmj.h1258>.
- Oliver, R.L. 2014. *Satisfaction: A Behavioral Perspective on the Consumer*. Routledge. <https://doi.org/10.4324/9781315700892>.
- Parasuraman, A., Zeithaml, V.A., Berry, L.L. 1985. A Conceptual Model of Service Quality

- and Its Implications for Future Research. *Journal of Marketing*, 49(4), pp. 41-50. <https://doi.org/10.1177/002224298504900403>.
- Rodrigues, A.M., O'Brien, N., French, D.P., Glidewell, L. and Sniehotta, F.F., 2015. The question-behavior effect: Genuine effect or spurious phenomenon? A systematic review of randomized controlled trials with meta-analyses. *Health Psychology*, 34(1), pp.61-78. <https://doi.org/10.1037/hea0000104>.
- Rothman, A.J., Bartels, R.D., Wlaschin, J., Salovey, P. 2006. The Strategic Use of Gain- and Loss-Framed Messages to Promote Healthy Behavior: How Theory Can Inform Practice. *Journal of Communication*, 56(suppl\_1), pp. S202-S220. <https://doi.org/10.1111/j.1460-2466.2006.00290.x>.
- Sallis, J.F., Calfas, K.J., Alcaraz, J.E., Gehrman, C., Johnson, M.F. 1999. Potential mediators of change in a physical activity promotion course for university students: Project grad. *Annals of Behavioral Medicine*, 21(2), pp. 149-158. <https://doi.org/10.1007/BF02908296>.
- Shrout, P.E., Fleiss, J.L. 1979. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), pp. 420-428. <https://doi.org/10.1037/0033-2909.86.2.420>.
- Smith, A.K., Bolton, R.N., Wagner, J. 1999. A Model of Customer Satisfaction with Service Encounters Involving Failure and Recovery. *Journal of Marketing Research*, 36(3), pp. 356-372. <https://doi.org/10.1177/002224379903600305>.
- Stauss, B., Weinlich, B. 1997. Process-oriented measurement of service quality: Applying the sequential incident technique. *European Journal of Marketing*, 31(1), pp. 33-55. <https://doi.org/10.1108/03090569710157025>.
- Stewart, B., Smith, A.C.T. 2014. The significance of critical incidents in explaining gym use amongst adult populations. *Qualitative Research in Sport, Exercise and Health*, 6(1), pp. 45-61. <https://doi.org/10.1080/2159676X.2013.766814>.
- Weber, R.P. 1990. Basic content analysis. Quantitative applications in the social sciences. Newbury Park: Sage.
- Yeoman, A., Cooper, J., Urquhart, C., Tyler, A. 2001a. Evaluating virtual outreach services: The VIVOS project experience. In: EAHIL Workshop. Alghero, Sardinia.
- Yeoman, A., Cooper, J., Urquhart, C., Tyler, A. 2001b. The Value and Impact of Virtual Outreach Services: Report of the VIVOS Project.
- Zeithaml, V., Parasuraman, A.P., Berry, L. 1990. Delivering quality service: Balancing customer perceptions and expectations. The Free Press, New York.