

Anlage 1

Qualitätsmesssystem für den bayerischen Schienenpersonennahverkehr (ab 2014)

Stand: Januar 2020

Allgemeines

Das Qualitätsmesssystem für den bayerischen Schienenpersonennahverkehr ersetzt das bisherige Qualitätsmesssysteme für Wettbewerbsstrecken (Regelbetrieb 2008/2009 – 2013) und das Messsystem, das gemäß dem Verkehrsdurchführungsvertrag mit DB Regio von 2005 bis 2013 für die von DB Regio erbrachten Verkehrsleistungen angewendet wurde. Das Qualitätsmesssystem ab 2014 orientiert sich an dem bisherigen Qualitätsmesssystem für Wettbewerbsstrecken und umfasst Fahrgastbefragungen (Kundenbefragungen), offene Tests in Zügen, verdeckte Tests in Zügen und Kundencentern (Beschwerdemanagement) und offene Tests an Stationen.

Grundlage des Qualitätsmesssystems ist die europäische Norm DIN EN 13816 „Transport, Logistik und Dienstleistungen – öffentlicher Personenverkehr – Definition und Festlegung von Leistungen und Messung der Dienstleistungsqualität.“

Im Qualitätsmesssystem für den bayerischen SPNV sind alle im Wettbewerb vergebenen Netze/Strecken sowie die im Rahmen des großen Verkehrsdurchführungsvertrages mit DB Regio erbrachten Leistungen integriert.

Durchführendes Unternehmen und Stand der Projektumsetzung

Mit der Durchführung des Projektes wurde die Firma Target Group GmbH (Dormitz) beauftragt, die auch das bisherige QMS für Wettbewerbsnetze konzipiert und bis Ende 2013 durchgeführt hat. Die Projektvergabe erfolgte im Rahmen eines europäischen Ausschreibungsverfahrens. Die Laufzeit beträgt 10 Jahre und endet im Dezember 2023. 2023 wird die Durchführung des Messsystems neu vergeben.

Erhebungsumfänge und -methode

Pro Erhebungszeitraum und Netz werden 200 offene und 200 verdeckte Tests (in sehr kleinen Netzen 100 verdeckte Tests) sowie 1.000 Fahrgastbefragungen durchgeführt. Neu ab 2016 (Kalibrierungsphase) bzw. 2017 (Regelphase) ist die Durchführung von 100 verdeckten Tests zum Thema Beschwerdemanagement je Netz und Erhebungszeitraum. Die Erhebungen sind gleichmäßig über die gesamte Erhebungsperiode (d. h. über ein bzw. zwei Jahre) verteilt. Die Verteilung der Erhebungen auf die jeweiligen Strecken (bei Netzen mit mehreren Strecken), Tageszeiten und Wochentage erfolgt in Abhängigkeit von den Fahrgastzahlen. Soweit Besonderheiten bestehen, z.B. hinsichtlich Fahrzeugeinsatz, wird kontrolliert, dass diese in der Stichprobe nicht übermäßig oder zu wenig repräsentiert sind. Erhoben werden sowohl die erste als auch die zweite Klasse.

Bei den Stationstests in Bayern werden innerhalb von 2 Jahren alle Stationen einmal erhoben (ca. 500 Stationen pro Jahr). Die Schichtung der Stichprobe erfolgt nach Bahnhofsmagements, Tages-, Wochen-, und Jahreszeiten, Netze/EVU, Größe der Stationen.

Alle Erhebungsarten werden mit „Handhelds“ durchgeführt, d.h. die Tester geben ihre Ergebnisse in die Handhelds ein.

Prüfpersonal und Qualitätssicherung der Ergebnisse

Die Qualitätssicherung der Erhebungsergebnisse erfolgt über Paralleltests, verdeckte Kontrollen und Plausibilitätsprüfungen.

Qualitätsdatenbank

Die Ergebnisse der Erhebungen fließen in eine Qualitätsdatenbank. Jedes beteiligte Unternehmen erhält einen Zugang zur Datenbank (Zugang zu einem zentralen Server über das Internet) und zu den für sein Unternehmen relevanten Daten. Für die offenen Tests, die Fahrgastbefragungen und die Stationstests sind standardisierte Auswertungen (in Tabellen- und graphischer Form) sowie sämtliche Rohdaten abrufbar, für die verdeckten Tests standardisierte Daten sowie solche Rohdaten, die keine Rückschlüsse auf einzelne Mitarbeiter der EVU ermöglichen.

Anreizsystem (Version Stand Januar 2020, Änderungen vorbehalten)

1. Folgende Kategorien (in nachstehender Tabelle fett dargestellt) fließen in das Anreizsystem ein. Sie werden durch definierte Indikatoren (in nachstehender Tabelle kursiv dargestellt) bestimmt. Die Erhebung der Messwerte erfolgt über drei Methoden: offene Tests, verdeckte Tests und Fahrgastbefragungen (ab 2017 zusätzlich verdeckte Tests zum Beschwerdemanagement). Die Ergebnisse der Stationstests fließen nicht in das Anreizsystem. Über welche Erhebungsmethode bzw. anteilige Kombination von Methoden die einzelnen Indikatoren ermittelt werden, zeigt die nachstehende Tabelle:

Kategorien / Indikatoren	offen	verdeckt	Befragung
Sauberkeit Zug außen			
<i>Wagen sauber (Freiheit von Schmutz)</i>	67%	0%	33%
<i>Wagen frei von Graffiti (sichtbare Seite)</i>	100%	0%	0%
Sauberkeit Zug innen			
<i>Sauberkeit Zug innen</i>	67%	0%	33%
<i>WC/ Sanitäreinrichtung</i>	67%	0%	33%
Fahrgastinformationen im Zug (Regelfall)			
<i>„Fahrgastinformationen im Regelfall“ (Zug) (Zugzielanzeiger, Liniennetzplan, optische Haltestelleninformation)</i>	67%	0%	33%
<i>Stationsansagen</i>	0%	67%	33%
<i>Ansage Ausstiegsseite</i>	0%	67%	33%
<i>Umsteige-Durchsagen vor Knotenbahnhof</i>	0%	67%	33%
<i>Verabschiedungs-Durchsagen vor Ausstieg</i>	0%	67%	33%
Fahrgastinformation bei Störungen			
<i>Infos im Zug</i>	0%	0%	100%
Funktionsfähigkeit der Ausstattung der Fahrzeuge			
<i>Innentüren</i>	100%	0%	0%
<i>Außentüren (als defekt beschildert)</i>	100%	0%	0%
<i>WC (als defekt beschildert)</i>	100%	0%	0%
<i>WC-Ausstattung</i>	100%	0%	0%
<i>Temperatur</i>	100%	0%	0%
<i>Müllbehälter (Kapazität nutzbar)</i>	100%	0%	0%
<i>Fenster ohne Kratzer</i>	100%	0%	0%
<i>Trennscheiben ohne Kratzer</i>	100%	0%	0%
<i>Funktionsfähigkeit Fahrscheinautomat im Zug</i>	100%	0%	0%
Verhalten Zugbegleiter / Service			
<i>Präsenz des Kundenbetreuers im Zug</i>	0%	100%	0%
<i>Präsenz des Kundenbetreuers an den Stationen</i>	0%	100%	0%
<i>Durchführung der Fahrscheinkontrolle</i>	0%	100%	0%
<i>Erscheinungsbild allgemein</i>	0%	67%	33%
<i>Sozial-kommunikative Kompetenz</i>	0%	67%	33%
<i>Fachliche Kompetenz</i>	0%	67%	33%
Kundenorientierung bei Beschwerden (allg.)			
<i>Erreichbarkeit/Verfügbarkeit/Wartezeit</i>	0%	67%	33%
<i>Sozial-kommunikative Kompetenz</i>	0%	67%	33%
<i>Fachliche Kompetenz</i>	0%	67%	33%

Das Bonus-Malus-System bezieht nicht alle im Rahmen der Messungen erfassten Indikatoren ein. Nicht enthalten sind z.B. Merkmale, die bereits über andere Vertragsmechanismen pönalisiert werden (z.B. Pünktlichkeit) und Merkmale, bei denen die Stichprobe und damit die statistische Sicherheit zu gering ist (z.B. Informationen im Störfall bei den offenen und verdeckten Tests). Die enthaltenen Indikatoren sind der beigefügten Gesamtübersicht/Excel-Tabelle zu entnehmen.

Die Indikatoren der offenen Tests werden in einer 100%-Skala (100=bester Wert, 0=schlechtester Wert), die Indikatoren der verdeckten Tests und der Fahrgastbefragungen in einer 3er- bzw. 5er-Skala (1=bester Wert, 5=schlechtester Wert) erfasst. In der Datenbank werden sämtliche Werte (aller Instrumente) auf einer 100%-Skala (100=bester Wert, 0=schlechtester Wert) dargestellt, d.h. skalierte Werte werden entsprechend umgewandelt.

2. Die Gewichtung der Qualitätskriterien ist im Grundsatz einheitlich für alle einbezogenen Netze. Es erfolgt lediglich eine Differenzierung im Hinblick auf das Kriterium „Verhalten Zugbegleiter/Service“, da hier unterschiedliche Vorgaben zu Zugbegleiterquoten bestehen und eine hohe Gewichtung des Kriteriums bei nur geringer Zugbegleiterquote nicht angemessen wäre. **Außerdem wird ab 2020 noch nach Vertriebsmöglichkeit im Fahrzeug mittels Automaten differenziert (Typ 1a bzw. 2a). Diese Differenzierung findet jedoch nicht auf der Kategorienebene statt, sondern lediglich auf der Indikatorebene innerhalb der Kategorie Funktionsfähigkeit der Ausstattung.** Definiert sind demnach zwei Netztypen: Netze mit einer Zugbegleiterquote von mehr als 50% (Typ 1 bzw. 1a) und Netze mit einer Zugbegleiterquote von 50 % und weniger (Typ 2 bzw. 2a). In Grenzfällen erfolgt die Zuordnung zum jeweiligen Netztyp durch die BEG. Der nachstehenden Tabelle sind die Gewichtungsverteilungen zu entnehmen:

Kategorie	Netztyp 1(a)	Netztyp 2(a)
Sauberkeit Zug außen	9%	10%
Sauberkeit Zug innen	22%	25%
Fahrgastinformationen im Zug (Regelfall)	18%	20%
Fahrgastinformation bei Störungen	4%	5%
Funktionsfähigkeit der Ausstattung der Fahrzeuge/ technische Verfügbarkeit	18%	20%
Verhalten Zugbegleiter / Service	25 %	15%
Kundenorientierung bei Beschwerden (allg.)	4%	5%

Daneben existieren besondere Netztypen für VDV-Netze und S-Bahnen.

3. Basis für die Ermittlung von Bonus- oder Malusansprüchen sind Nullwerte. Diese bestimmen für die jeweiligen Qualitätskriterien und -indikatoren das Qualitätsniveau, bei dem weder ein Malus noch ein Bonus anfällt. Einige Indikatoren wurden als arithmetisches Mittel zu Sammelindikatoren zusammengefasst (z.B. WC-Ausstattung).
4. Der maximale Bonus bzw. Malus ist auf die im Verkehrsvertrag festgelegte Höhe begrenzt. Für die einzelnen Qualitätskriterien bzw. -indikatoren sind jeweils Nullwerte und Zielwerte bzw. Kappungsgrenzen definiert. Aus diesen lassen sich indikatoren- bzw. kategorienspezifische Bonus- bzw. Malusbeträge errechnen. Deren rechnerische Summe kann bei Ausschöpfen aller Bonus- bzw. Malusmöglichkeiten den im Verkehrsvertrag festgelegten Maximalwert um den Faktor 1,5 überschreiten. Der tatsächlich abzurechnende Bonus- oder Malusbetrag ist jedoch in Höhe des im Verkehrsvertrag festgelegten Maximums gedeckelt.

5. Die jeweiligen Nullwerte sind der nachfolgenden Tabelle zu entnehmen:

Kundenbefragung		
Kategorie	Nullwert [100%-Skala]	skaliertes Wert [5er-Skala]
Wagen sauber (Freiheit von Schmutz)	79,38	1,83
Sauberkeit Zug innen (pauschal)	74,73	2,01
WC/ Sanitäreinrichtung	63,45	2,46
„Fahrgastinformationen im Regelfall“ (Zug) (pauschal)	81,50	1,74
Infos im Zug bei Störungen (rechtzeitig/ vollständig/ verständlich)	67,03	2,32
Erscheinungsbild allgemein	85,03	1,60
Hilfsbereitschaft, Freundlichkeit, Kompetenz des Zugbegleiters	81,30	1,75
Kundenorientierung bei Beschwerden	60,75	2,57
Verdeckte Tests (Zug / Kundencenter)		
Kategorie	Nullwert [100%-Skala]	Skaliertes Wert [3er-/5er-Skala]
Stationsansagen	98,00	1,08
Ansage Ausstiegsseite	98,00	1,08
Umsteige-Durchsagen vor Knotenbahnhof	97,80	1,09
Verabschiedungs-Durchsagen vor Ausstieg	97,40	1,10
Präsenz des Kundenbetreuers im Zug	85,53	1,58
Präsenz des Kundenbetreuers an den Stationen	88,40	1,46
Zub führt Fahrscheinkontrollen durch	86,48	1,54
Erscheinungsbild allgemein	76,90	1,46
Sozial-kommunikative Kompetenz	85,95	1,28
fachliche Kompetenz	82,50	1,35
Erreichbarkeit/Verfügbarkeit/Wartezeit (Beschwerdemanagement)	84,35	1,31
Sozial-kommunikative Kompetenz (Beschwerdemanagement)	93,39	1,13
fachliche Kompetenz (Beschwerdemanagement)	82,40	1,35
offene Tests		
Kategorie	Nullwert [100%-Skala]	-
Wagen sauber (Freiheit von Schmutz)	77,52	-
Wagen frei von Graffiti (sichtbare Seite)	98,00	-
Sauberkeit Sitzplätze	90,11	-
Fußboden ohne losen Schmutz	89,99	-
Fußboden ohne festen, feuchten, klebriger Schmutz	95,91	-
Sauberkeit Fenster	77,94	-
Sauberkeit Einstieg	89,46	-
WC/ Sanitäreinrichtung	83,48	-
Zugzielanzeige (außen, seitlich)	98,00	-
Liniennetzplan	98,00	-
Optische Haltestelleninfo (innen)	98,00	-
Innentüren	97,94	-
Außentüren (als defekt beschildert)	98,00	-
WC (als defekt beschildert)	98,00	-
WC-Ausstattung	97,56	-
Temperatur	91,53	-
Müllbehälter (Kapazität nutzbar)	98,00	-
Fenster ohne Kratzer	98,00	-
Trennscheiben ohne Kratzer	98,00	-
Funktionsfähigkeit Fahrscheinautomat IM ZUG	97,59	-

Die Basis für die Festlegung der Nullwerte bilden die Messergebnisse der Testphase im Jahr 2007 sowie der ersten Regelerhebungsperiode der daran beteiligten Netze. Als erste Regelerhebungsperiode gelten die Erhebungsjahre 12/2007-12/2008 und 12/2008 bis 12/2009. Außerdem erfolgte im Jahr 2013 eine Evaluierung des Systems mit dem Ergebnis, dass einzelne Nullwerte zielorientiert angepasst wurden (die aktuellen Werte sind in obiger Tabelle bereits enthalten). Der Nullwert für das neue Modul Beschwerdemanagement wurde im Rahmen der Kalibrierungsphase 2016 festgelegt und ist ebenfalls in obiger Tabelle dargestellt. **Der Nullwert für die Temperatur wurde im Rahmen einer Kalibrierungsphase 2019 neu festgelegt. Dabei wurde neben der Messung der Temperatur im Fahrzeug die aktuelle Außentemperatur in die Berechnung mit einbezogen. Bei hohen Außentemperaturen wird die zulässige maximale Temperatur im Fahrzeug nach oben verschoben.**

Es gelten folgende Bedingungen:

- Falls es gravierende Qualitätsprobleme in den einbezogenen Netzen gibt, können Sonderregelungen getroffen werden.
 - Es besteht die Möglichkeit der Revision z.B. nach drei oder fünf Jahren durch die BEG.
 - Eine spätere optionale Differenzierung nach Netztypen ist möglich.
6. Die Zielwerte werden je Indikator festgelegt beim Mittel zwischen Nullwert und Qualitätsoptimum: $(100\% + \text{Nullwert}) / 2$. Bei hohen Nullwerten wird der Zielwert ggf. nach oben (100%-Skala) bzw. unten (5er-Skala) angepasst, damit der minimale Toleranzbereich (siehe 7.) und die maximale Steigung der Bonus- bzw. Maluskurven (siehe 8.) eingehalten werden. Die untere Kappungsgrenze wird spiegelbildlich dazu gesetzt: $\text{Nullwert} - (\text{Zielwert} - \text{Nullwert})$. Bei den Indikatoren zur Sauberkeit innen (offener Test), wird auf der Malusseite von dieser Vorgehensweise abgewichen. Die Steigung der Maluskurve ist hier um den Faktor 4 reduziert, damit (hohe) Malusbeträge aufgrund struktureller Erschwernisse in diesem Bereich leichter vermieden werden können. Der Toleranzbereich sowie die Bonusseite bleiben davon unberührt.
- Um bei Indikatoren mit sehr niedrigem Nullwert den Wirkungsbereich nicht zu groß und damit die Anreizwirkung nicht zu schwach werden zu lassen, liegt der Zielwert maximal 8 % über dem Nullwert.
7. Die Nullwerte werden mit einem Toleranzbereich versehen, damit die bei einer Stichprobenziehung immer vorhandenen geringfügigen Zufallsschwankungen der Messergebnisse nicht sofort malus- bzw. bonusrelevant werden. Die Breite der Toleranzbereiche orientiert sich an den bei der Testphase ermittelten Konfidenzintervallen: In der Regel beträgt die Deckelung der Breite des Toleranzbereichs bei verdeckten Tests +/- 5 %, bei Kundenbefragung +/- 3 % und bei offenen Tests +/- 3 %.
- Ausnahmen sind die Indikatoren Sauberkeit Zug außen und Sauberkeit WC (Ermittlung durch offene Tests). Hier wird der maximale Toleranzbereich auf +/- 5 % festgelegt.
- Weil die Konfidenzintervalle im hohen Qualitätsbereich kleiner als diese Regeltoleranzbereiche werden und das System hier auch Wirksamkeit entfalten soll, verjähern sich dort die Toleranzbereiche auf +/- 20% der Differenz aus Skalenende und Nullwert zzgl. 0,6. Untergrenze für die Toleranzbereiche ist +/- 1.
8. Da bei Indikatoren mit einem sehr hohen Nullwert der Wirkungsbereich des Bonus-Malus-Systems sehr eng und schon bei geringen Qualitätsänderungen der maximale Bonus oder Malus erreicht würde, ist die Steigung der Bonus- bzw. Maluskurve so begrenzt, dass zwischen dem äußeren Rand des Toleranzbereichs (noch kein Bonus bzw. Malus) und dem jeweiligen Zielwert (voller Bonus bzw. Malus) eine Qualitätsdifferenz von mindestens 1 (100%-Skala) bzw. 0,04 (5er-Skala) liegt. Dabei verschie-

ben sich ggf. der Zielwert und die untere Kappungsgrenze bei Nullwerten größer 94,67% (bzw. kleiner 1,21 auf der 5er Skala) jeweils etwas nach oben bzw. unten. Durch die Begrenzung der Steigung der Bonus- und Maluskurven wird ein Mindestwirkungsspektrum sichergestellt.

9. Qualitätsveränderungen wirken sich nach Überschreitung des Toleranzbereiches bis zum Ende des Wirkungsbereiches prinzipiell linear auf den damit verbundenen Bonus/Malus aus. Innerhalb des Malusbereiches tritt jedoch bei Unterschreitung eines Schwellenwertes (Qualitätsniveau von 2/3 des möglichen Malus), der eine untere Akzeptanzgrenze für schlechte Qualität markiert, ein zusätzlicher Abschlag zur Erhöhung der Anreizwirkung ein. Der Abschlag greift nur, wenn der Messwert 92% unterschreitet (bzw. 1,32 auf der 5er-Skala überschreitet).

Der Abschlag je Indikator unter der Akzeptanzgrenze richtet sich nach dessen Gewichtung und beträgt 2%, 3 % oder 4 % bezogen auf die vertragliche Deckelungssumme, wie folgt:

Gewichtung von – bis	Abschlag
bis (einschl.) 2%	2%
bis (einschl.) 3%	3%
über 3%	4%

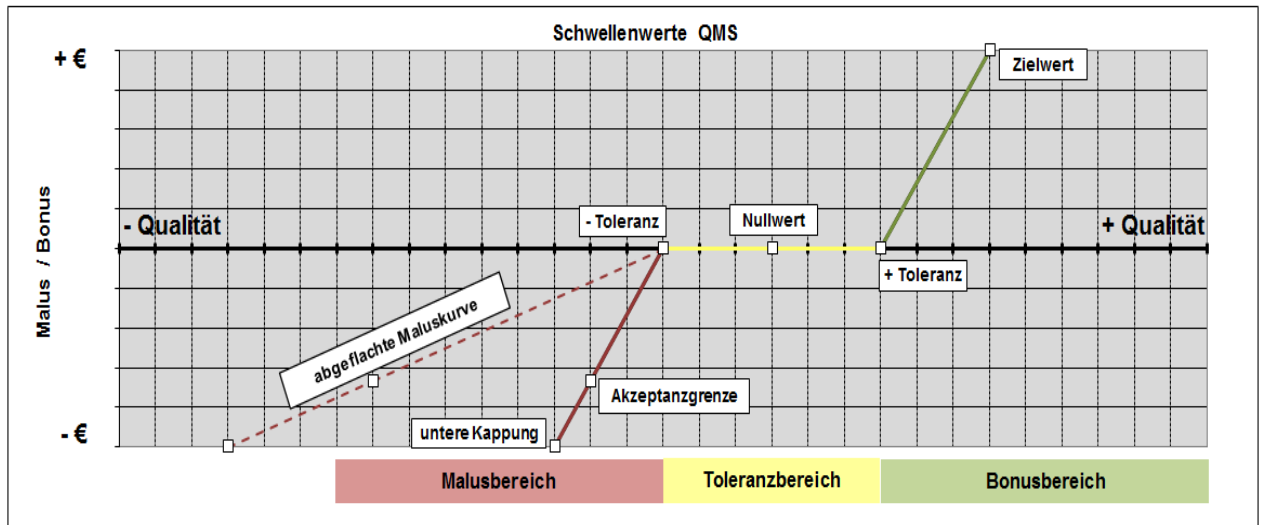
Die jeweiligen Gewichte der einzelnen Indikatoren variieren je nach Netztyp (siehe 2.). Dadurch können auch dieselben Indikatoren in Abhängigkeit des Netztyps in verschiedenen Abschlagsklassen eingruppiert sein. Zur Übersicht der Gewichte der Indikatoren sowie des jeweils zugeordneten Abschlagsgewichts siehe Netztypentabelle.

Eine Überschreitung der Deckelungssumme ist auch mit Schlechtleistungs-Abschlägen nicht möglich.

Abschläge werden nicht bei Ergebnissen der Fahrgastbefragungen fällig.

10. Die Qualitätserhebungen werden kontinuierlich durchgeführt. Die Abrechnung erfolgt bei größeren Netzen ab ca. 2,5 Mio. Zugkm. pro Jahr jährlich, bei kleineren zweijährlich. Die Festlegung, ob ein Netz einen ein-oder zweijährigen Erhebungsrhythmus hat, trifft die BEG.
11. Revision: Das Qualitätsmesssystem für den bayerischen SPNV ist im Grundsatz als dauerhaftes System angelegt, weshalb periodische Anpassungen nötig werden können. Nachsteuerungen der Nullwerte und Zielwerte behält sich die BEG daher vor. Bei Bedarf erfolgt eine Anpassung der Nullwerte unter Berücksichtigung der Entwicklung der durchschnittlichen Qualitätsentwicklung je Indikator. Perspektivisch muss sich die BEG auch die Änderung anderer Parameter des Systems und den Ersatz des Systems durch ein neues System vorbehalten.

Abbildung: Funktionsweise des Anreizsystems (Prinzipiskizze)



Abgeflachte Maluskurve relevant für die Kriterien

- Sauberkeit Einstieg
- Sauberkeit Fußboden
- Sauberkeit Sitzplätze
- Sauberkeit WC

(alles offener Test)