MOVING FORWARD WITH REGIOTAXI

A BENCHMARK OF PERFORMANCE AND AN EVALUATION OF SEVERAL DRT SERVICES IN THE NETHERLANDS

Master Thesis for Transportation Engineering & Management

R.P.C.Buysse

January 2014

adviseurs mobiliteit
Goudappel Coffeng

UNIVERSITEIT TWENTE.

Goudappel Coffeng

Onderzoek & Ontwikkeling
Snipperlingsdijk 4
7417 BJ Deventer
Postbus 161, 7400 AD Deventer

Mentors company:
Ing. J. Boxum
Drs. W. Korver

Universiteit Twente

Master: Civil Engineering and Management

Track: Traffic Engineering and Management

Faculty: Engineering Technology (CTW)

Drienerlolaan 5
7522 NB Enschede
www.utwente.nl/cem

Mentors university:
Dr. T. Thomas
Dr. K. Geurs

Author

R.P.C.Buysse
Smedenstraat 72d
7411 RG Deventer
rpcbuysse@gmail.com

Summary

This study is on the performance of Regiotaxi in the Netherlands. Eleven Regiotaxi services were benchmarked on a set of indicators and subsequently evaluated. Also, the transport authorities were asked to fill out a survey on their policy priorities and the relation between their performance and their policy priorities was studied.

The study was done for mobility consultancy Goudappel Coffeng. They had customer satisfaction surveys for different Regiotaxi services available and they wanted to benchmark the surveys and get an insight on customer satisfaction across regions. This was expanded to a full evaluation of in total 11 Regiotaxi services. Beside the customer satisfaction surveys, trip databases and management reports were also used for the evaluation. The main objective of the study was:

To assess the performance of Regiotaxi in the Netherlands, by (a) benchmarking performance indicators and finding internal relations between them, (b) evaluating different Regiotaxi systems and (c) determining the effectiveness of policy, given the policy objectives of the transport authorities.

Regiotaxi is a Dutch demand responsive transport (DRT) system for Wmo-indicated (Wmo is the law on social support) and it can be combined with other functions, such as public, school or work transport. A literature study was done for DRT and the evaluation of DRT. This literature study revealed that research has mostly focused on the financial effectiveness of DRT, for example, the effective scheduling of vehicles and cost structures. Therefore it was chosen to focus on evaluating performance in a broader sense, using the trip databases. An evaluation framework from Andrade (2008) was adapted to do so.

In order to get context for the study and to validate the adapted evaluation framework, interviews were conducted with a policy advisor, a policy maker, a director of a transportation company and a representative of public transport travelers.

The evaluation framework uses two levels of objectives to generate a single total score. The higher level objectives are perception, performance and economic durability. In the lower level these objectives are split into lower level objectives. These lower level objectives are described by indicators. For example, one of the lower level objectives describing perception is perceived safety. Perceived safety is then scored using the satisfaction about safety and driving style that were taken from the customer satisfaction survey.

The framework combines two multi-criteria decision algorithms to generate a final score: the Analytical Hierarchy Process (AHP) and the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS). AHP was used to set weights for the objectives on the different hierarchy levels. In the study policy priorities were used to weight the objectives and AHP made the setting of priorities less complex for the transport authority by reducing it to a series of one-on-one comparisons. The transport authority had to indicate to what extent one or the other objective had a higher priority. The result matrix could then be converted to a set of weights for the compared objectives.

Another advantage of AHP is that human errors can be processed and quantified, i.e. the answers do not have to be perfectly consistent and an . This was preferred since for comparing multiple objectives it would be very difficult to be perfectly consistent. So instead of forcing perfect consistency, some small amount of errors was allowed, this was quantified with a consistency ratio. When the consistency ratio would be too high, it would be requested the answers were revisited to make them more consistent. In the end all surveys were sufficiently consistent. In general, on the higher level hierarchy perception was deemed most important. For perception, perceived safety was most important and perceived comfort least important. For performance, reliability was by far considered most important.

TOPSIS was chosen because it allowed for a single service to be evaluated, without having to resort to relative ranking of services. This is done by independently setting an ideal and anti-ideal situation and determining how close the service is to both. The ideal and anti-ideal situation would be determined by experts, however, this did not work out in this study as it was advised by experts to use the data available instead of expert opinion as it was expected this would give better values. As a consequence, since the ideal and anti-ideal values were based on the scores of the participating services, the result was still a relative ranking. It was still chosen to continue with TOPSIS since there were no other disadvantages and the ideal and anti-ideal solutions generated in this study can be used to perform future Regiotaxi evaluations in the intended way.

In order to benchmark and evaluate the services, monitoring data had to be made comparable; this was one of the issues tackled in this study. The biggest problem was the variation in questions and survey scales used in the customer satisfaction surveys. Different conversion methods were considered to convert all survey scales – different 4- and 5-point scales – to a 10-point scale. In the end two methods were chosen: For converting the 5-point scales to a 10-point scale, scale profiles were matched. For both 5-point and 10-point scales a profile of usage was made, based on the data available from the customer satisfaction surveys. The profiles were then coupled, for example, if 15% of the answers were the first point on the 5-point scale, the 15% of lowest answers on the 10-point scale were averaged to generate an equivalent scale value for 1 on the 5-point scale for the 10-point scale. For converting the 4-point scales this was not possible, because nearly all questions with a 4-point scale were from the same survey, and therefore not enough variation was available to make a robust scale profile. The 4-point scales were therefore converted using anchored extremities, i.e. 1 on the 4-point scale was 1 on the 10-point scale, 4 was converted to 10, intermediate values were rescaled as if the distance between scale points was equal.

Besides the customer satisfaction surveys, also the data from the trip databases and management reports had to be made comparable. For several indicators a value per kilometer was used, however, some services work with a zone system and the services that register kilometers use different routing programs to generate calculate the trip distance. Therefore, since the kilometers came up for several indicators and the data were so diverse, reference trip distances and travel times were generated using the origin and destination from the trip database in Google maps. The transport authorities delivered subsidy numbers in very different formats. From the different formats a subsidy per passenger kilometer was distilled.

After all the indicator values were benchmarked, they could be used for the evaluation. The evaluation showed that no system scores bad or good across the board, so bad scores in one area are compensated in other areas. This suggests that there is room for the regions to improve by learning from each other.

The most important conclusions from correlating the benchmarks and scores were:

- Tariff and subsidy were positively correlated, meaning that systems with a higher tariff also had higher subsidies.
- Economic durability (the extent to which the system is durable and can go without subsidy) has only a weak correlation with performance and perception. This means that freeing up more money for the system does not necessarily lead to a better performing system.
- Travel time performance and travel time perception do not correlate.
- Tariff and perception and negatively correlated, suggesting that people are more than willing to pay for the service they're receiving.
- Perceived safety and perceived comfort strongly correlate. This result combined with remarks made in the interviews bring up the question whether travelers are able to correctly estimate their safety and that when surveyed on this subject, they are not actually reporting perceived comfort.

- The average trip distance is negatively correlated with trips per inhabitant. There are several possible explanations for this and it is recommended to get a better understanding of this relation, as it may be relevant for the travel budgets Wmo-travelers receive.
- Travel time satisfaction is the indicator that is correlated the best with overall satisfaction, so it is an important policy focus. However, since actual travel time performance does not correlate with travel time satisfaction, a better understanding of what makes up travel time satisfaction is needed.

The results of the assessment of the effectiveness of policy showed that the priorities for policy objectives were not reflected in the performance level of the policy objectives. Two possible causes for this were hypothesized: a poor translation from policy objectives to requirements for the transport operator or a lack of monitoring of the requirements for the transport operator.

The process and outcomes of the study led to several recommendations, both for the practical organization of Regiotaxi as for future research.

The main recommendation for practice is to start sharing experiences and data in a regular structural manner. In order to do this, data will have to be made more comparable, as both the data from the customer satisfaction surveys as the monitoring data that is gathered is very diverse at the moment. Uniformity can be achieved by creating a guideline, such as MIPOV, which is a guideline for monitoring data for conventional public transport. When more comparable data are available, sharing data and experiences could be done at a national platform for Regiotaxi systems. This will contribute to improving the relation between policy objectives and actual performance. The evaluation framework used in this study can be used for future evaluations, not only for a comparison, but also to evaluate a single system. Another recommendation to improve the relation between policy objectives and actual performance is for the transport authorities to reconsider how policy objectives lead to performance. Specifically the way the tender is set up and how the system is monitored should be examined.

A current topic for DRT in the Netherlands is budget cuts by the government. Two remarks are made about this subject. Firstly, although the results show that in some cases subsidy can potentially be reduced without affecting the performance, it is advised to be careful with cutting budgets. Problems may spread to other systems, for example conventional public transport. Secondly, budget cuts might be realized by reorganizing DRT systems in the Netherlands. At the moment, the division of DRT systems is based on how they are funded. However, there is potentially room to improve efficiency by basing the division on technical requirements. It should be researched whether the efficiency gains are higher than the increased costs of additional bureaucracy for redirecting funding streams.

As for future research, it is recommended to look into the causes of some of the relations found in this research. Also, it would be very useful to develop a valid rescaling method for customer satisfaction survey data. Finally, this research the weighting of objectives was done based on policy priorities, which leads to an evaluation score from the perspective of the transport authority. The policy priorities survey which was done to generate the objectives weights could also be set up for travelers or transportation companies to evaluate the system performance from their perspective.

Samenvatting

Deze studie handelt over de prestatie van Regiotaxi in Nederland. Elf Regiotaxidiensten werden met elkaar vergeleken op een stel indicatoren en vervolgens geëvalueerd. Tevens werden de vervoersautoriteiten gevraagd om een enquête in te vullen met betrekking tot hun beleidsprioriteiten. De relatie tussen hun prestatie en hun beleidsprioriteiten werd bestudeerd.

De studie werd uitgevoerd voor Mobiliteitsadviseurs Goudappel Coffeng. Zij beschikten over klanttevredenheidsonderzoeken van verschillende Regiotaxidiensten en zij wilden een vergelijking van de onderzoeken en tevens wensten zij inzicht te verkrijgen over de klanttevredenheid van de regio's. Dit werd uitgebreid tot een volledige evaluatie van in totaal 11 Regiotaxidiensten. Naast de klanttevredenheidsonderzoeken, werden tevens databases van de ritten en managementrapportages gebruikt voor de evaluatie. De belangrijkste doelstelling van de studie was:

Stel de prestatie van Regiotaxi in Nederland vast door: (a) vergelijking van prestatie-indicatoren en het vinden van interne relaties tussen de indicatoren, (b) evaluatie van de verschillende Regiotaxi-systemen en (c) het vaststellen van de effectiviteit van het beleid, de gegeven beleidsdoelen van de vervoersautoriteiten in aanmerking genomen.

Regiotaxi is een Nederlands systeem voor vraagafhankelijk vervoer (demand responsive transport, DRT) voor Wmo-geïndiceerden (Wmo is de Wet Maatschappelijke Ondersteuning) en dit kan worden gecombineerd met andere functies, zoals openbaar vervoer, leerlingenvervoer of werkvervoer. Er werd een literatuurstudie voor vraagafhankelijk vervoer uitgevoerd, evenals een evaluatie van vraagafhankelijk vervoer. Uit deze literatuurstudie bleek, dat onderzoek zich in het verleden vooral had gericht op de financiële effectiviteit van vraagafhankelijk vervoer, bijvoorbeeld de doelgerichte inroostering van voertuigen en kostenstructuren. Om die reden is ervoor gekozen om deze studie vooral te richten op de evaluatie van de prestatie in een breder perspectief, waarbij gebruik gemaakt werd van de databases van de ritten. Een evaluatieschema van Andrade (2008) werd hiervoor aangepast.

Om de context van de studie verkrijgen en het aangepaste evaluatieschema te bevestigen, werden interviews gehouden met een beleidsadviseur, een beleidsmaker, een directeur van een vervoersbedrijf en een vertegenwoordiger van een reizigersorganisatie voor openbaar vervoer.

Het evaluatieschema maakt gebruik van doelstellingen op 2 niveaus om te komen tot een afzonderlijke totale score. Op het hoge niveau van de hiërarchie zijn de doelstellingen: perceptie, prestatie en economische duurzaamheid. In het lage niveau zijn deze doelstellingen gesplitst in lagere niveau doelstellingen. Deze lagere niveau doelstellingen worden beschreven aan de hand van indicatoren. Bijvoorbeeld: Een van de doelstellingen op lager niveau, die de perceptie beschrijft, is hoe de veiligheid wordt ervaren. De perceptie van veiligheid is vervolgens vastgesteld met gebruik van de tevredenheid over veiligheid en de rijstijl, die werden verkregen uit het klanttevredenheidsonderzoek.

Het schema combineert twee multi-criteria beslissingsalgoritmen om de uiteindelijke score te bepalen: het "Analytical Hierarchy Process" (AHP) en de "Technique for Order of Preference by Similarity to Ideal Solution"(TOPSIS). AHP werd gebruikt om de gewichten vast te stellen voor de doelstellingen van de verschillende hiërarchische niveaus. In de studie werden de beleidsprioriteiten gebruikt om de doelstellingen te wegen en het gebruik van AHP bewerkstelligde, dat het vaststellen van prioriteiten voor de vervoersautoriteiten minder complex werd door het reduceren tot paarsgewijze vergelijkingen. De vervoersautoriteiten moesten aangeven, tot welke hoogte aan het ene of het andere doel een hogere prioriteit werd toegekend. De resultatenmatrix kon vervolgens worden bewerkt naar gewichten voor de vergeleken doelstellingen.

Een ander voordeel van AHP is, dat op deze wijze menselijke fouten in de enquête kunnen worden gekwantificeerd en verwerkt; dit betekent dat de antwoorden moeten niet perfect consistent moeten zijn. Hieraan werd de voorkeur gegeven, omdat het voor de vergelijking van meervoudige doelstellingen erg moeilijk zou zijn om perfect consistente antwoorden te geven. Derhalve werd ervoor gekozen om een klein aantal fouten toe te staan, in plaats van geforceerd perfect consistente antwoorden te verlangen. Dit werd gekwantificeerd met een consistentie-ratio. Als de consistentie-ratio te hoog was, dan werd verzocht om de antwoorden te herzien, zodat de uitkomsten meer consistent zouden zijn.

TOPSIS werd gekozen, omdat daarmee een afzonderlijk Regiotaxi-systeem kon worden geëvalueerd, zonder dat de toevlucht moest worden genomen tot de vergelijking van de verschillende diensten. Hiervoor werden onafhankelijk een ideale en een anti-ideale situatie vastgesteld en bepaald hoe dicht de dienst bij beide uitersten gepositioneerd was. De ideale en anti-ideale situatie diende te worden vastgesteld door experts, maar dit is in deze studie niet gebeurd, omdat de experts adviseerden om de beschikbare data te gebruiken, in plaats van de mening van de experts, aangezien werd verwacht dat dit zou leiden tot betere waarden. Dientengevolge – gezien het feit dat de ideale en anti-ideale waarden werden gebaseerd op de scores van de deelnemende diensten – was het resultaat toch een relatieve vergelijking. Toch is besloten om met TOPSIS te blijven werken, omdat er geen andere nadelen werden vastgesteld en de ideale en anti-ideale waarden, die in deze studie werden gehanteerd, kunnen worden gebruikt om de toekomstige evaluaties van Regiotaxi op de goede manier uit te voeren.

Om de benchmark en de evaluatie van de diensten mogelijk te maken, moesten de monitoringsgegevens vergelijkbaar zijn; dit was een van de problemen, die in deze studie moesten worden opgelost. Het grootste probleem werd gevormd door de grote variatie in vraagstelling en de inschaling van de onderzoeksgegevens, die werden gebruikt in de klanttevredenheidsonderzoeken. Verschillende methodes om deze om te zetten werden onderzocht – de gebruikte schalen varieerden van 4- en 5-puntsschalen tot een 10-puntsschaal. Uiteindelijk werd gekozen voor twee methodes: Om de 5-puntsschaal om te zetten naar een 10-puntsschaal, werden de schaal-profielen met elkaar in overeenstemming gebracht. Voor zowel de 5-puntsschaal als de 10-puntsschaal werd een gebruiksprofiel gemaakt, gebaseerd op de beschikbare gegevens van de klanttevredenheidsonderzoeken. De profielen werden vervolgens gekoppeld. Bijvoorbeeld: als 15% van de antwoorden op de 5-puntsschaal het eerste punt vertegenwoordigden, dan werd 15% van de laagste antwoorden op de 10-puntsschaal gemiddeld om te komen tot een equivalente schaalwaarde voor ‘1’ op de 5-puntsschaal. Dit was niet mogelijk voor de omzetting van de 4-puntsschaal, want bijna alle antwoorden van de 4-puntsschaal kwamen uit hetzelfde onderzoek en er was derhalve niet voldoende variatie beschikbaar om een betrouwbare schaalverdeling te maken. De 4-puntsschaal werd zodoende geconverteerd door het gebruik van verankerde uitersten, dat wil zeggen ‘1’ op de 4-puntsschaal was ‘1’ op de 10-puntsschaal; ‘4’ werd omgezet naar ‘10’ en tussenliggende waarden werden opnieuw ingeschaald alsof de afstanden tussen de schaalpunten gelijk waren.

Naast de klanttevredenheidsonderzoeken, moesten ook de gegevens van de rittendatabases en de managementrapportages vergelijkbaar gemaakt worden. In diverse indicatoren werd een waarde per kilometer gebruikt, maar enkele diensten werken met een zone-systeem en de diensten die de kilometers registreerden, gebruikten verschillende route-programma’s om de ritafstand te berekenen. Als gevolg hiervan waren de ritafstanden in kilometers zo divers, dat is besloten de referentie ritafstanden en reistijden van de ritten om te rekenen door de herkomst- en bestemmingsadressen uit de database te gebruiken in Google maps. De vervoersautoriteiten leverden getallen over subsidie op zeer diverse wijzen. Uit deze verschillende gegevens werd een subsidie per reiskilometer/passagier gegenereerd.

Uiteindelijk konden alle indicatorwaarden worden vergeleken en konden zij worden gebruikt voor de evaluatie.

Nadat een benchmark voor alle indicator-waarden was vastgesteld, konden deze worden gebruikt voor de evaluatie. De evaluatie toonde aan, dat geen van de diensten over de gehele linie slecht of goed scoorde, omdat slechte scores op het ene gebied werden gecompenseerd door goede scores op een ander gebied. Dit geeft aan, dat er voor de regio's ruimte voor verbetering is door van elkaar te leren.

De belangrijkste conclusies uit de correlaties tussen de benchmarks en scores waren:

- Tarieven en subsidie hadden een positieve correlatie, wat inhoudt, dat de systemen met een hoger tarief ook een hogere subsidie hadden.
- Economische duurzaamheid (de mate van duurzaamheid van een systeem en waarin het in stand gehouden kan worden zonder subsidie) heeft enkel een zwakke correlatie met prestatie en perceptie. Dit betekent, dat het besteden van meer geld aan het systeem niet noodzakelijk leidt tot een systeem dat beter presteert.
- De werkelijke reistijd en de perceptie van de reistijd hebben geen correlatie.
- Tarief en perceptie hebben een negatieve correlatie, wat suggereert dat mensen bereid zijn om te betalen voor de service die zij krijgen.
- De perceptie van veiligheid en de perceptie van comfort hebben een sterkte correlatie. Dit resultaat, gecombineerd met opmerkingen die werden gemaakt tijdens de interviews, doen de vraag rijzen of reizigers in staat zijn om op juiste wijze hun veiligheid in te schatten en of zij, wanneer zij dit moeten aangeven in een enquête, niet eigenlijk hun comfortniveau rapporteren.
- De gemiddelde reisafstand is negatief gecorreleerd met het aantal reizen per inwoner. Er zijn hiervoor diverse verklaringen mogelijk en het is aan te bevelen om een beter inzicht in deze relatie te verkrijgen, omdat dit relevant zou kunnen zijn voor de budgetten, die WMO-geïndiceerden ontvangen.
- Tevredenheid over reistijd is de indicator die het beste correleert met de algemene tevredenheid, dus dit is een belangrijke factor voor beleid. Echter, omdat de werkelijke reistijd niet correleert met de tevredenheid over de reistijd, is het nodig om een beter begrip te krijgen van wat van belang is bij de tevredenheid over reistijd.

De resultaten van het onderzoek naar de effectiviteit van het beleid tonen aan, dat de prioriteiten voor beleidsdoelen niet terug te zien zijn in het prestatieniveau van de beleidsdoelen. Twee mogelijke oorzaken hiervoor kunnen worden verondersteld: een slechte vertaling van de beleidsdoelstellingen naar de gestelde eisen voor de vervoerders of een tekort aan toezicht op de invulling van de gestelde eisen.

Het proces en de uitkomsten van de studie leidden tot een aantal aanbevelingen, zowel voor de praktische organisatie van Regiotaxi als voor toekomstig onderzoek.

De belangrijkste aanbeveling voor de praktijk is om ervaring en gegevens te delen op een regelmatige en gestructureerde manier. Om dit te kunnen doen, moeten gegevens vergelijkbaar gemaakt worden. Op dit moment zijn zowel de klanttevredenheidsgegevens als de gegevens uit toezicht op het systeem zeer divers. Uniformering kan worden bereikt door een richtlijn op te stellen, zoals ook is gebeurd in het openbaar vervoer door de invoering van MIPOV, welke een richtlijn is voor welke gegevens nuttig zijn voor het toezicht houden op een conventioneel openbaar vervoerssysteem. Als beter vergelijkbare gegevens beschikbaar komen, dan kunnen gegevens en ervaringen worden gedeeld in een nationaal platform voor Regiotaxidiensten. Dit zal zeker bijdragen aan het verbeteren van de relatie tussen beleidsdoelstellingen en werkelijke prestatie. Een andere aanbeveling om dit te verbeteren is dat de vervoersautoriteiten zouden kunnen heroverwegen hoe beleidsdoelstellingen moeten leiden tot prestatie. Met name de manier waarop de aanbesteding wordt gedaan en hoe het systeem wordt opgevolgd zou moeten worden overwogen.

Een actueel onderwerp voor DRT in Nederland is de bezuinigingen door de regering. Over dit onderwerp worden 2 opmerkingen gemaakt: Allereerst wordt aanbevolen om zorgvuldig om te gaan met het invoeren van

bezuinigingen alhoewel de resultaten aangeven, dat in bepaalde gevallen subsidie mogelijkerwijs zou kunnen worden teruggebracht zonder dat dit de resultaten beïnvloedt. Er zouden problemen kunnen ontstaan in andere diensten, bijvoorbeeld in het conventionele openbare vervoer. Ten tweede kunnen de bezuinigingen mogelijk gerealiseerd worden door een herindeling van vraagafhankelijke vervoerssystemen in Nederland. Momenteel is de indeling van dergelijke systemen gebaseerd op hoe ze gefinancierd worden. Er is echter potentieel ruimte om de efficiëntie te verbeteren door de indeling op technische eisen te baseren. Het wordt aangeraden om te onderzoeken of de winst uit verbeterde efficiëntie opweegt tegen de kosten van de extra bureaucratie voor het herverdelen van de middelen.

Voor toekomstig onderzoek wordt aanbevolen om verder te kijken naar oorzaken voor de verbanden die in dit onderzoek zijn gevonden. Daarnaast zou het bijzonder nuttig zijn om een valide herschalingsmethode te ontwikkelen voor de data uit de klanttevredenheidsonderzoeken. In dit onderzoek is de weging in de evaluatie gedaan op basis van de gerapporteerde beleidsprioriteiten van de vervoersautoriteiten. Ten slotte wordt daarom aanbevolen om deze evaluatie ook uit te voeren vanuit het perspectief van de reiziger en de vervoerder, door de beleidsprioriteitenenquête ook door hen in te laten vullen en hun weging toe te passen.

Dankwoord

Op deze plaats wil ik graag wat mensen bedanken die een aandeel hebben gehad aan het schrijven van dit rapport.

Tom Thomas, mijn dagelijkse begeleider van de UT, wil ik bedanken voor de uitgebreide sessies die we hebben gehad om mijn onderzoek vorm te geven. Vaak waren deze sessies diepgaand en op de een of andere manier kwam ik er altijd met een betere focus en overlopend van ideeën uit. Bovendien maakte hij altijd tijd om te bespreken, zelfs al probeerde ik het soms op wat korte termijn te organiseren.

Verder wil ik mijn begeleiders bij Goudappel Coffeng bedanken. Wim Korver voor de nuttige feedback op mijn deelstukken en voor de hulp bij het leggen van contact met de juiste mensen die ik nodig had om dit onderzoek uit te voeren. Jantine Boxum voor het meedenken over praktische problemen en voor het zorgen dat ik mijn weg binnen het bedrijf kon vinden. Daarnaast natuurlijk ook voor de gezellige kamer om te werken op kantoor.

Mijn hoofdbegeleider van de UT, Karst Geurs, wil ik bedanken voor zijn feedback en voor het niet te gemakkelijk te zijn over mijn eerste versies van het onderzoeksplan. Terugkijkend heeft de extra iteratie voor het onderzoeksplan mijn onderzoeksproces een stuk soepeler laten lopen.

Verder wil ik Tim en mijn ouders bedanken voor de morele ondersteuning en hun reviews. Tim voor de inhoudelijke reviews en de vormgeving van het rapport. Mijn ouders voor het fungeren als klankbord. Ik vind het geweldig om te zien hoe betrokken jullie er bij raakten.

Naast begeleiding en reviews had ik ook bronmateriaal nodig voor deze studie. Daarom wil ik de mensen bedanken die tijd hebben vrijgemaakt om zich te laten interviewen en mij de benodigde informatie en context hebben gegeven om deze studie uit te voeren.

Als laatste wil ik mijn collega's bij O&O bedanken voor de prettige werksfeer. De professionele omgeving heeft me geholpen om tot dit eindproduct te komen.

Roland Buysse
Deventer, januari 2014

TABLE OF CONTENTS

ABBREVIATIONS

AHP	Analytical hierarchy process
Awbz	Algemene wet bijzondere ziektekosten (General law on exceptional medical expenses)
CBS	Central bureau for statistics
CSS	Customer satisfaction survey
DRT	Demand Responsive Transport
TOPSIS	Technique for Order of Preference by Similarity to Ideal Solution
Wmo	Wet maatschappelijke ondersteuning (Law on social support)

1 INTRODUCTION

This study is on the performance of Regiotaxi, a type of demand responsive transport (DRT) system in the Netherlands. It is a master thesis as conclusion of the master Transportation Engineering & Management at the Twente University. The study was performed at mobility consultancy Goudappel Coffeng in Deventer.

Goudappel Coffeng is already involved with Regiotaxi by performing customer satisfaction surveys for the transport authorities. These surveys sparked an interest to benchmark the performance of Regiotaxi systems in the Netherlands. Additionally there's an interest in discovering how different performance indicators are related to each other and to both perceived and actual performance of the Regiotaxi systems. Different customer satisfaction surveys were available at Goudappel Coffeng. However, these surveys were all internal studies from the transport authorities and there was no overarching insight into the different regions into customer satisfaction.

Regiotaxi systems all include transport for Wmo-travelers. Wmo is the Dutch law on social support. This law requires municipalities to organize transportation for people with a Wmo-indication. In many regions in the Netherlands municipalities have collaborated - sometimes also with the province - to organize this transport. These collaborations are established to minimize costs by increasing scale; therefore a lot of systems are also publically available. In collaborations with provinces the public transport part is usually subsidized by the province. Figure 1 shows the area of all the current Wmo systems and the types of transport service that are included in the system. For the largest part these systems include both Wmo- and public transport. In the Randstad area, a lot of systems are still managed by the municipalities themselves, while outside the Randstad many systems are set up in collaboration with the province.

Research into DRT systems has been done both on actual and perceived performance. The research on the actual performance of DRT has focused on creating cost-effective systems, by looking at planning and employment of vehicles and chauffeurs. It is common for authorities to survey the satisfaction of users of their DRT system. What seems to be missing is an overarching insight in customer satisfaction over DRT systems and insight in the quality of the systems outside of the financial optimization. This study has been done to create such insight.

In the next chapter, the context of the study is clarified by expanding on the setup of DRT in the Netherlands and by providing an overview of literature. In chapter 3, the study setup is presented, including the study objective and questions. Chapter 4 summarizes the results from two sets of interviews that were conducted. Chapter 5 characterizes each Regiotaxi service and their region. In chapter 6, the evaluation framework is presented and the objectives and indicators are explained. The rescaling that was done to make the customer satisfaction scores comparable is elaborated in chapter 7. Chapter 8 explains the use of weights in the evaluation model and presents the results of the policy priority survey that was held among the participating transport authorities. This is then used in chapter 9 for the evaluation of the regions, in this chapter also benchmarks are provided for every region for every indicator. In chapter 10 the policy priorities and the results of the evaluation are set against each other to consider the effectiveness of policy. Finally, chapter 11 contains the conclusion and recommendations.

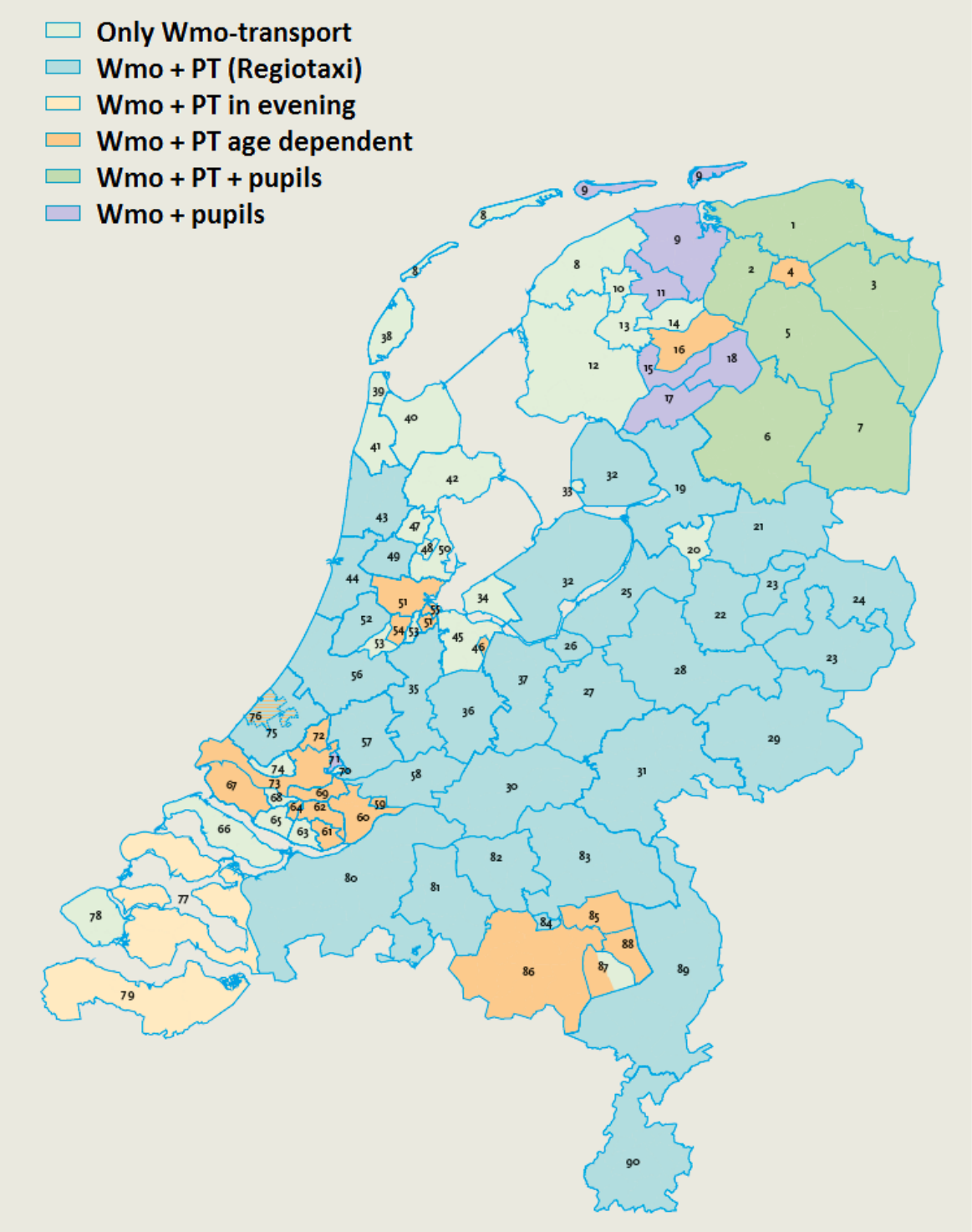

Figure 1 Wmo transport systems in the Netherlands (Kennisplatform Verkeer en Vervoer, 2013)

2 STUDY CONTEXT

This chapter provides the context of the study. The study focuses on Regiotaxi; therefore a definition of DRT and an explanation of the concept of Regiotaxi are given to start this chapter off. Secondly, different classifications for DRT systems are discussed. The third paragraph focuses on DRT in the Netherlands. Regiotaxi is categorized based on the categorizations of the previous paragraph. In the study an evaluation of Regiotaxi is performed, so in the fourth paragraph literature on the evaluation of DRT is presented. Finally, the issues with DRT are discussed using the PESTLE framework.

2.1 DEMAND RESPONSIVE TRANSPORT

For several decades now, DRT has been used as a custom made solution to transport in many areas. It can be employed for different reasons, such as: providing affordable transport in thinly inhabited areas, functioning as a feeder transport mode for other public transport systems or providing special transport for people with specific needs for their transport like impaired or elderly people. At its simplest DRT can be defined as *any form of transport where day to day service provision is influenced by the demands of users* (Nelson & Phonphitakchai, 2012). A more applied definition is given by Grosso et al. (2002):

"Demand Responsive Transport (DRT) is an intermediate form of transport, somewhere between bus and taxi and covers a wide range of transport services ranging from less formal community transport through to area-wide service networks"

There are many terms for DRT; it is also referred to in literature as: demand responsive transit or service, flexible transport service (FTS), dial-a-ride (DAR or DART), Special Transport Services (STS), paratransit and several more. The design and focus of these systems varies slightly but they all fall under the former definition.

Originally, the application of the concept grew when countries started to develop legislation for providing transport to impaired and disabled people in the 1970s and 1980s (Nelson, Wright, Masson, Ambrosino, & Naniopoulos, 2010). Conventional fixed route public transport at that time was often not designed for use by the disabled and DRT systems were used to provide transport. More recently countries are expanding their view on who is a potential user, beyond impaired and disabled people. The application of DRT is becoming increasingly common for all types of users.

Up until two decades ago, most DRT systems in Europe were created and maintained locally, evolving from a local need for transportation, without strong involvement of higher levels of government. Since then, an interest has been rising to coordinate the organization of DRT from a higher level. The system is sometimes included in social policies, as a means for helping socially excluded people in areas with a low population density and low accessibility. The growing interest of governments has coincided with the ability to order and plan trips using telematics, making DRT more viable in a lot of situations. Since the end of the 1990s, DRT became more technically viable because of advances in software and digital maps for in-vehicle computers, remote communications and the use of GPS (Enoch, Ison, Laws, & Zhang, 2006). In Europe, the SAMPO (ANIMATE, 1997) and SAMPLUS (ANIMATE, 1999) projects researched the impact of these new technologies on DRT and showed clear benefits.

In a DRT service, flexible routes are planned based on pre-ordered trips. These trips are combined into a route to make the DRT service more cost-efficient than regular taxi services. A consequence of this method is that, possibly, a detour will be made when taking customers to their desired location. Therefore the trip will take longer and there will be more uncertainty about the total trip duration, the pickup time and the arrival time. This makes the service less suitable for people with a limited time budget. Nelson & Phonphitakchai (2012) found that for the DRT system in the region of Tyne and Wear in the UK nearly 80% of the users had a trip

purpose of shopping, entertainment, leisure, and visiting friends and relatives. This shows that the service is a better match for leisure trip motives associated with larger time budgets.

Generally, DRT services are set up in low density rural areas with little access to the fixed-route public transport system. However, there have been systems that worked in an urban context (Nelson & Phonphitakchai, 2012). DRT services in rural or urban areas can differ in which performance indicators are of importance for the service. In the US, the transit cooperative research program (TCRP) made separate guidebooks for measuring and assessing performance of rural and urban DRT services (Transit Cooperative Research Program, 2008),(Transit Cooperative Research Program, 2009). Also, in the UK, a distinction has been made between rural and urban bus operations (Enoch, Ison, Laws, & Zhang, 2006). Both urban and rural DRT are under performance pressure. The rural system has issues with managing costs of providing transport to a low density, low demand, stretched out region. Urban services can have issues with managing high demands efficiently and with offering high service to a relatively bigger share of disabled and impaired customers.

2.2 CATEGORIZATION OF DRT AND REGIOTAXI

In order to evaluate the performance of DRT, the differences between services need to be appreciated. Services differ in the environment in which they are set up, the goals they have and the markets they operate in.

For the INTERMODE consortium, Enoch, Potter, Parkhurst and Smith (2004) propose four composite typifications of DRT: Interchange DRT, Network DRT, Destination-Specific DRT and substitute DRT.

1. The interchange DRT is connected to the public transport system with integrated timetabling and tickets with connecting services and guaranteed connections. It feeds the general public transport system. The service is meant to be close to taxi services, with high comfort, because choice users are a target group. The fares need to compete with perceived motoring costs.
2. Network DRT has similar characteristics as interchange DRT, but it can also serve to replace existing public transport. Network DRT is not necessarily interconnected with existing public transport and can be a service of its own.
3. Destination-specific DRT is a specialist form of DRT that serves a particular location, like an airport or employment location.
4. A substitute DRT-service is when DRT is the main public transport service and therefore does not feed, or complement the existing system, but rather is a ‘reinvention’ of the public transport system at a location. Substitute DRT-service can also be for specialist forms of transport, such as school buses.

Another report that tries to place DRT is written by the Derek Halden Consultancy (2006). They identify four main markets for DRT:

1. Premium value services: these services are high-end, focusing on reduced travel times, customer care and comfort. An example is an airport transfer service.
2. High value to agency services: these are specialist services, tailored for the needs of their particular target group; for example school transport or employment transport.
3. High care needs: this category is quite diverse. It includes services for disabled, non emergency patient transport, social services transport and community transport.
4. Best value public transport: the markets for this type of DRT are low density areas and generally anywhere where the DRT is cheaper or better suited than conventional fixed-route public transport.

The difference between these typifications is that Enoch et al. (2004) focus on the position of DRT in the complete public transport network, comparing what the DRT service does to the conventional bus services. The report from Derek Halden Consultancy (2006) considers the differences between DRT services as a more

isolated entity, where the user markets are considered, but secondary markets from substitute transport services are not as important.

Another categorization that is made by Enoch et al. (2006) is a sliding scale of financial viability, going from commercially viable to acceptable subsidy, justifiable higher subsidy and financially unsustainable. Most DRT systems will fall in the second or third category, because DRT often serves niche markets that cannot exist in a fully commercial environment. DRT systems generally have a too low income and complex user requirements. In the third category, a reason for subsidy to be justifiably higher is because a system is designed for specific user groups, like people who are impaired or experiencing social exclusion. Some commercially viable systems do exist; especially in the USA this is more common. Financially unsustainable systems logically disappear automatically over time.

2.3 DRT IN THE NETHERLANDS

In the Netherlands, involvement of the national government in DRT systems started in the mid-1990s, as a means to reduce the amount of unprofitable rural bus lines. The province of Gelderland, and other transport authorities after them, used this development to combine publically accessible DRT systems and care transport, using means from the 'Wet voorzieningen gehandicapten' (Wvg, law on facilities for disabled) to support the system. During the 2000s the amount of services kept growing. The base for the current Regiotaxi system is 'Wet personenvervoer 2000' (Wp2000, the law on passenger transport); this law states that public transport authorities are responsible for the public transport element in Regiotaxi systems. The Ministry of Infrastructure and the Environment is responsible for providing a framework of legislation and resources that the PT-authorities can use to set up the system (Kennisplatform Verkeer en Vervoer, 2009). The objective of Regiotaxi is to supplement fixed route public transport services and/or replace discontinued public transport lines, especially outside peak hours and in areas with a low population density.

Because the service suffered from low conspicuousness and not being well known, one brand and logo were introduced by the Ministry of Infrastructure and the Environment for DRT throughout the country under the name 'Regiotaxi' in 2006. Along with creating one brand, a product formula was agreed upon by 'het Nationaal mobiliteitsberaad' (National Mobility Deliberation), with indications and bandwidths for the availability, pre-registration time, call-back service, departure time, detour time, service offered by the chauffeur and arrival time insurance.

Regiotaxi is a form of DRT which is open to the public, but it is often combined with social mobility programs, which provide mobility options for groups like school kids, impaired people and people working in sheltered workshops. Regiotaxi has a share of 32% of the total turnover of approximately 1 billion euro for all DRT systems in the Netherlands. Other main DRT systems are: school transport (24% of total turnover), AWBZ- (a more comprehensive version of the Wmo, which includes care and treatment for people with a long term illness) and Wsw-transportation ('wet sociale werkvoorziening', work transport, 27%), ambulance services (11%) and the supra-regional transportation of handicapped; Valys (6%) (SEO economisch onderzoek, 2011).

Table 1 Operational categorization of DRT (Enoch, Potter, Parkhurst & Smith, 2004; Round & Cervero, 1996)

Characteristic	Alternatives
Scheduling type	Fixed schedule, demand responsive, unscheduled
Route type	Fixed route, route deviation, flexible route
Client type	Specialized, general public
Number of trip segments	Transfer, no transfer
Ride-Sharing	Shared, exclusive
Vehicle type	Minicab, taxi, minibus, midibus
Origin-Destination relationship	One-to-one, one-to-many, many-to-one, many-to-many
Origin-Destination service	Door-to-door, Stop based
Real-time information access	Accessible, not accessible
Service goals	Efficiency, equity

The operational categorization of DRT can be based on different characteristics; these are shown in Table 1. Regiotaxi systems are generally demand responsive, with a flexible route. The vehicles that are used can vary, depending on the needs of the customer, since the client type can both be from the general public or clients with specialized needs, like a wheelchair lift or extra leg room. The origin-destination relation is many-to-many and door-to-door: this means there are no transfers. It should be noted though, that some transport authorities set a limit on the use of Regiotaxi for the general public, in the sense that it can only be used for trips which cannot be completed by regular public transport. This means that for some trip requests, the client will be asked to partially make the trip by public transport and hence there will be a transfer. That the service is generally door-to-door is a consequence of the provision of transport to often mobility-impaired Wmo-passengers. The category of real-time information access is somewhat outdated, as it stems from the paper from 1996 and alludes to the availability of route-information. This is nowadays widely available in the Netherlands, so the Regiotaxi services also have access to this. As for passengers they have no access to real-time trip information. The service goal is equity of accessibility for everyone. Of course all DRT-systems seek to be efficient, but Regiotaxi is not a profit-oriented service and seeks to provide mobility for both the general public and passengers with special needs.

Regiotaxi is an overarching term, and transport authorities have some freedom to include or exclude elements for their service. This means that Regiotaxi can fall in different classifications of paragraph 2.2, depending on the local system design. The classification used by INTERMODE (Enoch, Potter, Parkhurst, & Smith, 2004) is not very well suited for Regiotaxi, as Regiotaxi can fill any of the roles of the classification.

In some Regiotaxi regions, travelers are required to use conventional public transport if a good connection is available. In that case Regiotaxi only serves as a feeder for the public transport service. This is similar to interchange DRT. There are also areas with Regiotaxi, where small unprofitable bus lines have been canceled and Regiotaxi has taken on a part of that transportation, as in Network DRT. Regiotaxi is always a many-to-many type of system, however, some services have special Regiotaxi pick up spots, specifically at hospitals, retirement homes or public transport stations. These examples correspond to characteristics of destination-specific DRT systems. Finally, some Regiotaxi services arrange transportation for pupils to school, which falls in the definition of a substitute DRT service.

For the categorization of the Derek Halden Consultancy (2006), all classifications can once again apply except for the premium value services. High value to agency fits in the cases where school transport is provided. ‘High care needs’ is a classification that applies to all Regiotaxi services, as it is part of the Wmo. Regiotaxi is set up as a social support service for either people with low mobility or people with poor access. It is therefore not meant to be high-end transportation, or at least only in so far to fulfill the basic needs of special user groups. This is also why in the classification of financial viability from Enoch et al. (2006) Regiotaxi would not fall in the

commercially viable category. In that classification Regiotaxi falls into category 2 or 3. The public transport element of Regiotaxi generally is in the acceptable subsidy range; some subsidy is provided to give people in rural areas means for transportation. The Wmo-transportation element of Regiotaxi falls in the justifiable higher subsidy category, as it provides mobility for a vulnerable social group.

2.4 LITERATURE ON THE EVALUATION OF DRT SERVICES

When evaluating a service, the question it boils down to is: What is the quality of the service? Quality for service can be defined in different ways, for example Crosby (1979) defines it as "conformity to requirements". This is a rather technical definition, looking at how a service performs compared to what is required of the service. However, the definition of quality can also take a consumers point of view. In this case the perception of performance defines quality. Grönross (1990) specifies quality as: "The outcome of an evaluation process, where the consumer compares his expectations with the service he perceives he has received." Finally, quality can also be used as a synonym for excellence. In that case quality does not exist until it is perceived. In the other definitions, quality can be high or low, but in the last definition quality refers to what is considered high quality in the other definitions. In that case, low quality does not exist, but instead is the absence of quality (Paquette, Cordeau, & Laporte, 2009). The last definition is not of interest for this study, but the other two definitions of quality, either based on actual performance or on perception, will be considered next.

2.4.1 PERFORMANCE

In this paragraph the method of evaluating performance in literature is of interest, in order to inspire the setup of this study.

Enoch et al. (2006) evaluated DRT in Wiltshire, called Wigglybus. This system runs in four different areas and the different areas have different policy objectives and markets for which they are set up. Three of the systems use flexible routes between set origins and destinations for the bus. The fourth system offers door-to-door transport from and to the local hospital. In the evaluation the systems were put through a side by side comparison on how they were set up. Categories that were taken into consideration were: scheme context, scheme design and scheme performance. Scheme performance was evaluated using patronage, costs, revenues, subsidy per month, cost cover, average fare and several different costs per passenger. The performance evaluation therefore was very much focused on the financial side. These financial indicators were also compared to the performance of conventional bus lines in the same area. The recommendations which came out of the evaluation were mostly aimed at making the service more cost-effective, for example by simplifying routes and fare structures or using different kind of vehicles

Crainic et al. (2008) point out that Demand-Adaptive transit Systems (DAS, a type of DRT where optional stops and detours are incorporated in a system that is further similar to a conventional bus line) are built up from a set of complex design choices and therefore an evaluation framework should be tailored to a specific system. Their evaluation is focused on the effective planning of the vehicles. They offer an evaluation framework for the effectiveness of the planning of DAS lines and compare the planning of DAS to the planning of conventional bus lines and Dial-a-Ride systems, a fully flexible DRT service.

Brake et al. (2007) looked at flexible transport services (FTS) in Europe and tried to distil success factors from different systems and use them as recommendations for other systems. They define FTS as a broader service than DRT, DRT being a subset of FTS. However, they do not provide a further explanation of the difference between the two types and continue to use the terms interchangeably. They identify success factors in the subareas of economic viability, technology, service and system design, management, and marketing and promotion. Their recommendations contain some less remarkable advice like *'there is a clear need to identify the most appropriate type of scheduling system'* (technologies) and *'(the service design) should reflect the outcome of a comprehensive user requirements exercise'* (service design). They recognize the potential in

expanding services and creating partnerships, but warn for several problems, like the dispatch center not being equipped for the combined services. Finally, they advise to divert a lot of attention to maintaining visibility, because for a more flexible service, the way booking, routing works and the type of vehicles that belong to the service become less clear to potential clients.

Diana, Quadrifoglio & Pronello (2007) compare the emissions of fixed route transit systems with many-to-many type demand responsive services for different:

- types of city lay outs on a surface of 25 km^2
- service quality levels
- demand densities

They show that DRT systems can potentially do much better concerning emissions in situations with low demand, where high service quality is sought.

There has been a series of European projects, which looked at different dimensions of DRT. CONNECT (co-ordination of concepts for new collective transport) was a collaboration on centralizing information on DRT in Europe. A knowledge portal was developed for FTS where current research on current practices for FTS was gathered and best-practice was developed for practitioners. The project ended in 2006. In 2008 and 2009 this work was continued in the FLIPPER project (flexible transport services and ICT platform for Eco-Mobility in urban and rural European areas).

The SAMPLUS project followed the SAMPO project in the mid and late 90s. The SAMPO project attempted to develop DRT systems supported by ITS in four European countries: Belgium, Finland, Italy and Sweden. The project successfully attempted to identify hidden demand and plan routes for this demand in Travel Dispatch Centers (TDC) (ANIMATE, 1997). In the SAMPLUS project the overall aim was to demonstrate and evaluate the usefulness of telematics technologies for DRT. This was done in the same countries as the SAMPO project. They concluded that the use of telematics makes DRT systems much more viable and user friendly (ANIMATE, 1999).

The MASCARA project (deMand reponsive trAnsport service for increasing Social Cohesion in urbAn and Rural Areas) ran from 2005 to 2007 and aimed to assess the current viability of DRT and identify a best practice for making DRT work across Europe. To do this, the potential of integrating DRT in regional transport strategies was investigated. In the MASCARA project, Andrade (2008) worked on an evaluation framework for DRT systems. Within the MASCARA project there was interest in the performance of the criteria: social inclusion, sustainable mobility and technical performance. The underlying objectives and indicators can be picked per case, depending on the aim of the evaluation. A multi-level evaluation framework was developed using two existing multi-criteria analysis methods: Analytic Hierarchy Process (AHP) and the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS). With this evaluation framework, DRT systems can be evaluated per separate case.

2.4.2 PERCEPTION

Just as important as the technical performance of DRT systems is the perception of this performance by the user. Although it is generally assumed that these are strongly correlated, they can be far apart. The most convenient way of determining customer satisfaction is through surveys. Several researchers have attempted to list the dimensions and underlying attributes, which can be used to survey satisfaction for DRT services. Falcocchio (1979) suggested three dimensions: convenience, comfort and safety and connected eleven attributes to these dimensions, which are listed in Table 2. The dimensions are still a little scant. Since then, it has been suggested that dimensions to use in satisfaction surveys could be taken and adapted from the satisfaction survey of conventional public transport.

Table 2 Dimensions used by Falcocchio

Dimensions	Attributes
Convenience	1. Reliability 2. Waiting time 3. Transfers 4. Ease of entry and exit 5. Walking distance
Comfort	6. Heating and ventilation 7. Noise 8. Sudden stops and turns 9. Having a seat
Safety	10. Fear of falling 11. Fear of muggings

Pagano and McKnight (1983) developed a list for dial-a-ride services based on attributes originally established for public transportation. Besides that, they also used existing literature for dial-a-ride services, like the work of Falcocchio and own observations. They developed 8 dimensions with 41 corresponding attributes. The list of Pagano and McKnight (1983) is quite extensive and can be found in appendix A. They additionally tried to discover which dimensions and attributes are most important. To do this, they surveyed different user groups, based on age and mobility. They discovered that importance of the different items is not the same for different user groups. For example, the group under 65 valued reliability greatly while the group over 65 was more concerned with safety. They conclude that measuring quality for a dial-a-ride service is complex and the difference between user groups should be taken into account when doing so. The most important attributes they found are highlighted in the table in appendix A. Knutsson (1999) also tried to make a division of dimensions for quality. His list can also be found in appendix A. His work is partly inspired by the work of Pagano and McKnight; yet he makes a very different grouping. His dimensions are: information, dignity, comfort, travel time and fare. Compared to Pagano and McKnight the attributes for information and fare are totally different. Dignity partly corresponds to driver characteristics and travel time corresponds partly to reliability. Comfort surprisingly does not correspond to the attributes used by Pagano and McKnight at all; instead they focus more on vehicle access, extent of service and responsiveness.

2.5 ISSUES WITH DRT

There are several issues for DRT services to flourish. The issues are touched on in this section using the PESTLE framework (Political, Economic, Social, Technical, Legal, and Environmental factors). The issues are taken from Enoch, Ison, Laws & Zhang (2006), Brake, Nelson & Wright (2004) and Ferreira, Charles & Tether (2007).

Politically, there are few issues for DRT. Government agencies are generally well aware of the existence and possibilities DRT offers. It's important to keep interest from politicians though, so services can continue to develop and do not die out. Also, there's a need for community ownership of the DRT systems, so an ongoing consultation with user groups and the community can help the quality of the service greatly.

Economic viability is one of the main issues for DRT. Income is generally low. The system suffers from not being widely known, due to it serving niche markets. These niche markets usually cannot exist in a competitive tendering process, because the market is too small or because the specific demands of the users are too complicated and diverse. Furthermore, ticket prices are often kept artificially low for special user groups. Costs are high, because a large part of the costs of running a vehicle are personnel costs. DRT services generally run smaller vehicles with relatively a lot of personnel. Vehicles require special investments for specific user groups, like an investment in a wheelchair lift.

As for social issues, in some cases, public DRT services have suffered because of a bad image. The service is seen as only for impaired or elderly people, which means a portion of potential ridership is lost. Another cultural barrier can be the aversion of traveling in a shared vehicle, due to the proximity to strangers. Especially in smaller vehicles this may be an issue. Also, DRT systems are often poorly recognized, especially many-to-many setups, because of the lack of physical infrastructure like central stops. Ridership can also be

disappointing due to a lack of 'public transport culture'. In areas with high car ownership, where taking public transport is not commonly done, this habit will also be reflected in the ridership of DRT services.

Technological issues are not very prevalent in DRT anymore. In the past years, the implementation of telematics for booking and planning trips has greatly aided the development of DRT systems. Also, in-car systems have been developed to help communication with chauffeurs and using effective routes during the trip. Identifying the most appropriate scheduling system is the main technological challenge.

The impact of legal barriers can vary greatly from country to country. The law does not always clearly define the position of DRT systems. As the services are somewhere between buses and taxis, it is not always clear what rules and arrangements apply to them. This leads to them not being able to profit from arrangements for either group. Sometimes it is possible to make use of arrangements, but an operator has to go through complex bureaucracies to be able to benefit from them. Another barrier can be the integration of conventional public transport and DRT services. The integration of DRT in the public transport network can be very effective, however, there needs to be clear regulation in place to separate the roles of bus, taxi and DRT service.

Environmental issues do not really apply to DRT services. A demand for the use of environmental friendly vehicles could be included in the tender, if authorities assign high priority to this. Potentially DRT services could perform worse than conventional public transport, when applied to high demand areas. When this occurs, however, the threshold of economic viability for a conventional bus line will already have been crossed and the service will be replaced, especially since there's still a preference for fulfilling demand with conventional bus lines by transport authorities.

In the next chapter the study setup will be presented, including the core problem that is being studied, the main objective of the study and the research questions.

3 STUDY SETUP

3.1 PROBLEM DEFINITION

The literature overview shows that some work has been done in evaluating DRT systems. The focus in these studies has mostly been on the economical side, the effective planning of trips and vehicles (as a subset of the economical evaluation) and internally on perception through customer satisfaction surveys. There are some other possible angles that so far have been neglected. Three issues will receive a closer look:

First of all, the focus of customer satisfaction surveys has been on internal control only. All transportation services for the Wmo are by law obligated to perform such a survey at least once a year. However, there has not been a cross-regional comparison of these surveys. As a consequence of each transport authority performing their own customer satisfaction survey, there's a lack of uniformity in the design of the surveys. This makes it difficult to compare the satisfaction of users between regions and to evaluate the perception of users of DRT services in the Netherlands as a whole.

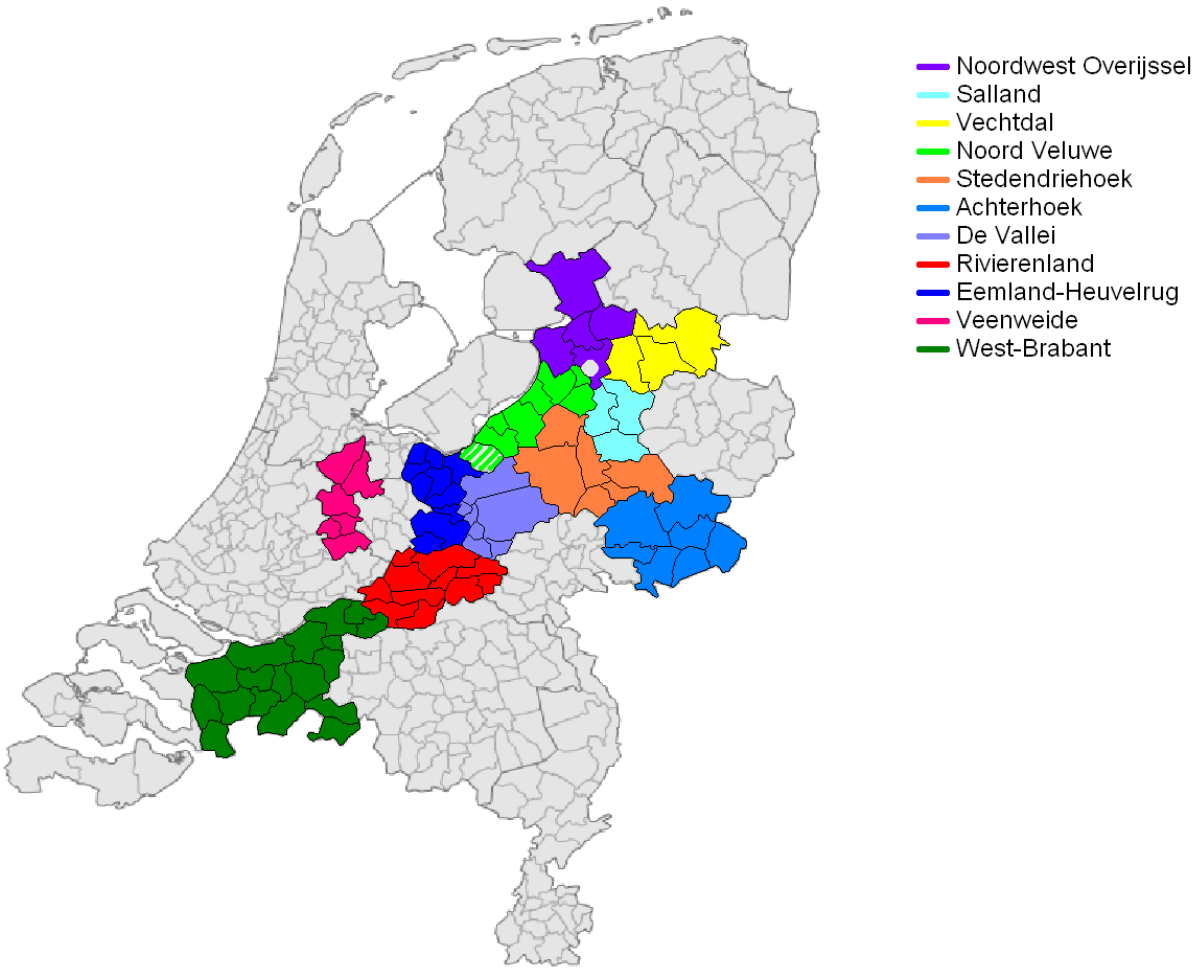

Figure 2 Participating Regiotaxi services

Secondly, a different angle would be to evaluate the internal performance of DRT. Internal performance here is defined as the performance of the system itself, as opposed to external performance, which would show the effect on other systems. For example, internal performance could be measures like tariff, customer satisfaction or travel time. Examples of external performance are effects on congestion or emissions. External performance measures the effect on other systems (traffic, the environment), while internal performance relates to the DRT

system itself. So far, the evaluation of internal performance in literature is mostly focused on the economical side, i.e. the amount of subsidy per passenger or kilometer and operational costs. As for Regiotaxi, other performance measures like punctuality and travel time are monitored. However, once again there has not been a cross-regional comparison of these performance measures.

Finally, the evaluation scores give insight into the performance in different areas for the Regiotaxi services. It will show the areas that have potential for improvement. This improvement will generally be shaped by policy, through the requirements from the tender and the management of the Regiotaxi service after the contract has been procured. However, this only works if policy is having its intended effects. A third issue that is interesting to consider is therefore: to what extent are policy priorities reflected in the evaluation results?

3.2 STUDY OBJECTIVE & RESEARCH QUESTIONS

Given the gaps in research mentioned in the previous paragraphs, transport authorities were approached to get cooperation for an overarching evaluation of several Regiotaxi services. The provinces of Utrecht, Gelderland, Overijssel and the region West-Brabant were willing to provide data to conduct such a study. In total, these four transport authorities manage eleven different Regiotaxi services. Figure 2 shows a map containing the municipalities in the Netherlands with the regions participating in the study colored. With enough data available, the main objective of the study is:

To assess the performance of Regiotaxi in the Netherlands, by (a) benchmarking performance indicators and finding internal relations between them, (b) evaluating different Regiotaxi systems and (c) determining the effectiveness of policy, given the policy objectives of the transport authorities.

The study thus looks to accomplish a deeper understanding of how Regiotaxi is performing and how its performance is being influenced. In order to do this, the following questions are posed:

(a) What are benchmarks for relevant performance indicators and how are these indicators related?

The study is looking to get an indication on what can be expected of Regiotaxi services by looking at what bandwidth they are currently operating in. For this, customer satisfaction surveys, monitoring data and management data of the regions are used. By getting a better insight in what the performance on separate performance indicators is and how they are related, there is a basis to consider the next question.

(b) How are different Regiotaxi systems performing overall and how do they perform relatively to each other?

For this research question, an evaluation framework is designed based on the framework of Andrade (2008), who developed an evaluation framework in the European MASCARA project. After performing this evaluation, the results can be put against the policy objectives of the different transportation authorities to find out the following:

(c) To what extent are policy objectives of the transport authorities reflected in the performance of Regiotaxi?

For this research question, the transport authorities are given a short survey to discover their priorities for the different objectives in the evaluation framework. The relative score on these objectives is then compared to the relative importance the transport authorities give to the performance objectives.

4 INTERVIEWS

At the start of the study, an understanding about the context of Regiotaxi was developed. Besides using literature, two sets of interviews were conducted among different stakeholders of Regiotaxi. The first set of interviews was used to further focus on issues in the Regiotaxi and to develop context for the study. Also, input was asked for the evaluation model presented in chapter 6. The second set of interviews was done with the participating regions, in order to get better insight into their specific regions and the data they had available.

4.1 INTERVIEW SUBJECTS

In order to obtain the maximum amount of information, the interviews in the first set were conducted in an open, conversational way. A topic list was prepared beforehand, with specific questions that were of interest. However, it was attempted to let the interview flow and to pick up new topics by connecting them to the answers given. Perspectives from four people in the field were sought: an expert in the field, a contractor, a traveler's representative and a transport authority.

The expert interview was conducted with Guy Hermans, who works as the program manager collective transportation of the Transport Knowledge Resource Centre (KpVV). KpVV supports all professionals in the various decentral tiers of government involved in the development and implementation of traffic and transport policy. KpVV develops and disseminates knowledge and expertise in this specific field. From this interview, other potential interview targets were established, from transport representation and public transport user representation.

Dirk Schenk, the director of Vloettax, was interviewed for the contractor's perspective. Vloettax carries out the trips for Regiotaxi Flevoland. For the traveler's perspective, an interview was conducted with Gerrit Douma, the secretary of Rocov Overijssel. Rocov Overijssel is a representative body for travelers in public transport, advising the provincial government and public transport contractors, and monitoring the public transport in the region. Finally, to separate the first and second set of interviews, a governmental body, which did not contribute data, was sought. Regio Twente was a good candidate; they were approached to help with data in first instance, but denied this request. With their rejection they did offer to help in other ways and hence were included in the first set of interviews. Marco Berloth from Regio Twente was interviewed. Regio Twente is a partnership between the municipalities in the region Twente; they manage Regiotaxi Twente.

The second set of interviews was conducted with:

- Joost Pullens, Regio West-Brabant
- Klaas Veenma (general contact), Björn Edelenbos (Policy), Jaap de Kleine (Data), province of Overijssel
- Johan Wyma, province of Utrecht
- Gijs Pelsma, province of Gelderland

The purpose of this second set of interviews was threefold. First of all, the interviews were used to take inventory of the monitoring data that was available. The evaluation framework in the next chapter depended on the information that was readily available, because there was no option to gather additional data during the study. The second purpose was to learn more about how the Regiotaxi systems were organized. Finally, the interviews gave the transport authorities the opportunity to learn more about the study and directly ask questions and give input on what would be interesting to focus on.

4.2 EXPERT INTERVIEWS

The topics discussed in the expert interviews varied greatly and gave good qualitative information. In this paragraph, information that is directly relevant within the delineation of the study subject is presented. From

the topics that were discussed that fall outside of the final delineation, the most interesting remarks are also discussed.

4.2.1 DEVELOPMENTS IN REGIOTAXI

To start the interviews off, recent developments in Regiotaxi were discussed. The environment in which Regiotaxi operates is changing quickly, because of reforms and budget cuts by the National Government. According to Guy Hermans this mainly affects the care-side of Regiotaxi, as there are changes in Wmo and Awbz.

According to Gerrit Douma, in order to streamline the organization of Regiotaxi, the provinces were becoming more and more involved in the tender and the management. Coupled to the changes in Wmo and Awbz, a reverse trend is now observable, where provinces are reconsidering whether they want to have a large responsibility for a system that mostly should be the responsibility of the municipalities. Wmo travelers are the responsibility of municipalities and for many systems make up 70%-90% of the total amount of Regiotaxi users. Provinces are usually only responsible for the remaining 10%-30% of travelers in the Regiotaxi: the public transport travelers.

In Twente, Marco Berloth mentioned that there is a decreased use of Regiotaxi, while the target user group is growing. There is no explanation so far why this is happening.

The interviewed were also asked whether they thought Regiotaxi had successfully managed to take over unviable conventional bus lines. The general opinion was negative on this topic. Marco Berloth felt that the system was not ‘collective’ enough to replace buses, i.e. in Twente only a combinationfactor of trips of 1.2 is achieved, making the system incomparable to conventional buses. Guy Hermans pointed out that the high ratio of Wmo users to public transport users suggests that public transport travelers have not found their way to Regiotaxi. Dirk Schenk mentioned that when unviable bus lines were removed in his region, there was a surge in interest for Regiotaxi. Since the accessibility in his region is low, people needed to find some alternative. Compared to other Regiotaxi services, his area serves a relatively high amount of public transport travelers (~50%). However, it is hard to say whether there are choice-users who made the switch from bus to Regiotaxi.

4.2.2 FINANCING, ORGANIZATION & MANAGEMENT

Regiotaxi has its basis in the Wmo; it’s a component that all Regiotaxi systems have. Several of the people interviewed mentioned the misuse of Awbz travelers of Regiotaxi. The Awbz has its own transportation system and this system is financed from different sources than Regiotaxi. Misuse can occur when Awbz-travelers are referred to Regiotaxi by Awbz-institutions, whereas people with an Awbz-indication are supposed to use Awbz transport. Regiotaxi and Awbz transport are designed to be two separate systems. When assigning the travelers to Regiotaxi instead of Awbz transport this puts an extra strain on the finances of Regiotaxi.

Marco Berloth suggested it could potentially be beneficial for DRT in the Netherlands to look for more opportunities to combine services. The misuse by Awbz travelers of the Regiotaxi shows that different DRT systems are separated based on where the money is coming from to finance the systems. However, Awbz transport and Wmo transport are not so different from each other and an efficiency gain might be obtained by combining the systems. Gerrit Douma mentioned that the combination of the Wmo and public transport function of the Regiotaxi is currently not causing any issues for the travelers.

On the question of how the current role of Regiotaxi in the public transport system could be improved, different ideas popped up. First of all, because of budget cuts, many governmental bodies are looking to shrink the current Regiotaxi system to cut costs. Guy Hermans pointed out that this trend could potentially reap fewer gains than is intended. For example, Wmo travelers, including for example people in wheelchairs, switching to conventional bus lines can put pressure on time-tables because more time is needed to enter and exit the bus.

So it should be considered that the advantages for some actors do not necessarily outweight the disadvantages of other actors. Also, Dirk Schenk pointed out that shrinking the service means that fewer trips can be combined and that the costs per trip rise, voiding part of the cuts made. Marco Berloth pointed out the importance of the public transport part of Regiotaxi for people who do not have a Wmo indication, but do feel they are limited in their mobility. This user group may not feel comfortable using regular public transport and has a need for Regiotaxi to participate in a social environment. The public transport part of Regiotaxi is a great alternative for them.

There's disagreement on whether price differentiation would be a good tool to create a more effective system. Guy Hermans and Marco Berloth both saw potential in using this to improve the system. The latter gave a few examples like asking an additional fee for special needs like a wheelchairlift or extra leg space. Another option would be to create a discount for people who can be flexible in the time at which they travel. For example someone requiring Regiotaxi to do groceries at 11.00 could get a discount if they indicate they can be picked up between 10.00 and 13.00 instead of having a specific pickup time. This would allow transport operators to better combine trips and lower the costs. Dirk Schenk did not think price differentiation has potential, on account of it being too complex for the travelers.

4.2.3 TENDER

The length of contracts between the transport operator and transport authority is in practice in the Netherlands about 3 to 5 years. In the interviews it is recognized that for the contract length, a balance must be found between on the one hand contracts short enough to enable market effects and to adapt for policy changes, and on the other hand contracts long enough to get a return on investment for transport operators and consistency in the system for travelers. Generally it is thought that the current balance is good for contract lengths. Only Dirk Schenk would like to see longer contracts made possible when there's already a satisfactory relation between the transport operator and authority.

There are some doubts about the current effect of the tendering process. Guy Hermans mentioned that the design of the tender currently does not stimulate transport operators to increase quality beyond the minimum set in the tender. He feels this is mainly because it is the main condition transport authorities look at. Transport operators feel like they have no chance to win a tender based on quality only, even if that quality justifies a potentially higher price. There are also doubts about the effectiveness of bonus-malus arrangements. According to Marco Berloth, there is no proof that the bonus-malus arrangement is working as an incentive to increase quality beyond the minimum. However, he explained, the arrangement can be used by the transport authority to build some goodwill with the transport operator, by setting attainable targets and hence giving the occasional bonus, so that a pleasant relation is maintained. Dirk Schenk indicated that 'doing something extra' is not valued enough. According to him, it is best for a transport operator to stick to the specifications of the contract. He also pointed out that when customer satisfaction surveys are used for a bonus-malus arrangement, it is important that the survey is robust and complete and that all parties agree that the survey is fair towards the transport operator.

Gerrit Douma noticed a trend that transport authorities are making the demands in the tender increasingly specific. This leads to a loss of the advantages from the tendering process, since it is supposed to stimulate transport operators to come up with smart, cost reducing solutions. Making the tender very specific leaves little room for the transport operator to do this. Marco Berloth acknowledged the importance of giving the transport operator enough freedom to supply the transport. He pointed out that the tender needs to be very specific on the topic of communication, i.e. all parties must know who gathers which information and communicates this to which party in a specific way. This includes both communication between organizations and the harmonization of computer systems.

In general the interviewees agreed that transport authorities do a good job in determining their priorities in the design of the system. However, sometimes this is not translated into good criteria in the tender. Another problem can be the lack of evaluation. Guy Hermans and Dirk Schenk both reported situations where a demand for a maximum age of the vehicles was not fulfilled by the transport operator, and no action was undertaken by the transport authority.

4.2.4 OBJECTIVES

Three important objectives were given in the interviews. Gerrit Douma said that it was important to maintain Regiotaxi as core transport system to give full access to everyone to public transport. As for internal objective, Guy Hermans stated that all time-related factors are important, i.e. punctuality, detour time. Finally, several interviewees stressed the importance of informing the traveler well on the rules and possibilities of the system. Gerrit Douma refers to a study, which shows improvement of the perception of travelers of the system after an information campaign.

In the Netherlands there exists a guideline for public transport, which advices transport authorities on which data should be gathered for the monitoring of the system: 'Model informatieprofiel openbaar vervoer' (MIPOV) (TransTec adviseurs BV, 2008). In the interviews it was asked whether it would be useful to create such a guideline for DRT systems, or Regiotaxi specifically. Guy Hermans indicated that this could be useful, it make the comparison between regions possible by standardizing between the regions. However, Marco Berloth found that for their region they do not have any problems setting up the monitoring and he did not see the added value in such a guideline. Also, if such a guideline was to be made, it would be important that there is no obligation to follow this guideline and the transport authorities continue to have the freedom to arrange it themselves, in accordance with their own views on what is important for their system.

4.3 INTERVIEW WITH THE TRANSPORT AUTHORITIES

The interviews with the transport authorities were part of the process of setting up the evaluation. The main topics were the organization of their service and the data that they collected. The results of those interviews will not be discussed here, but they are used in chapters 5 and 6.

5 DESCRIPTION OF THE REGIONS

In this chapter the Regiotaxi systems and their respective regions are introduced. First some general system characteristics of the Regiotaxi systems, then characteristics of the region that could be relevant for the use and the performance of Regiotaxi are presented and finally the system setup and rules are discussed.

5.1 GENERAL SYSTEM CHARACTERISTICS

The general system characteristics in this section are taken from the trip databases. The trip database for each region contains the trips made in the period of September to November 2012. This period was chosen at the start of the study, because it is the period gives a sufficient, but manageable amount of data. It was the most recent trip data available, except for December. The choice was made to avoid December, because in the interviews it was pointed out that December is different from other months concerning usage, because of the holiday season.

5.1.1 TRIP VOLUME

The trip volume varies strongly per region. The largest amount of trips for the considered period are from Regiotaxi West-Brabant with just over 260.000 trips, while the smallest amount of trips is from Regiotaxi Veenweide with just short of 18.000 trips. West-Brabant is ahead of the pack by a long way, as the next biggest region is the Achterhoek with 107.000 trips. Most regions produce between 20.000 and 100.000 trips over the 3 months period. There is a significant negative correlation between trips per inhabitant and the average trip distance (-0.65). It seems that if relatively more trips are made, the additional trips are short trips. In Salland by far the most trips per inhabitant are made and they have the shortest average trip distance, whereas Veenweide has the fewest trips per inhabitant and the longest average trip distance.

Table 3 Characteristics for the Regiotaxi systems

	Total trips*	Trips per 1000 inhabitants	Average distance [km]	Distance standard deviation [km]	Avg. (free flow) Travel time [minutes]
Noordwest	19706	594	10.3	9.3	13.7
Salland	59730	2,287	8.4	7.4	11.7
Vechtdal	37452	981	10.9	8.1	13.6
Noord Veluwe	57529	1,324	9.6	7.8	12.5
Stedendriehoek	20762	264	9.7	8.0	12.7
Achterhoek	106909	1,619	9.5	6.9	12.7
De Vallei	92896	1,250	8.7	7.4	12.7
Rivierenland	88094	1,383	10.3	8.2	12.6
Eemland-Heuvelrug	96015	957	10.3	8.5	13.5
Veenweide	17911	0.369	14.0	9.4	16.9
West-Brabant	261578	1.516	9.1	7.9	12.3

*For the period of September-November 2012

5.1.2 USER PROFILE (AGE, GENDER, WMO)

The user profile is not complete for all the regions. For Noordwest, Salland and Veenweide there are no traveler data available. For the other regions, the average user seems to be 60-65 years old, with about 2/3rds of the travelers being female. Eemland-Heuvelrug sticks out with both the oldest average travelers at 67 years and the smallest portion of male travelers at less than 30%.

For some of the data in table 4, only user characteristics were available if a travel pass was used. This means that these data points are potentially skewed, because people who have travel passes are likely regular users. The profile of the typical regular user is not necessarily the same as of the incidental user.

Table 4 Characteristics for the Regiotaxi users

	% of trips for WMO	Age	Std. dev. Age	% Male
Noordwest	60.18%			
Salland	57.30%			
Vechtdal	63.03%	*63.3	*21.0	*39.2
Noord Veluwe	84.83%	58.3	18.8	37.6
Stedendriehoek	92.82%	60.1	16.8	33.0
Achterhoek	81.34%	60.1	17.7	35.0
De Vallei	82.00%	57.9	18.7	36.8
Rivierenland	73.42%	58.3	19.3	38.0
Eemland-Heuvelrug	90.81%	**67.9	**19	**29.1
Veenweide	92.02%			
West-Brabant	85.88%	64.7	19.9	31.0

*Only users with travel pass (~30% of trips)

**Only users with travel pass, based on birthyear instead of birthday

5.1.3 GENERATING TRIP DISTANCES

Trip distances were generated for all regions, based on origin and destination information from the trip data. In four of the databases, the trip distance was available in kilometers. In the others distance was not given or only reflected in zones traveled. The databases with distances in kilometer available did not all use the same program for determining the travel distance. For comparing data, it was considered very relevant to include the average distance of trips. After all, it is easier to keep within the boundaries of a maximum fifteen-minute extra travel time if relatively a lot of short trips are made. A matlab script was used to generate the travel distances in Google maps based on the postal codes of the origin and destination. This allows generating insightful variables, like subsidy per passenger kilometer, as well as comparing performance indicators to travel distance. The slightly adapted script was taken from a fellow student, Tim van der Kruijs, who had used the script to generate distances that visitors of leisure facilities travel to those facilities, for about 1000 trips. The script also suited the need in this study to generate the distances of the approximately 858.000 trips in this research.

Initially only the regions without trip distances in kilometers had their distances generated, with the region Noordwest Overijssel as extra region to check how different the results of different trip planning programs were. The results of this test were that the differences can be big. Table 5 shows that the distances in maps were generally longer than for the routing programs used in the trip databases. Note that the small number of equal distance is caused by the data from the trip table having 2 decimals, while the google maps distance are rounded to 1 decimal. 13083 of the compared trip distances were within 1 kilometer, 14890 were within 2 kilometers. This means that nearly 20% of the generated trips have a larger discrepancy than 2 kilometers.

Also, the mean distances differed strongly, being 9.2 kilometers for the default program and 10.5 kilometer for Google maps.

Because of this test, it was decided that in order to get consistent data, all trips had to be generated in the same manner. The trip distances present in the trip databases were discarded. The other regions had more congruent distances compared to the maps distances than Noordwest though (Table 5)

Table 5 Distance generation comparison

Region	Noordwest (trips)	West-Brabant (trips)
Google maps shorter distance	3987	110339
Equal distance between the programs	529	3653
Google maps longer distance	14016	136562
Total trips	***18532***	***250554***

There was a complication while running the script, as it did not return answers for 7,6% of the trips. Manually going through the first two hundred of these error values, it appeared that part of the errors occurred because trips below 1 km could not be processed. This happened in 78 of the 200 manually reviewed values. Some postal codes from the file either did not exist or at least did not exist in Google maps: this led to 24 of the 200 missing values. Finally, the last part seemed to have something to do with the formatting of the postal codes, as it seemed to not return an answer for specific postal codes. All trips to or from Barneveld in the manually reviewed trips for example, did not return a value (98 out of 200 values).

The consequences of the missing values are limited; the biggest problem is the missing values below 1 kilometer. This was solved by creating dummy values based on the sample that was taken from the errors. The average distance of trips less than one kilometer was 700 meters. The missing values therefore were assigned a value between 400 and 1000 meters in a uniformly distributed manner. The non-existent postal codes were simply void and no compensating actions were performed to adjust for this. Finally, the sample of possibly bad formatted trips, showed similar characteristics to the 92.4% of good data, therefore no adjustments were made for these missing values either.

5.2 REGIONAL CHARACTERISTICS

The regional data were taken from the CBS. This section is important for providing context for the other data, for interpreting the performance and satisfaction scores. Where possible, the data were taken from the end of 2012, however, some data are not generated periodically and can be from earlier years. Table 6 and Table 7 show the characteristics for the regions.

5.2.1 POPULATION AND SURFACE AREA

The population in the Regiotaxi service areas range from 100k to 690k, with most regions having less than 300k inhabitants, as can be seen in Table 6. Eemland-Heuvelrug and West-Brabant stand out as more populous areas. At the same time, West-Brabant also stands out for having a large surface area, while Eemland-Heuvelrug is actually rather small for the population it has.

5.2.2 URBANISATION

Urbanization is the ratio of inhabitants who live in urban areas. This definition can be ambiguous, as the definition of what constitutes an urban area can strongly vary. The data that are used in this study use the definition of ‘Begrenzing bebouwd gebied 2000’ for built-up area, which is used in the ‘Nota ruimte’. Appendix B contains a figure which explains the determination of what constitutes a built-up area. Urbanization is

Table 6 Regional population characteristics

	Population 2012	Population 2011	Population 2008	Population 2008 BBG	% of Population in BBG	% of Population in Wmo
Noordwest	132744	131881	130570	99671	76.3%	2.74%
Salland	104462	103925	102361	67508	66.0%	2.69%
Vechtdal	152763	152927	151706	121994	80.4%	2.40%
Noord Veluwe	173866	172781	170041	133054	78.2%	2.53%
Stedendriehoek	315147	314358	312745	257919	82.5%	3.25%
Achterhoek	264086	265077	265764	204087	76.8%	3.03%
De Vallei	297375	293250	289454	241188	83.3%	2.45%
Rivierenland	254876	253354	251851	194471	77.2%	2.08%
Eemland-Heuvelrug	401302	397784	391151	347663	88.9%	2.30%
Veenweide	194163	193238	182724	151774	83.1%	2.04%
West-Brabant	690277	685888	678885	590896	87.0%	2.55%

*BBG = Built-up area

defined as the percentage of inhabitants living in built-up area. Urbanization of all regions is between 76% and 89%, with the exception of Vechtdal, which distinguishes itself from the group with only 66% urbanization.

5.2.3 DISTANCE TO FACILITIES

Another regional characteristic that was taken into account is the average distances in a region to facilities often frequented using Regiotaxi. The facilities that were taken into consideration are: general practitioners office, the hospital and railway stations. In general, the average distance to the GP or hospital does not vary as strongly as the average distance to a railway station. Most notable are the relative large distance in Vechtdal to the GP office and to the hospital in Noord Veluwe, and the short distance in Salland to railway stations.

5.2.4 VEHICLE OWNERSHIP

Vehicle ownership is similar for most regions. Only West-Brabant stands out a little by having slightly more motor vehicles per person.

Table 7 Regional accessibility characteristics and vehicle ownership

	Average distance GP office (km)	Average distance hospital (km)	Average distance railway station (km)	Motor vehicles per person
Noordwest	1.472	5.482	6.414	0.448
Salland	1.178	4.478	2.431	0.466
Vechtdal	1.793	6.644	6.018	0.508
Noord Veluwe	1.158	7.794	3.767	0.468
Stedendriehoek	0.993	5.436	3.797	0.483
Achterhoek	1.361	6.231	4.586	0.498
De Vallei	1.006	3.937	3.349	0.449
Rivierenland	1.216	6.321	5.444	0.497
Eemland-Heuvelrug	0.813	4.124	3.828	0.493
Veenweide	0.976	5.936	5.595	0.474
West-Brabant	0.949	5.503	6.227	0.584

5.3 SYSTEM SETUP

The trip data, obtained from the trip databases, which were made available by the participating regions, contain the following information about the trip:

- characteristics of the traveler (e.g. age, Wmo-indication, gender),
- characteristics of the reservation (e.g. reserved departure/arrival time, special requirements),
- characteristics of the trip (actual departure/arrival time, trip distance, address of departure/arrival)
- payment (trip price customer, subsidy).

The type of data that is gathered differs from service to service. In several cases data were collected in different units. The most obvious example of this is that Gelderland and Utrecht worked with a zonal system while the other services used distance in kilometers to determine price.

5.3.1 DEPARTURE AND ARRIVAL TIMES

These items were available for all services, except for West-Brabant, which didn't have a specific time, only a date and hour-stamp.

5.3.2 TRAVEL TIME

Travel time was available for all regions, although for some it was not readily available and had to be derived from the departure and arrival times.

Except for Regiotaxi Veenweide, all systems have an agreed maximum detour time they can use for combining trips. The transporter will try to remain within these boundaries. The detour rules for the different regions are shown in Table 8.

Table 8 Detour rules and conditions per Regiotaxi service

Regiotaxi region	Condition	Detour rule
Noordwest	> 2 zones	50% of the shortest travel time
Salland	> 2 zones	50% of the shortest travel time
Vechtdal	<10km or >10km	10minutes – 20minutes
Noord Veluwe	=< 2 zones or > 2 zones	15minutes – 30minutes
Stedendriehoek	=< 2 zones or > 2 zones	15minutes – 30minutes
Achterhoek	=< 2 zones or > 2 zones	15minutes – 30minutes
De Vallei	=< 2 zones or > 2 zones	15minutes – 30minutes
Rivierenland	=< 2 zones or > 2 zones	15minutes – 30minutes
Eemland-Heuvelrug	=< 2 zones or > 2 zones	+50% detour – +100% detour
Veenweide		No detour rule
West-Brabant		30 minutes

6 EVALUATION FRAMEWORK

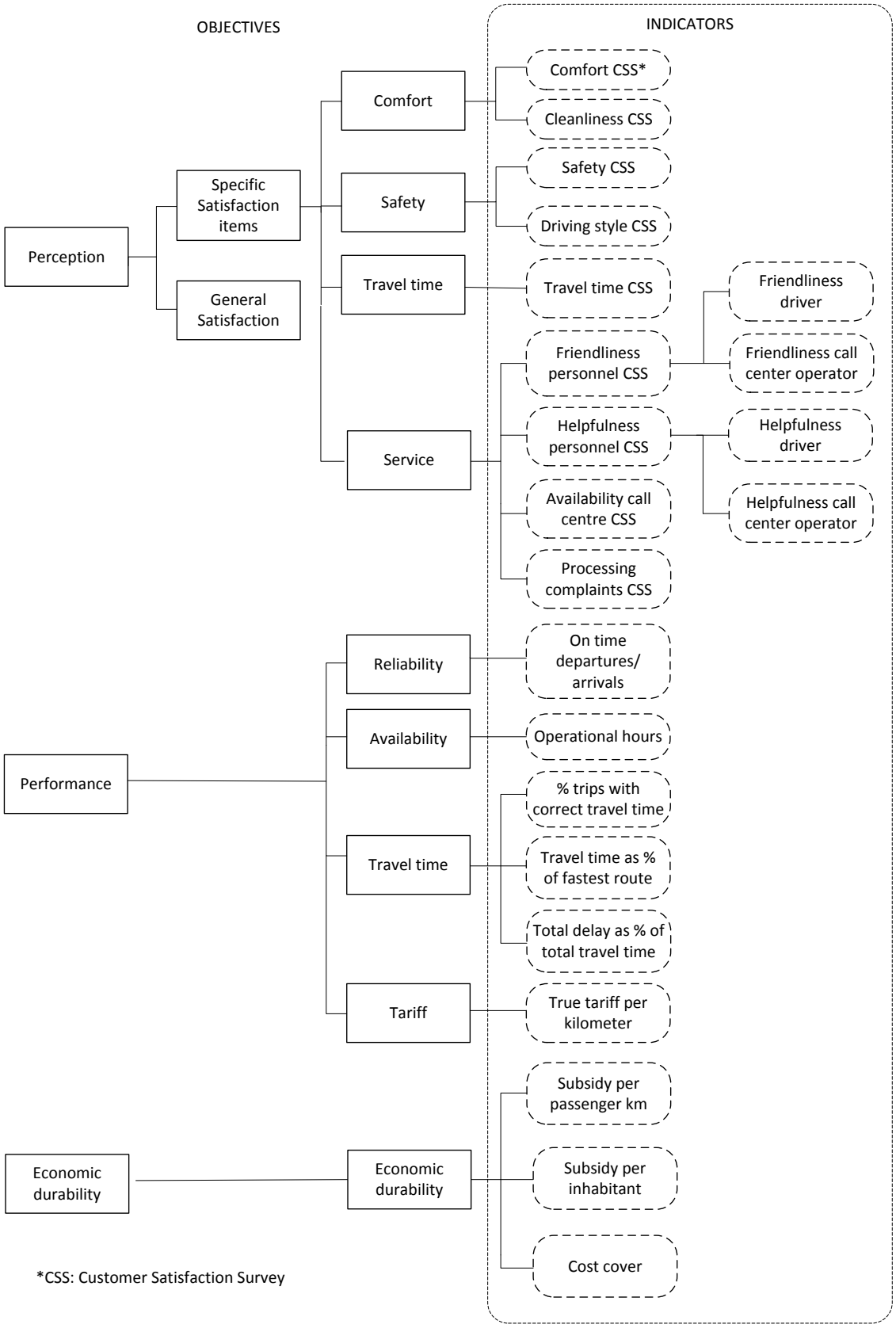

Figure 3 Evaluation model

The regions are evaluated using a multi-criteria analysis method. This chapter presents the objective tree and the indicators that describe the objectives for evaluating the Regiotaxi systems with parts of the multi-criteria analysis algorithms of AHP (Analytical hierarchy process) and TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution). The method was taken and slightly adapted from the evaluation framework of Andrade (2008) and is used to generate a single score per system without ranking the different systems. This means that the evaluation framework can be used to rate single systems and the priorities per individual region can be included.

6.1 OBJECTIVES

6.1.1 OBJECTIVE TREE

A list of policy objectives for performance of Regiotaxi was generated and indicators were matched to these objectives to measure them. The list of goals was presented to four people in the field in the first set of interviews described in chapter 4. They had no additions or suggestion for changes except for one. It was suggested to use information provision as objective for perception or performance. This suggestion was taken to heart, but in the end was not included in the model, this is explained in paragraph 6.1.2. The objective hierarchy is shown in Figure 3.

At the top of the hierarchy, three main objective-groups are defined: perception, performance and economic durability. These objectives cover the most important internal objectives for the Regiotaxi. Economic durability is important because the system needs to have a cost structure, which allows it to be sustained over a longer period. Except this is worth little if transport is slow or unreliable; therefore a certain performance level is also required. Also, in order to keep attracting travelers, the perception of the users is important.

Different indicators for economic durability can be used for example: cost cover and subsidy related indicators. Not every region was prepared to release their management reports for the Regiotaxi, however, therefore options were limited. An important reason for including economic durability to the evaluation framework was to balance out performance indicators. Especially the price per kilometer for the traveler can be directly influenced by the amount of subsidy available for the Regiotaxi service.

The objective of performance can be split up in several parts: reliability, availability, travel time and tariff. The first three sub-objectives are taken from monitoring data. Usually, as part of the contract, the transportation company provides this information to the public transport authority for monitoring purposes.

Perception is measured in the customer satisfaction surveys. The use of these surveys, however, brings up a predicament. There are two ways to come to determine the overall satisfaction of travelers: with a single or multi-item measure. The single item measure is the answer to the question in the survey: “Rate how satisfied you are with the Regiotaxi system.” The multi-item measure looks at the overall satisfaction as the result of satisfaction of sub-objectives like travel time or service. These individual items do not necessarily predict the overall satisfaction well. This asymmetry is found in several studies. A good overview is given by Slevitch & Oh (2010). So the single-item overall satisfaction and the combined multi-item overall satisfaction indicators are not the same, but which one is the correct measure? Bergkvist and Rossiter (2007) give a nice overview of the advantages and disadvantages of using either multiple- or single-item measures for a construct (like satisfaction). The advantages of a multiple-item measure are: higher reliability, better discrimination in categories of the attribute, and it is useful if the construct is abstract. The arguments against using multiple-items are: it is unnecessary if the object is concrete singular and additional items might tap into another predictive attribute. Bergkvist and Rossiter studied the validity of both methods and concluded that both methods have equally high predictive validity. For this study, satisfaction can be obtained both from multiple-items and an overall satisfaction score. They are both used in this study and given equal weight.

The specific objectives in the 'perceived-tree' cover several objectives, the items are similar to those used for example in the 'OV Klantenbarometer' (Kennisplatform Verkeer en Vervoer, 2013) and 'Benchmark OV Vlaanderen 2006' (TransTec adviseurs BV, 2006).

The 'OV Klantenbarometer' uses:

- general satisfaction
- information and safety
- driving comfort
- time and traffic flow
- tariff

'Benchmark OV Vlaanderen' uses:

- facilities
- information and safety
- tariff
- driving comfort
- time and traffic flow

Of these, facilities is not relevant to the Regiotaxi, as all systems under consideration provide door to door transport and the only facilities of the Regiotaxi are designated locations with a 'taxi-post', which can be used to order a ride.

6.1.2 EXCLUDED OBJECTIVES

The model presented in Figure 3 is tailored for this research. First of all, that means that only internal objectives, which are objectives that directly relate to the system, are considered. The data for external objectives (for example: reducing congestion or improving social inclusion) are not available and thus were not included. Secondly, the objectives in the lower part of the hierarchy could potentially be expanded. The presented model is limited by the data that were available and because of that not complete. There are several other objectives that could be of interest. Examples of additional objectives for perception can be taken from the organization of satisfaction by Pagano & McKnight (1983), who use items like 'vehicle access' and Knutsson (1999), who uses 'information'. Some of the items they use can be re-organized to the objectives used in the model of Figure 3, while others, like 'information', are objectives that had to be left out because there were too little data available.

To illustrate this lack of data, 'information' will be used as an example. The level of information provision or satisfaction about information provision in particular would be a very interesting objective to take into consideration, because the level on which travelers are informed, would strongly indicate their happiness. However, the data available on this topic are too scattered across the regions to include it. Overijssel does not have an item regarding 'information' in its survey. Utrecht and West-Brabant both ask whether the traveler is familiar with the call-back services. On top of that Utrecht asks the opinion about the information provision when ordering a ride and West-Brabant also asks about familiarity with the option to order a priority trip. While knowledge of these specific items (call-back service and priority trips) could be an indicator of how well travelers are informed, it is a meager measure and does not suffice. It is for example possible that the public transport authorities actively promoted these options, while the information provision of the general system is lacking. Gelderland asks the opinions on the completeness and accuracy of the information provision as a whole. These are usable indicators, however, they are not sufficient to include 'information' in the objectives hierarchy, as more than half the systems would require dummy-variables for this objective. So overall, not enough uniform data are available to include 'information' in the evaluation.

The evaluation focuses on the performance of the Regiotaxi service at trip level, so this means: the satisfaction about trips, the performance of trips and the costs/subsidy per trip. Overhead costs are not included in the evaluation, thus excluding costs for maintaining the project bureau which monitors the service and puts out the tenders.

The next three paragraphs elaborate on the indicators that were chosen to describe the objectives in the lower hierarchy.

6.2 PERCEPTION

The data for the indicators for perception all came from the customer satisfaction surveys (CSS). There is no uniformity of surveys for the different regions and similar questions were gathered to use as indicators. The precise selection and conversion process of the satisfaction indicators is presented in the next chapter. An overview of which questions were used from the customer satisfaction surveys to measure the indicators can be found in appendix C.

6.2.1 COMFORT

The objective of comfort is build up from comfort and cleanliness. Satisfaction about cleanliness was polled in every customer satisfaction survey, satisfaction about comfort was missing in Overijssel and Gelderland.

6.2.2 SAFETY

Safety is described by the direct item from the CSS for safety, but also by driving style. Driving style was a difficult indicator to match to an objective, as an argument can be made for placing it under either safety or comfort. As a result of discussion in the interviews it was elected to assign driving style to safety. Remarkable is that both Overijssel and West-Brabant did not poll the attitude of travelers towards safety.

6.2.3 TRAVEL TIME

Travel time satisfaction is taken from a single-item question for all surveys, except West-Brabant.

6.2.4 SERVICE

Service is the most complex sub-objective for perception. Service is defined as the extent to which the transport operator cares for his passengers. This is split up in the helpfulness and friendliness of the personnel, the availability of the call centre and the satisfaction about how complaints are processed. The personnel are split up in drivers and call centre operators. The service indicators are generally well measured. Only Overijssel is missing several items: Helpfulness of personnel and processing complaints. Gelderland and West-Brabant do not have an question for polling the friendliness of the call center operator and Utrecht does not poll the helpfulness of the call center operator. It should be noted that the satisfaction of processing complaints only applies to travelers who filed a complaint and therefore the response for the CSS is significantly lower for that item.

6.3 PERFORMANCE

6.3.1 PUNCTUALITY

There are two points to measure punctuality: departure and arrival. All the regions use the condition that a trip is punctual if the pick-up and drop off are within 15 minutes before or after the pre-arranged time. For most trips, punctuality is only dependant on departure and there's no pre-arranged arrival time, because this can vary depending on the combination of trips. The latest allowed arrival time results from the detour rules in the region. Some systems offer a guaranteed time of arrival, which can be used for traveling to an important or time-sensitive event, like a funeral or to a railway station in order to be in time for a connection to a train. In that case the transporter can set the departure time. Only a very small portion of trips uses this however, for

example: of all the trips in Gelderland, only 0.03% of the trips used a guaranteed arrival time. The condition is taken into account where applicable.

6.3.2 OPERATIONAL HOURS

Table 9 Operational hours

Region	Operational hours	Hours/day
Gelderland	6.00 – 1.00 mon-sun	19
Overijssel	6.00 – 1.30 mon-sun	19.5
Utrecht	6.00 – 24.00 mon-sun	18
West-Brabant	6.00 – 1.00 mon-sun	19

The regions have similar opening hours. These are listed in Table 9. The operational hours of a standard week are used, as there are some exceptions to these times, for example on New Year's Eve most systems have deviating operational hours.

6.3.3 TRAVEL TIME

The extent to which the transporter succeeds at staying within the detour rules from Table 8, would be an obvious first choice for travel time performance, in part because it confers with an expectation the traveler has of the trip. However, the variety in rules means that using the percentage of correct travel times is useless without context. Therefore another approach would be to use the mean ratio of actual travel time to free flow travel time or the ratio of total actual travel time to total free flow travel time. These approaches have the disadvantage that they can be seen as favouring regions with respectively more short or long trips. As a detour on a short trip will sooner cause high ratio, because a similar absolute delay has a higher effect on ratio of a short trip than a long trip; this approach favours a region with a lot of long trips. However because per trip, a long trip would have a bigger influence on the second approach than a short trip, it could potentially also cause imbalance.

The three indicators together give a good insight into travel time performance, as the approaches using ratios can serve as context for the percentage of trips within the boundaries. That is, the percentage of trips represents the extent to which the expectations that are created by the detour rules are met. The approaches using ratios give a measure as to how quick travelers are actually transported, and include this interest of the traveler.

6.3.4 TRIP PRICE

Tariff is a difficult indicator to compare the systems. All of the regions use a call out fee, but beyond that the tariff becomes difficult to compare. Gelderland and Utrecht use a zone system, while Overijssel and West-Brabant use a price per kilometer. On top of that, the different regions define different user groups for special tariffs; different rates can be found for elderly, people living in the countryside or people with a location specific origin or destination like a taxipoint or a railway station. Even within one system tariffs can vary, as sometimes municipalities will subsidy certain user groups. Because of the variety of tariffs, a straightforward comparison of the different prices is not possible. There are two ways around this and both methods make use of applying the tariff to trips and then comparing those results. This method allows for even very complex tariff structures to be boiled down to a single number, although it is then no longer really a comparison of tariffs, but of trip prices paid. The first method is by simulating different trip scenarios and calculating the price of every region for these virtual trips. These prices can then be compared for the different scenarios. The second option is to use the trips that are actually made. This leads to a slightly different indicator, i.e. the price *current users* pay as opposed to price of an arbitrary trip.

The downside for the second option is that price and user profile are related and therefore the comparison does not take into account how attractive the tariff is for a random potential user. This means that if a

Regiotaxi has a very cheap tariff for a user group like elderly people, the low price will cause relatively more elderly to use the service and therefore the mean price that is paid to drop. On the one hand the better score for tariff is justified, because elderly get a lower price. On the other hand, there may be a large potential user group who do not make use of the service because of the high price they have to pay, and these non-users are no longer accounted for in the comparison. Another weakness of using the actual trip prices is that between regions, trip length matters. A region with on average long trips will have a smaller influence of call out fees and appear cheaper in the price per kilometer than a region with shorter trips.

The downside of the first method is that insight in the potential users becomes relevant. After all, with the scenarios, a Regiotaxi service could achieve a very good score by having a low tariff for a target group who barely makes use of the service, even with the special low price.

Table 10 Tariffs overall and split for the user groups Wmo-travelers and Public transport traveler

	Tariff/km [euro]	Tariff/km (WMO) [euro]	Tariff/km (PT) [euro]
Noordwest	0.3498	0.2201	0.4964
Salland	0.4471	0.2875	0.6139
Vechtdal	0.2854	0.2225	0.3767
Noord Veluwe	0.3033	0.2235	0.7152
Stedendriehoek	0.2748	0.2351	0.6685
Achterhoek	0.3442	0.2402	0.6946
De Vallei	0.3796	0.271	0.7167
Rivierenland	0.3846	0.2549	0.6472
Eemland-Heuvelrug	0.2797	0.2283	0.6538
Veenweide	0.33	0.2952	0.5946
West-Brabant	0.3273	0.2688	0.6889

All in all, the whole tariff comparison requires a lot of context. Because the trip data of every region also contain the fare that was paid, it was opted to use the fares from the trip databases for comparing tariffs. In order to make some distinction between focus group travelers and others, the distinction between WMO travelers and public transport travelers is made; this distinction is present in each region. Public transport travelers are defined as all the non-WMO travelers, so special rates for seniors and such exceptions are included in the public transport price. Note that systems with more trips outside the region will on average have higher trip prices, because the customer pays full price for these zones. The tariffs are presented in Table 10.

6.4 ECONOMIC DURABILITY

Subsidy was chosen as indicator because it represented the interest of governments to provide a maximum amount of travel at the lowest possible cost. Every system under consideration is subsidized. In order to see what governments are getting for the subsidy, a similar indicator is used as for trip prices, i.e. subsidy per kilometer and subsidy per inhabitant. These indicators are a counterweight to trip price. Trip price on itself is a poor indicator of cost of the system; a heavily subsidized system might have cheap trip prices, but objectively it doesn't make its performance better. Also, as noted for trip prices, the average trip price can go up if a lot of trips are made outside the region. By using subsidy compared to the total amount of kilometers driven, this is balanced, since an increase of trips outside the region will lead to a reduced amount of subsidy/km. The data used to deduce this indicator varied greatly between the regions. In appendix D the method of deducing

subsidy per kilometer is explained per region and from that, subsidy per inhabitant could be calculated. Because of the varying data and the sometimes complex conversion to comparable subsidy numbers, the subsidy should be used with care, as the results may not be very robust. The total subsidy is shown in table 11. Cost cover was the third indicator used to describe economic durability.

Table 11 Total subsidy

	Total subsidy [x1 mln euro]
Noordwest	1.32
Salland	1.35
Vechtdal	0.49
Noord Veluwe	1.22
Stedendriehoek	2.76
Achterhoek	3.11
De Vallei	2.65
Rivierenland	2.27
Eemland-Heuvelrug	4.11
Veenweide	1.53
West-Brabant	2.60

7 CONVERTING SURVEY SCORES

7.1 CUSTOMER SATISFACTION DATA

According to the Wmo the municipalities are obligated to hold at least a yearly customer satisfaction survey, for Wmo-facilities. All the participating Regiotaxi services are (also) a Wmo-facility and the customer satisfaction surveys are organized system-wide, not per municipality. The customer satisfaction surveys were all performed in 2012. The response for each region is shown in Table 12.

Table 12 Survey response of the customer satisfaction surveys

Bureau	Region	Response
Gelderland call centre survey	All 5 regions in Gelderland	458
Gelderland travel survey	Achterhoek	453
	Noord-Veluwe	280
	Rivierenland	444
	Stedendriehoek	468
	de Vallei	328
Overijssel	Noordwest	132
	Salland	125
	Vechtdal	120
West-Brabant		429
Utrecht	Eemland-Heuvelrug	482
	Veenweide	491

The surveys for the Regiotaxi services in Gelderland were held in September 2012. The total survey was split into three surveys: a travel survey, a call centre survey and a bureau survey, surveying respectively the satisfaction about the trips, about the call centre and the project bureau. For the evaluation, only the first two sub-surveys were used; the bureau survey was not taken into account, as it did contain relevant questions for this research. The questions that were used can be found in appendix C. Since all the services use the same call center, the survey was not held separately for each region.

The response to the survey in Overijssel is by far the lowest of the different regions. This is probably caused by the choice to use a telephone survey instead of a paper survey. This choice also limited the amount of questions that was asked in the survey. As a consequence, Overijssel has the most missing values for the evaluation. Another difference with the surveys of the other regions is that the one for Overijssel was done throughout the year and not at a specific time. Every two weeks several users were called for the survey.

For Eemland-Heuvelrug and Veenweide, two surveys were held: one in spring and one in autumn. The last survey of 2012 was used. West-Brabant only had a single survey.

7.2 CUSTOMER SATISFACTION SCORES

The customer satisfaction survey data are used in the generation of scores and for exploring the main drivers for satisfaction. There are 5 different surveys that are used: Overijssel, Gelderland, West-Brabant, Veenweide and Eemland-Heuvelrug. The two systems in Utrecht, Veenweide and Eemland-Heuvelrug, both have their own survey, although the two surveys share a lot of questions and differ only slightly. All the PT-authorities incorporate WMO-travel in their systems and because of the Wmo they are required to survey satisfaction at least yearly. Even though all systems with Wmo-travel have to survey satisfaction yearly, there is no template or guideline for what should be asked in these surveys. The only requirement is that the method of surveying is deliberated with representative organizations. This causes the surveys to differ on what is asked, how the

questions are formulated and how the answers are designed. Obviously, the lack of uniformity causes problems for the comparison in this research. Luckily, most of the PT-authorities ask roughly the same question. Table 13 illustrates the problem for the comparison of satisfaction between regions, by using the question concerning the driving style of the chauffeur.

Table 13 Example of variation in questioning

PT-authority	Question	Answer
Gelderland	How satisfied are you about: The driving skill of the chauffeur? Report which grade reflects your experience with Regiotaxi Gelderland best.	Tick a box between 1-10, 1 being lowest and to the left. An option for 'no opinion' is available.
Overijssel	Please indicate to what extent you're satisfied or dissatisfied about: The driving style of the chauffeur?	Circle a number from 1 to 5, 1 being lowest and to the left. An option for 'no opinion' is available and indicated with a '6'.
West-Brabant	Are you satisfied or dissatisfied about the driving behaviour of the chauffeur?	Tick a box from 'very satisfied' to 'very dissatisfied' (4-point scale). Very satisfied is the most left option. There's a fifth box for 'no opinion' available.
Utrecht	How do you judge the following aspect of the Regiotaxi Eemland-Heuvelrug/Veenweide: The driving behaviour of the chauffeur?	Fill in a number between 1-10 as in a report mark scale (10 being highest).

Although all regions ask more or less the same, there are slight differences between the questions which could lead to different answers.

Firstly, there are problems with semantics and definitions. The four different questions ask three slightly different things: driving skill, driving style and driving behaviour. Generally, all these questions ask more or less: How satisfied are you about how the driver has performed his driving task? But in order to do this, the survey-makers chose slightly different, but closely related terms to ask this question.

Secondly, the rating scales vary. Every rating scale uses a different system. Even the two 10-point scales are different, as one asks to tick a box corresponding to a scale-point number and the other one asks for a mark. Of the three grade scales, two have the most positive item at the right and one is reversed and has it at the left side. The grade scales give the option for 'no opinion' while the marking method does not. So of the many design dimensions available for rating scales, it seems that along almost every dimension there is a survey that deviates from the other surveys.

There is a third potential comparison problem, which does not appear in the example. In the previous example, all the scales that are used are theoretically balanced. They use the semantic bipolar pair of 'satisfied-dissatisfied'. However, one of the questions that was encountered in the surveys was a skewed scale, with an additional item at the positive end of the scale. The question: "What is your general impression of safety during a trip with the Regiotaxi?" has to be answered on the scale: outstanding – good – sufficient – insufficient – bad. Semantically, good – bad and sufficient – insufficient are polar ratings, making 'outstanding' a positive outlier. However, it can be argued that on evaluating a performance, 'sufficient' is the point where the performance goes from bad to 'good enough' and is therefore a neutral point, since being just slightly worse than sufficient would indicate a negative rating.

Recognizing these shortcomings in the data, there are no options available to solve many of the problems and the data have to be worked with as best as possible. The issue concerning definitions is hard to quantify and it

is not expected to find large differences between surveys based on slight semantic disparities. For example, it is to be expected that a negative opinion on driving style will lead to a similarly negative opinion on driving behaviour. The issue of skewed scales is also hard to deal with, but luckily it only occurs for one of the relevant questions. It is opted for to simply use different conversion methods and compare the different results for lack of a 'best option'. The issue of scale conversion will be expanded on in the next section.

7.2.1 SCALE CONVERSION

While there is literature on the influence of ranking order of scale points, amount of scale points, the availability of an 'opt out' and of using a marking method compared to a grading scale, none of these effects have large influences and it seems that for the purpose of this research the margin of error in these different methods is irrelevant. This leaves the problem of rescaling the 4, 5 and 10-point scales to a uniform scale. In other research, this rescaling is often done in a mathematical way, for example using:

$$\frac{(Rating-1)}{(number\ of\ response\ categories-1)} * 100 \qquad \text{(Preston \& Colman, 2000)}$$

or by anchoring the top and bottom value and matching the intermediate values by using the amount of scale items in the new scale at equal distances between scale-points (Dawes, 2002). These are easy, clear cut solutions, but they fall short in several ways:

1. The assumption of equal value of scale extremities for anchored top and bottom values.
2. Arbitrariness of scale extremities for anchored midpoint values of the scales.
3. The assumption of an equal distance between adjacent scale points.

The listed issues will be illustrated for the rescaling of a 5-point scale to a 10-point scale.

When survey subjects are presented with questions which they have to report their attitude on, on both a 5-point and a 10-point scale, how will the 5-point scale logically translate into the 10-point scale? That is, what values will be entered on the latter scale, given the answer on the former scale? Imagine that the subject wants to report that he is *almost completely* satisfied. On the 5-point scale, if 3 represents 'neither satisfied nor dissatisfied', 4 represents 'satisfied' and 5 represents 'very satisfied', 5 is the clear option here. On the 10-point scale, however, because there are more options, there's more room for nuance. 10 may seem as the option for really being completely satisfied, so where a 4 on the 5 point scale was too far out for this nuance, it's possible that the 5 will translate into a 9 on the 10-point scale. When different subjects fill in the question, some will choose 9, some will choose 10. This suggests that the translated value of 5, lies somewhere between 9 and 10 on the 10-point scale. Therefore anchoring the extremities of the two scales will lead to an artificially inflated standard deviation when converting the 5-point scale to the 10-point scale, and a deflated standard deviation when converting the other way.

A different approach could be to anchor the middle point. After all, on any scale going from very dissatisfied to very satisfied, the middle point of 'neither satisfied nor dissatisfied' will be equal; a zero-point. In combination with the consideration of the previous paragraph, extremities could then be picked in a way that larger scales have extremities farther away from the midpoint. Intermediate values could then be assigned based on equal distance between scale points between the extremity and the midpoint. However, there is nothing to base the setting of the extremities on, making any scale arbitrary. To further illustrate this, 3 different rescaling methods were applied to answers from the customer satisfaction survey of Overijssel. The different conversions from 5-point scale to 10-point scale are in Table 14 and the results of the different conversions are shown in Figure 4. In Table 14 the upper row shows the original 5-point scale points. The three bottom rows show the values of these scale points after the different rescaling methods.

Table 14 Different rescaling methods from a 5-point to a 10-point scale

Original value	1	2	3	4	5
Method	Rescaled values				
Anchored extremity	1	3.25	5.5	7.75	10
Anchored mid, 2.5-8.5	2.5	4	5.5	7	8.5
Anchored mid, 1.5-9.5	1.5	3.5	5.5	7.5	9.5

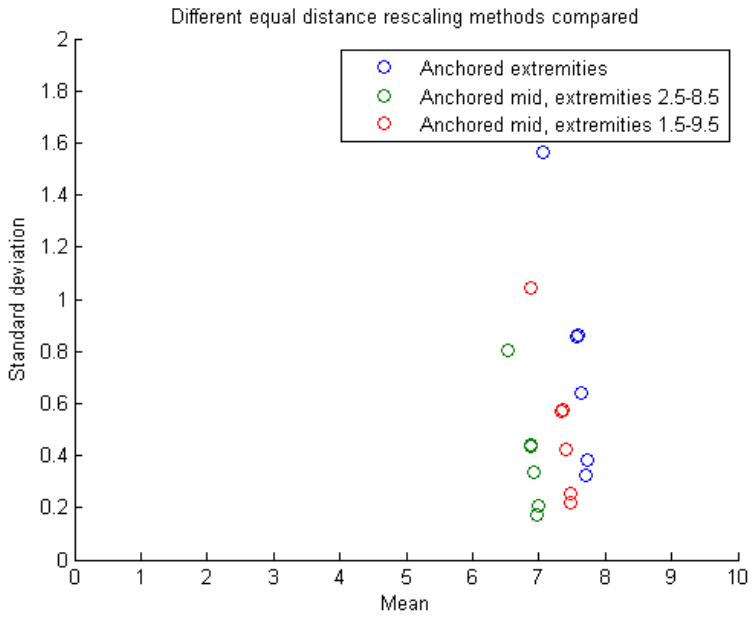

Figure 4 Different equal distance rescaling methods compared

In the figure, every point is a rescaled 5-point question. The results of 6 questions were used. The order of big to small standard deviations is the same for every rescale method, so the highest blue point is rescaled from the same question as the highest green and red point. The figure shows, that a higher gap between the midpoint and the extremity of a scale leads to a more extreme mean and a higher standard deviation.

The final issue is that of the assumption of equal distance between adjacent scale points. As Kennedy, Riquier & Sharp (1996) show, the meaning a survey creator assigns to a scale in a survey, is not necessarily the same as the interpretation that the survey subject has of the scale. Their research finds that in a scale from 1 to 5, the scale points actually correspond with values of: 1, 2.2, 3.1, 4.1 and 5. This is not necessarily problematic for comparing same-scale questions, as the interpretation of the scale for all questions remains the same. However, it inserts an extra uncertainty in the rescaling. Where with two equal distance scales, the rescaling of the intermediate points is done proportionately; if the interpretation of different scales is skewed differently from the equal distance scales, this conversion cannot be done proportionately.

To solve the previously stated problems, it was attempted to make use of the large quantity of data that were available. In order to rescale, a profile of each scale was made, using the propensity of people to use a certain scale point. These profiles can then be matched to create rescaling values. These profiles are built up from all the answers of all the questions on the same scale; the resulting frequency-graph is the profile. Figure 5 shows the graphs for a 5-point and a 10-point scale.

The main downside to this method is that it is biased by the fact that it polls the attitude towards Regiotaxi. The expectation, when applying this to more varying and truly independent questions, would be a semblance of a uniform distribution. However, since all questions relate to Regiotaxi, a normal distribution is to be expected. The use of this method can still be justified because it is also applied to attitudes of Regiotaxi and the two will have the same bias. However, the conversion factors cannot be used to convert other attitude surveys from one scale to the other.

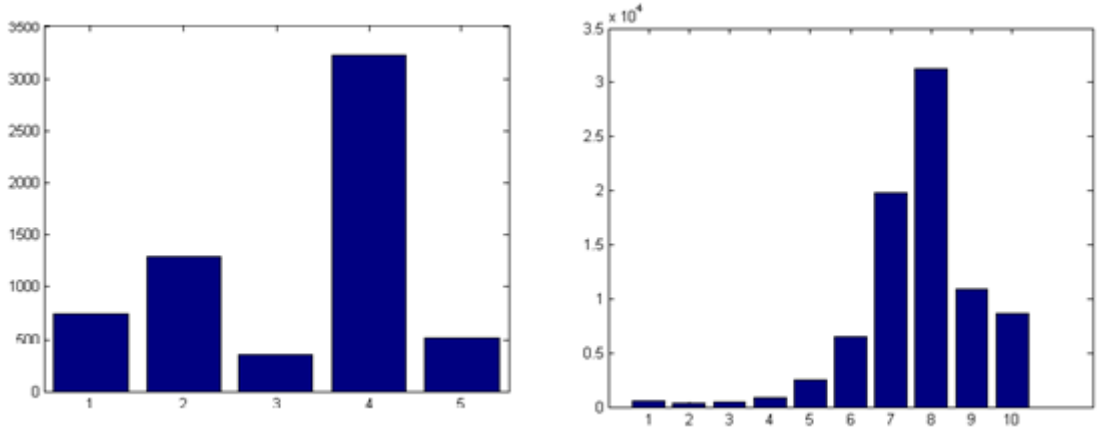

Figure 5 Frequency of scale-point use of 5-point (left) and 10-point scales

The figure for the use of the 10-point scale shows a negatively skewed normal distribution, which was expected since firstly, Regiotaxi could be considered as a luxury good when set against other public transport options. Secondly, for Wmo-users it provides a certain independence of other people and increased accessibility. Thirdly, in general the users of Regiotaxi are elderly and of this group it is known they report a more positive attitude in surveys than younger people. Therefore users will generally have a positive attitude toward the system and a negatively skewed distribution is a logical resulting distribution. In Figure 5 however, the distribution of the 5-point scale does not resemble a normal distribution. Given the propensity to use the scale point '4', it would be expected that 3 would be used more than 2.

Because of this unexpected result, the setup was reconsidered. The distribution of answers depends on the variety of questions which are used. However, in the previous set up, not the variety of questions, but the variety of answers was used. This means that a question with more responses was weighted higher than a question with few responses. Therefore, new profiles were generated, only this time every question was weighted the same, so that questions with larger responses would not dominate other questions. Figure 6 shows the new distributions.

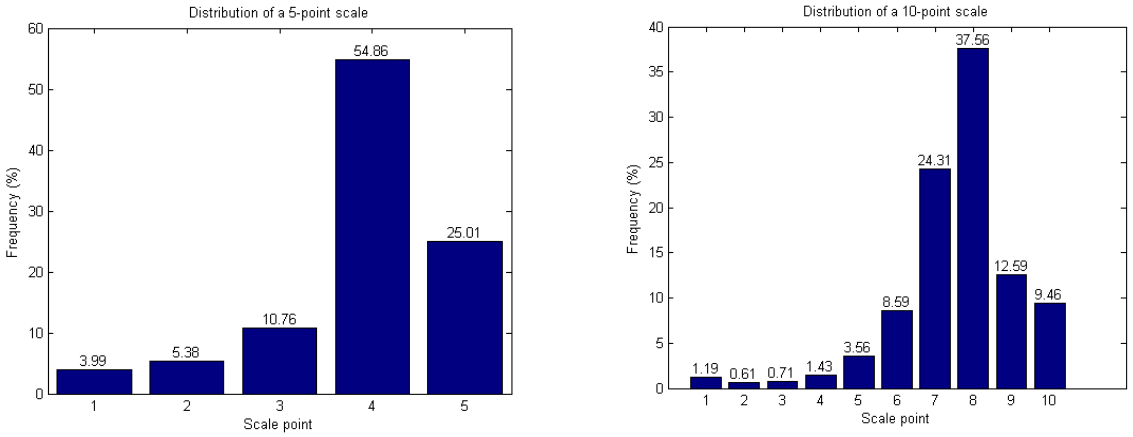

Figure 6 Frequency of scale-point use of 5-point (left) and 10-point scales revised

The distribution of the 10-point scale is still roughly the same, but the distribution of the 5-point scale is now more in accordance with expectations.

For converting the 5-point scale to the 10-point scale, a weighted rescale is applied. This means that the answers of '1' on the 5-point scale will be 2.66 on a 10-point scale, because:

$$\frac{1.19*1+0.61*2+0.71*3+1.43*4+0.05*5}{3.99}=2.66$$

The same is done for the other 5 scale points and the rescale then looks as follows:

5-point scale	1	2	3	4	5
10-point scale	2.7	5.3	6.4	7.6	9.3

The results for the regions that had to have answers rescaled are presented in Table 15 (original values) and Table 16 (rescaled values).

Table 15 Original values from the CSS

	NW*	SL	VD	EH	V	WB
Comfort	7.79	7.79	7.79	7.90	7.51	1.96
Cleanliness	4.02	4.02	4.03	7.99	7.91	1.88
Driving style	4.11	4.11	4.12	7.85	7.76	1.96
Safety				1.98	2.16	
Friendliness call center	4.06	4.08	4.09	8.03	8.06	
Friendliness chauffeur	4.12	4.14	4.14	8.07	8.09	1.84
Helpfulness call center						1.99
Helpfulness chauffeur				7.80	7.84	1.95
Availability call center	3.79	3.81	3.85	7.75	7.81	2.07
Processing complaints				2.65	2.92	2.62
Travel time satisfaction	4.07	4.06	4.07	7.13	6.89	2.43
Overall Satisfaction	8.09	8.06	8.04	7.92	8.11	7.78

*NW = Noordwest, SL = Salland, VD = Vechtdal, EH = Eemland-Heuvelrug, V = Veenweide, WB = West-Brabant

Table 16 Rescaled values for the CSS

	NW*	SL	VD	EH	V	WB
Comfort	7.79	7.79	7.79	7.90	7.51	7.37
Cleanliness	7.58	7.58	7.58	7.99	7.91	7.60
Driving style	7.62	7.62	7.62	7.85	7.76	7.36
Safety				7.75	7.49	
Friendliness call center	7.52	7.53	7.54	8.03	8.06	
Friendliness chauffeur	7.50	7.50	7.50	8.07	8.09	7.71
Helpfulness call center						7.41
Helpfulness chauffeur				7.80	7.84	7.60
Availability call center	7.21	7.22	7.27	7.75	7.81	7.01
Processing complaints				6.46	6.31	7.14
Travel time satisfaction	7.57	7.57	7.57	7.13	6.89	6.57
Overall Satisfaction	8.09	8.06	8.04	7.92	8.11	7.78

*NW = Noordwest, SL = Salland, VD = Vechtdal, EH = Eemland-Heuvelrug, V = Veenweide, WB = West-Brabant

8 SETTING WEIGHTS USING A POLICY PRIORITIES SURVEY

One of the objectives of this research is to say something about the current performance of Regiotaxi. In order to do this, a systematic method is needed to determine whether the system is performing well or not. There is more than one way to look at 'performance', however. The most obvious way is to use a list of parameters to gauge the performance. The better the system scores on these parameters, the better its performance is. However, this approach ignores the objectives and the perception of the system. For example: If the users want a comfortable ride, but are less concerned about timeliness, a system that is scored using the 'objective parameters' might score poorly, but still be a success, if it manages to offer what the user wants. So it is important to include priorities in the analysis. This is done by introducing a weighting system in the analysis.

After generating a list of indicators (see the evaluation model in Figure 3), the PT-authorities were approached to determine their priorities. However, some of the objectives differ in scale and abstraction and are difficult to compare. Also, the amount of items is much higher than the 7+/-2 that people can process according to Miller's Law (Miller, 1956). Therefore, the problem of assigning priorities is broken down in levels. The introduction of the hierarchy allows for smaller sets of objectives to compare priorities.

8.1 SETTING WEIGHTS USING ANALYTICAL HIERARCHY PROCESS

In order to determine the weights for the different objectives in a hierarchical structure, the use of the scale of measurement and the pairwise comparison of an analytical hierarchy process (AHP) is appropriate. This method is based on the following set of axioms:

Axiom 1 Given any two items i and j out of the set of objectives A, the PT-authority is able to provide a pairwise comparison a_{ij} of these objectives under any criterion c from the set of criteria C on a ratio scale which is reciprocal; i.e., $a_{ji} = 1/a_{ij} \forall i, j \in A$

Axiom 2 When comparing any two objectives $i, j \in A$, the PT-authority never judges one to be infinitely better than another under any criterion $c \in C$; i.e., $a_{ij} \neq \infty \forall i, j \in A$

Axiom 3 The objectives-problem can be formulated as an hierarchy

Axiom 4 All relevant objectives to the problem are represented in the hierarchy.

When these axioms are fulfilled, weights can be generated.

One of the major advantages of AHP is that, instead of forcing a perfect pairwise-scoring, it allows for human errors that may occur. It is difficult for the PT-authority to exactly estimate the distances between their priorities for different items, so that a result of the pairwise comparison of items a, b and c might be: a = 2b, b = 2c, a = 5c. This kind of noise in the data is a natural variation, caused by human cognitive limits. It is preferred to make this variation insightful and dealing with it, compared to forcing a perfect rational priority matrix, which will still contains the human error, except there's no information about this error anymore. Still, circular prioritizing (a > b > c > a) should be avoided. When this occurs, the PT-authority is asked to revise its priorities. Allowing inconsistency does introduce a problem, in the case that the priorities are perfectly consistent, i.e.,

$$a_{ik} a_{kj} = a_{ij} \forall i, j, k = 1,2,..., n$$

each column j could simply be normalized to yield the final weights:

$$w_j = \frac{a_{ji}}{\sum_k^n a_{ki}} \forall j = 1,2,\ldots, n$$

However, because the priorities will likely not be perfectly consistent, the weights would differ per normalized column, so which column is the right one? Saaty (2003) computes the principal right eigenvector of the matrix A:

$$Aw = \lambda_{\max} w$$

Where $\lambda_{\max}$ is the maximum eigenvalue of the matrix, or:

$$w_j = \frac{\sum_j^n a_{ji} w_i}{\lambda_{\max}} \forall j = 1,2,\ldots, n$$

The way AHP makes the error from the pairwise comparison insightful is by introducing a consistency ratio (CR). The consistency ratio is constructed from the consistency index (CI) and the random inconsistency index (RI). Saaty defines the CI as:

$$C.I. = \frac{\lambda_{\max} - n}{n-1}$$

Where n is the size of a square matrix and $\lambda_{\max}$ is the maximum eigenvalue.

The random index is given by Saaty and is a characteristic of an n by n matrix, the R.I. values are given in Table 17

Table 17 R.I. values for a n-by-n matrix

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
R.I.	0.00	0.00	0.58	0.9	1.12	1.24	1.32	1.41	1.45	1.49	1.51	1.48	1.56	1.57	1.59

The consistency ratio is subsequently:

$$C.R. = C.I./R.I.$$

Saaty advises a CR ≤ 0.1 to be considered acceptable. For higher values, the PT-authority is asked to reduce inconsistencies.

8.2 WEIGHTED SCORING

After the weights are determined, the AHP method is not completed. Since the weights are set per system and not for the overall evaluation, it is not prudent to rank the different systems, instead, it is attempted to evaluate each system on its own merits by generating a score. TOPSIS is a multi-criteria analysis algorithm which is well suited for scoring individual systems instead of ranking them. TOPSIS uses ideal and anti-ideal values for the indicators that are used. In the evaluation framework of Andrade (2008) the ideal and anti-ideal values are set by experts. However, in the interviews that were conducted, it was advised not to use expert opinion, because there simply are no experts available who could make decent estimations of these values. The advice was given to generate ideal and anti-ideal values based on the data that are available. This consequently means that the purpose of TOPSIS is lost in this study, because the ideal and anti-ideal values are not determined independently from the data. However, there was no disadvantage in continuing with this method, so no changes were made. Also, by using this method, future Regiotaxi evaluations can be judged on their own merits using the ideal and anti-ideal values from this study.

The second and third research questions involve the evaluation of the participating Regiotaxi services. The evaluation is performed twice, once the weights for the objectives are split evenly, so no particular weighting is applied. In the second case, weights are applied that were obtained through a survey of the policy priorities of the transport authorities. First the results of the survey and then the two evaluations are presented and the evaluation scores are compared.

8.3 RESULTS SURVEY

The participating Regiotaxi regions were asked to fill out their priorities in the model of Figure 3. The most left column of objectives, including perception, performance and economic durability, will be referred to as the higher or upper hierarchy level and the right side of the objectives as lower hierarchy level. The surveys were filled out by one person for each transport authority, so in total four surveys were finished. It was assumed for every transport authority that the survey properly reflects their policy priorities and that it was not necessary to survey several people from the same transport authority. For the reporting of policy priorities by the transport authorities, the pair wise comparison of items, as also used in the AHP methodology, was used. Three sets of weights were generated, on two hierarchy levels. The questionnaire can be found in appendix E. Because an indirect ranking system was used, a consistency rating was calculated in order to test whether consistency was acceptable. The three sets of weights were returned by four transport authorities. Of the total twelve sets of weights, three were returned for revision because inconsistencies were too large, one involving a circular a>b>c>a type of error. After the revision, all the sets of weight were considered consistent enough to be used, i.e. they had a consistency ratio smaller than 0.1.

Table 18 Weights of higher hierarchy level

	Gelderland	Overijssel	Utrecht	West-Brabant
Perception	63.7%	45.5%	48.1%	23.9%
Performance	25.8%	45.5%	40.5%	13.7%
Economic durability	10.5%	9.1%	11.4%	62.5%
Consistency ratio	0.0332	0	0.0251	0.0572

What stands out in Table 18 is the difference between West-Brabant and the others. West-Brabant values economic durability above perception and performance; the reverse is true for the others. The representative from Gelderland mentioned in conversation that economic durability was seen by the Province as a precondition for the service, but that beyond being able to keep the system alive, the other factors were more important. Figure 7 shows that perception is generally considered most important, then performance and economic durability is considered least important.

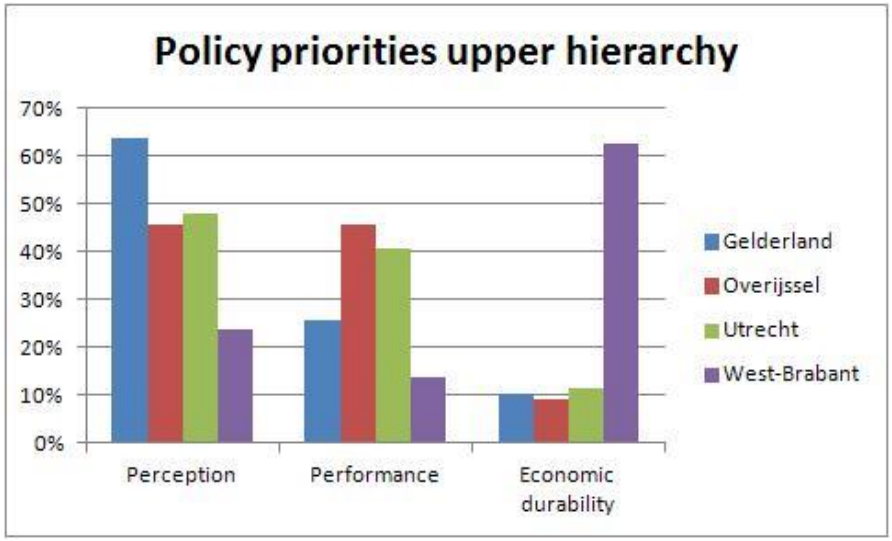

Figure 7 Policy priorities for the higher hierarchy

Table 19 Policy priorities for perception

	Gelderland	Overijssel	Utrecht	West-Brabant
Comfort	10.5%	14.7%	6.7%	4.0%
Information	27.3%	11.7%	21.5%	7.5%
Travel time	5.7%	14.7%	35.3%	28.8%
Safety	29.1%	35.1%	19.5%	28.8%
Service	27.3%	23.7%	17.0%	30.5%
Consistency ratio	0.0587	0.0441	0.0621	0.0297

Five items were included in the survey for the weights for perception. As mentioned in chapter 6, the objectives included in the model are not complete. Although information provision is only one of several objectives that could potentially be added to the perception tree of the model, it is considered the most relevant objective that is missing. In order to get a more complete overview of policy priorities it was included in the survey. However, for the evaluation score the item was excluded and the other weights were redistributed proportionately. Table 19 shows the policy priorities for perception. Table 20 shows the weights as they were used in the evaluation.

Table 20 Perception weights used in the evaluation

	Gelderland	Overijssel	Utrecht	West-Brabant
Comfort	14.5%	16.7%	8.5%	4.8%
Travel time	7.8%	16.7%	45.0%	31.1%
Safety	40.1%	39.7%	24.9%	31.1%
Service	37.6%	26.9%	21.7%	33.0%

What stands out in Table 19 for the weights given to the perception objectives is that besides comfort, each factor is highly valued by one region or the other. Utrecht seems to have more priority on travel time and information provision, whereas the other regions all value safety and service. West-Brabant also has a focus on travel time and less on information provision, while the reverse is true for Gelderland.

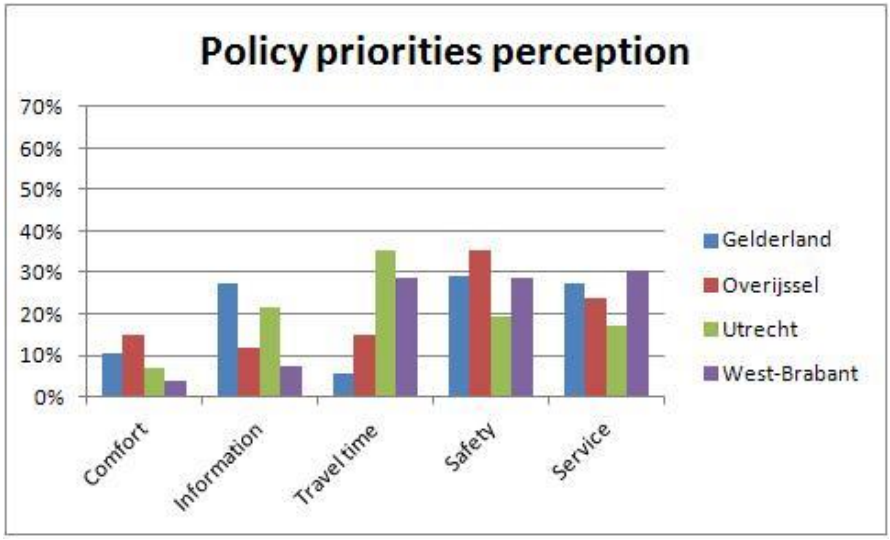

Figure 8 Policy priorities for perception objectives

Figure 8 shows that in general service and safety are the more important objectives and comfort is not highly prioritized.

Table 21 Weights for performance

	Gelderland	Overijssel	Utrecht	West-Brabant
Reliability	30%	64.0%	56.5%	61.3%
Travel time	10%	23.5%	26.2%	16.9%
Availability	30%	6.3%	5.5%	9.6%
Tariff	30%	6.3%	11.8%	12.3%
Consistency ratio	0	0.0774	0.0433	0.0572

For performance in Table 21 it is clear that all regions put emphasis on reliability. Travel time is also very important to all regions except for Gelderland. This is consistent with the priorities of Gelderland in perception, where they also had a low priority on travel time. This is reflected in Figure 9

The weights were used in the evaluation and set against the unweighted evaluation scores. The results of these evaluations are elaborated in the next chapter.

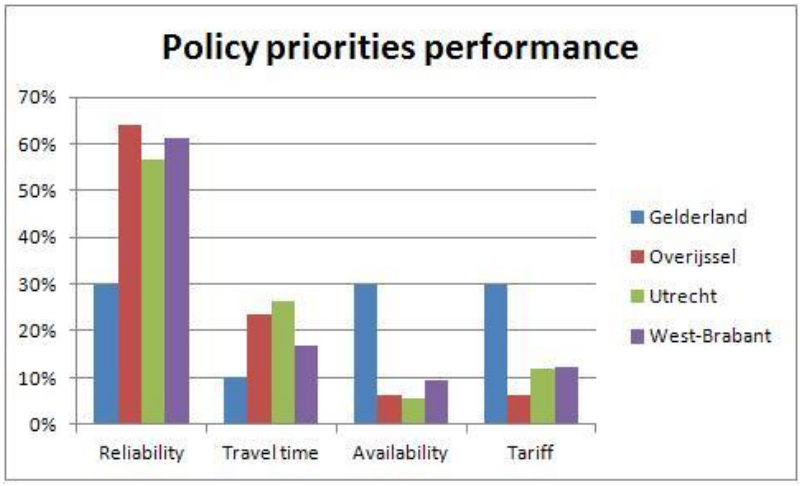

Figure 9 Policy priorities for performance objectives

9 REGIOTAXI EVALUATION

9.1 BENCHMARKS

The first research question concerned the performance of Regiotaxi services on different indicators:

What are benchmarks for relevant performance indicators and how are these indicators related?

In Table 22 the benchmarks of the indicators from the evaluation framework are presented. For the empty cells there are no data available. The services from Overijssel and West-Brabant have 4 missing values. Eemland-Heuvelrug has the most complete data, only missing 1 benchmark, the other system miss 2 benchmarks.

For the satisfaction benchmarks, most values are between 7 and 8. The main exception to this is 'processing complaints' which scores lower than 7 almost across the board. Generally, satisfiers like comfort and cleanliness seem to score better than dissatisfiers like call center availability, complaint processing and travel time.

It is surprising to see the variation in prices. Figure 10 sets the Regiotaxi services against each other. Especially there is a large variation in the public transport tariff, ranging from 38 to 72 cent per kilometer. The figure also shows that the tariff does not seem to have an effect on subsidy. In fact, the two variables are even positively, but not significantly, correlated. This is surprising as it is expected that a higher subsidy leads to lower trip prices.

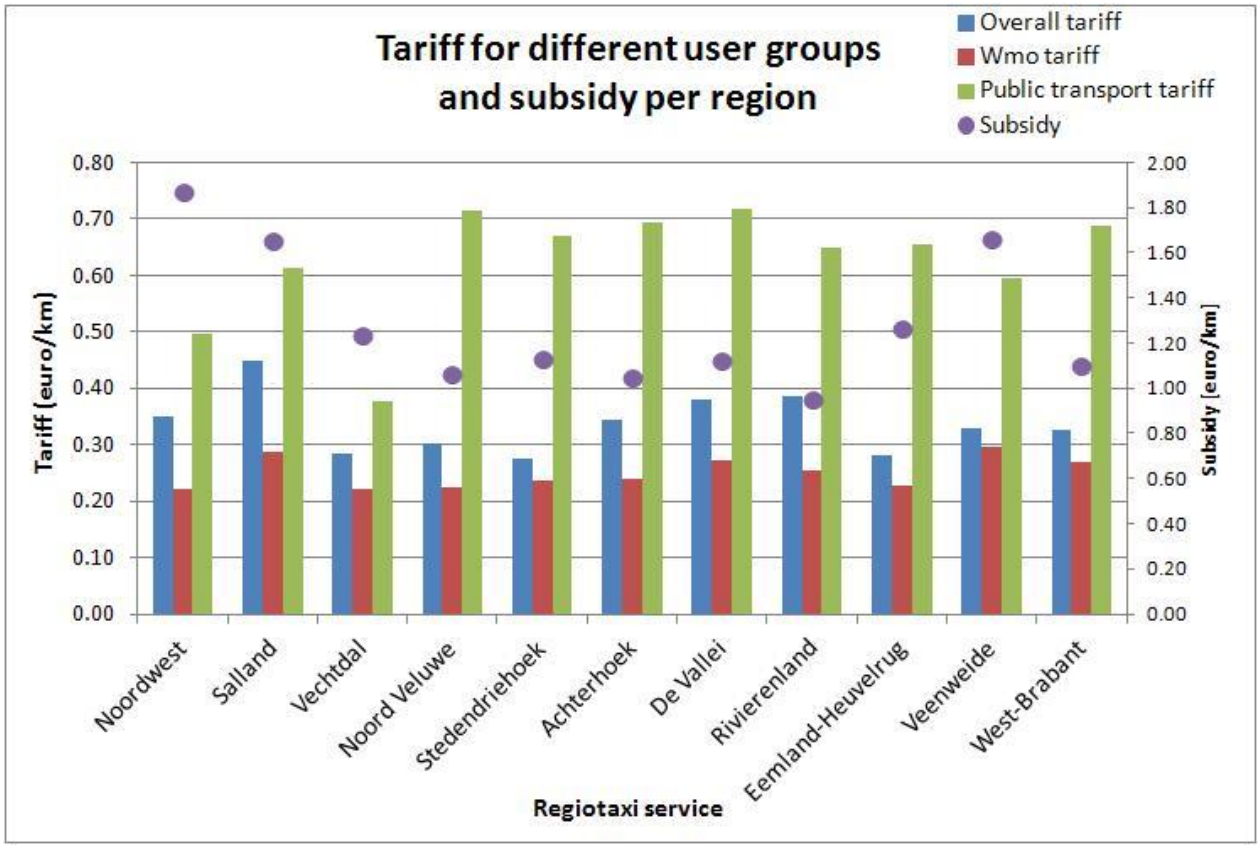

Figure 10 Tariff for different user groups and subsidy per Regiotaxi system

Table 22 Benchmarks

	Unit	NW***	SL	VD	NV	SD	AH	VA	R	EH	V	WB
Comfort	CSS*	7.7863	7.7920	7.7917	-	-	-	-	-	7.8962	7.5057	7.3679
Cleanliness	CSS	7.5836	7.5812	7.5790	8.1724	8.0257	7.9157	8.0233	8.3491	7.9888	7.9146	7.6018
Driving style	CSS	7.6200	7.6194	7.6189	8.1082	8.0818	8.0512	8.0064	8.3168	7.8490	7.7622	7.3606
Safety	CSS	-	-	-	8.2383	8.0991	8.1308	8.1039	8.3952	7.7524	7.4873	-
Friendliness call center	CSS	7.5183	7.5331	7.5405	-	-	-	-	-	8.0291	8.0622	-
Friendliness chauffeur	CSS	7.4991	7.5038	7.4971	7.9762	7.5102	7.8722	7.9737	7.8913	8.0742	8.0895	7.7107
Helpfulness call center	CSS	-	-	-	7.8947	7.3659	7.9412	7.8551	7.7442	-	-	7.4135
Helpfulness chauffeur	CSS	-	-	-	7.9964	7.8166	7.9724	7.8608	8.1517	7.8044	7.8411	7.6046
Availability call center	CSS	7.2129	7.2220	7.2661	7.5476	6.8723	7.1111	7.2133	7.2660	7.7500	7.8061	7.0148
Processing complaints	CSS	-	-	-	6.8571	5.9231	6.4839	6.1818	5.8421	6.4597	6.3058	7.1391
Travel time satisfaction	CSS	7.5720	7.5693	7.5665	7.6520	6.9368	7.3995	7.2880	7.8501	7.1256	6.8903	6.5749
Overall Satisfaction	CSS	8.0916	8.0560	8.0417	7.9773	7.4734	7.8076	7.7143	8.1484	7.9212	8.1062	7.7759
Punctuality	ratio	0.9583	0.9534	0.9106	0.9430	0.8445	0.8672	0.9003	0.8684	0.9637	0.9549	-
Availability	hours/ day	19.5	19.5	19.5	19.0	19.0	19.0	19.0	19.0	18.0	18.0	19.0
Travel time within rules	ratio	0.7165	0.6589	0.8853	0.9685	0.9614	0.9732	0.9411	0.9778	0.9704	-	0.9716
Travel time ratio	ratio	1.6186	2.3095	1.2789	1.4179	1.4654	1.4307	1.6542	1.3196	1.5057	1.8822	1.8963
Travel time total ratio	ratio	1.4168	1.9505	1.2368	1.3317	1.3796	1.3461	1.5908	1.2389	1.2535	1.7265	1.8469
Tariff	euro/ km**	0.3498	0.4471	0.2854	0.3033	0.2748	0.3442	0.3796	0.3846	0.2797	0.3300	0.3273
Subsidy	euro/km	1.8748	1.6610	1.2414	1.0653	1.1335	1.0557	1.1269	0.9582	1.2670	1.6675	1.1081
Subsidy	euro/ inhabitant/year	9.92	12.90	3.23	7.01	8.76	11.78	8.91	8.91	10.25	7.87	3.76
Cost cover	ratio	0.1706	0.2442	0.1924	0.2768	0.2795	0.2989	0.2889	0.2884	0.1621	0.1073	0.2309

*CSS = Rescaled score from the customer satisfaction survey

**Indicators per km use the total amount of kilometers driven for passengers (including non-subsidized km)

***NW = Noordwest, SL = Salland, VD = Vechtdal, NV = Noord-Veluwe, SD = Stedendriehoek, AH = Achterhoek, VA = de Vallei, R = Rivierenland, EH = Eemland-Heuvelrug, V = Veenweide, WB = West-Brabant

In paragraph 6.3.4 it was mentioned that the way tariff is generated, trip length should influence the scores. Now that the scores have been generated this is checked. Tariff and average trip length have a negative correlation (-0.34). That indicates that systems with a longer average trip indeed get a lower tariff. However, the correlation is not significant, so although there seems to be an influence from trip length, it is not definite.

9.2 COMPARING SYSTEMS USING TOPSIS

TOPSIS assumes there are *m* alternatives (regions) and *n* objectives and there's a score x_{ij} and a weight w_{ij} for each objective *j* for each region *i*. There's a matrix X = (x_{ij}) which is an *m* by *n* matrix. K is the set of ideal values and K' is the set of anti-ideal values.

First the scores x_{ij} are normalized

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_i {x_{ij}}^2}} \forall\, i = 1, \dots, m; j = 1, \dots, n$$

The weight vector w_j per system was already generated with AHP. The weighted normalized matrix *V* can now be generated. These values are the scores for their respective objectives per region.

$$v_{ij} = w_{ij} * r_{ij}\ \forall i = 1, \dots, m; j = 1, \dots, n$$

The ideal and anti-ideal values are set by using the available data. The ideal value $k_j \in K$ and the anti-ideal value $k'_j \in K'$ are set as the mean of all the scores of the region plus or minus two times the standard deviation.

$$k_j = \mu_j + 2\sigma_j\ \forall\, j = 1, \dots, n$$

$$k'_j = \mu_j - 2\sigma_j\ \forall\, j = 1, \dots, n$$

Where μ_j is the mean score of objective *j* and σ_j is the standard deviation. By using these boundaries, it is possible to have values better than ideal or worse than anti-ideal. Per definition of standard deviation, 5% of values could be outside these boundaries. Because the definition of ideal requires it to be the most optimal value, values better than ideal or worse than anti-ideal are equated to the ideal and anti-ideal values respectively. Because for some indicators this method does not work, some exceptions are made. In paragraph 9.5 a sensitivity analysis of the boundaries is performed and possible issues are discussed.

The next step is to calculate the separation *S* of the scores from the ideal and anti-ideal scores.

$$S_i = \sqrt{\sum_j (k_j - v_{ij})^2}\ \ \forall\, i = 1, \dots, m$$

$$S'_i = \sqrt{\sum_j (k'_j - v_{ij})^2}\ \ \forall\, i = 1, \dots, m$$

Finally, the relative closeness to the ideal solution C_i is determined using the separation from the ideal and anti-ideal values.

$$C_i = \frac{S'_i}{(S_i + S'_i)}$$

This analysis is performed three times per region, for perception, performance and economic durability. These three scores are then combined using their respective weights to yield a final score F.

$$F_i = C_{i,1} * w_{i,1} + C_{i,2} * w_{i,2} + C_{i,3} * w_{i,3}$$

A score of 0 means the system is performing at an anti-ideal level, a score of 1 means an ideal performance.

9.3 EVALUATION SCORES

TOPSIS generates a score by finding ideal and anti-ideal values and determining the distance to both. The score becomes higher, as the real value is closer to the ideal or further from the anti-ideal.

Table 23 shows the normalized ideal and anti-ideal values for the indicators from the model in Figure 3.

The ideal value can be both a higher or lower number than the anti-ideal, depending on the unit that was used for the indicator. For example, for any satisfaction measure, a higher grade is better. However, for evaluating tariff, a lower tariff is better than a higher tariff (for the purpose of this evaluation).

As for the perception scores, ideal and anti-ideal scores are seemingly close together. This is possibly due to the small part of the scale that is used for customer satisfaction.

The boundaries in Table 23 are used to generate a score per indicator which is then combined into a score for the objectives lower in the hierarchy. These scores are shown in Table 24. The scores in Table 24, Table 25 and Table 26 are between 0 and 1, where a score of 1 is a perfect score and a score of 0 equals an anti-ideal situation. Table 25 shows the higher hierarchy scores and Table 26 the final scores per Regiotaxi.

For creating a final score, two methods were used side by side: An unweighted method and a weighted method. In the unweighted method all lower hierarchy scores count equally for forming the higher hierarchy objective score. For the weighted score the weights from policy priorities provided by the transport authorities are taken into account.

Figure 11 shows that Noord Veluwe and Rivierenland stand out with their perception scores. As for the performance scores in Figure 12, the Regiotaxi systems of Overijssel are doing the best. Because of the reasons discussed in paragraph 8.2 the figures cannot be used to judge the objective general performance of the Regiotaxi systems on the different objectives.

Table 23 Ideal and anti-ideal boundaries

	Ideal	Anti-ideal
Comfort	0.3128	0.2901
Cleanliness	0.3217	0.2810
Driving Style	0.3232	0.2795
Safety	0.3193	0.2835
Friendliness personnel call center	0.3154	0.2875
Friendliness chauffeur	0.3202	0.2825
Helpfulness personnel call center	0.3153	0.2875
Helpfulness Chauffeur	0.3118	0.2911
Availability Call Center	0.3253	0.2773
Processing Complaints	0.3358	0.2663
Travel Time CSS	0.3331	0.2691
Overall score CSS	0.3172	0.2857
Punctuality	0.3289	0.2735
Availability	0.3180	0.2848
% Correct travel time	0.3732	0.2257
Average ratio actual/ free flow travel time	0.3972	0.1983
Ratio total travel time and free flow travel time	0.3915	0.2049
Tariff	0.2054	0.3911
Subsidy per kilometer	0.1552	0.4331
Subsidy per inhabitant	0.0855	0.4864
Cost Cover	0.4533	0.1297

Table 24 Lower hierarchy scores

	Comfort	Safety	Service	Travel Time	Punctuality	Availability	Travel time	Tariff	Economic durability
Noordwest	0.3486	0.3821	0.4460	0.6677	0.7485	0.7611	0.4109	0.4385	0.2840
Salland	0.3502	0.3818	0.4491	0.6660	0.7194	0.7611	0.0532	0	0.3054
Vechtdal	0.3490	0.3816	0.4586	0.6642	0.4654	0.7611	0.7075	0.7454	0.6299
Noord Veluwe	0.6936	0.7215	0.7516	0.7192	0.6577	0.5218	0.6501	0.6601	0.6532
Stedendriehoek	0.5995	0.6465	0.1962	0.2589	0.0730	0.5218	0.6024	0.7959	0.5725
Achterhoek	0.5219	0.6455	0.5148	0.5567	0.2078	0.5218	0.6395	0.4652	0.4953
De Vallei	0.5978	0.6110	0.4286	0.4850	0.4043	0.5218	0.4382	0.2964	0.5780
Rivierenland	0.7731	0.8966	0.3449	0.8468	0.2149	0.5218	0.7500	0.2726	0.6039
Eemland-Heuvelrug	0.6511	0.3897	0.6441	0.3804	0.7806	0.0430	0.6464	0.7726	0.3586
Veenweide	0.4498	0.2852	0.6003	0.2290	0.7283	0.0430	0.3142	0.5328	0.3461
West-Brabant	0.1945	0.2894	0.6095	0.0260	0.5000	0.5218	0.3331	0.5457	0.6912

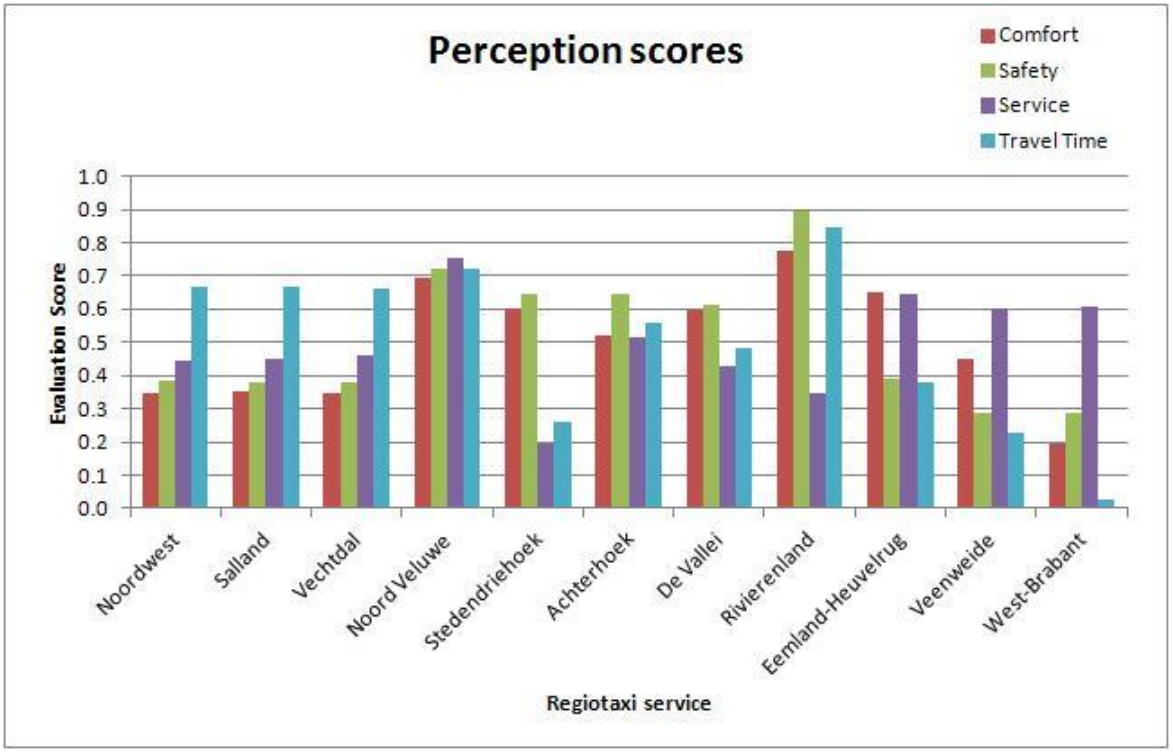

Figure 11 Perception scores per region

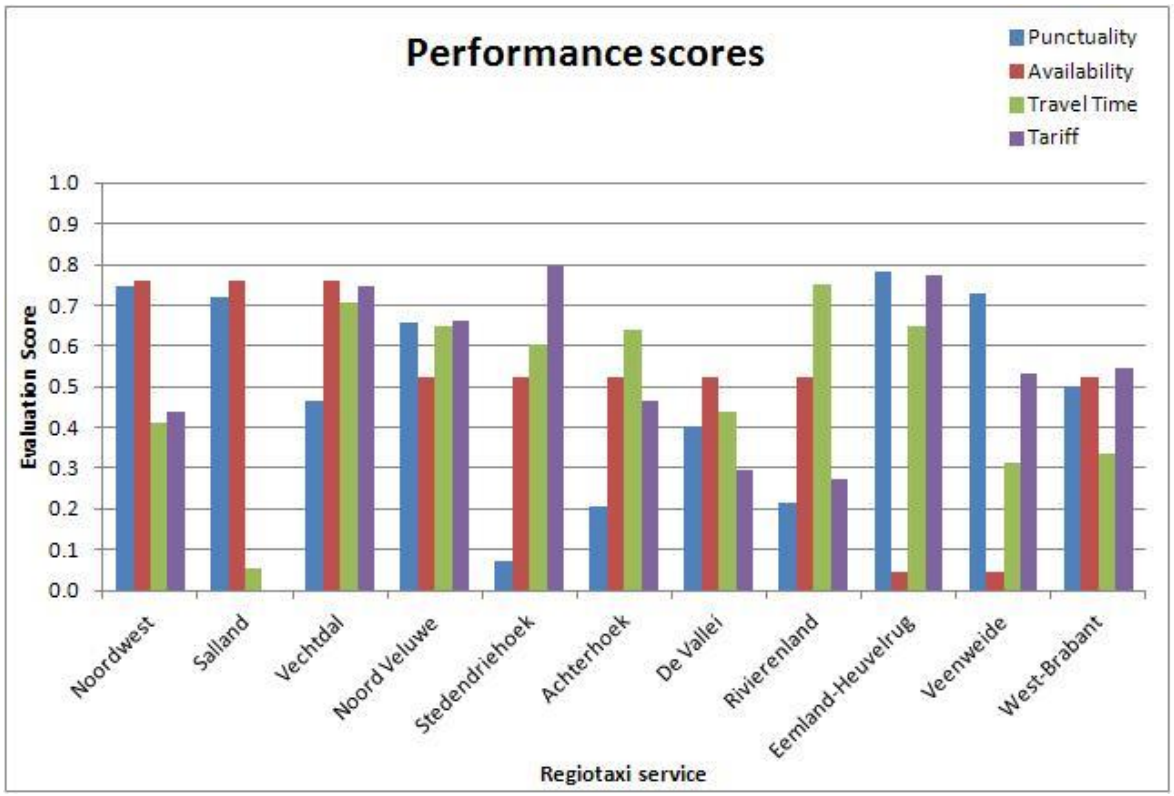

Figure 12 Performance scores per region

Table 25 Higher hierarchy scores

	Unweighted			Weighted		
	Perception	Performance	Economic durability	Perception	Performance	Economic durability
Noordwest	0.1949	0.1966	0.0947	0.2612	0.2957	0.0258
Salland	0.1878	0.1278	0.1018	0.2516	0.2365	0.0278
Vechtdal	0.1852	0.2233	0.2100	0.2481	0.2538	0.0573
Noord Veluwe	0.2152	0.2075	0.2177	0.4136	0.1593	0.0684
Stedendriehoek	0.0709	0.1661	0.1908	0.1402	0.1233	0.0600
Achterhoek	0.1541	0.1529	0.1651	0.2983	0.1091	0.0519
De Vallei	0.1305	0.1384	0.1927	0.2493	0.1060	0.0605
Rivierenland	0.2487	0.1466	0.2013	0.4600	0.0976	0.0632
Eemland-Heuvelrug	0.1697	0.1869	0.1195	0.2319	0.2853	0.0409
Veenweide	0.1861	0.1349	0.1154	0.2566	0.2266	0.0395
West-Brabant	0.1011	0.1584	0.2304	0.0757	0.0655	0.4942

Table 26 Final scores

Regiotaxi	Unweighted	Weighted
Noordwest	0.4861	0.5827
Salland	0.4174	0.5159
Vechtdal	0.6184	0.5592
Noord Veluwe	0.6404	0.6413
Stedendriehoek	0.4278	0.3235
Achterhoek	0.4720	0.4592
De Vallei	0.4615	0.4159
Rivierenland	0.5966	0.6208
Eemland-Heuvelrug	0.4762	0.5580
Veenweide	0.4364	0.5226
West-Brabant	0.4899	0.6354

9.4 RELATIONS BETWEEN THE SCORES

In this paragraph the correlation between the indicator scores and evaluation scores are considered in order to answer the last part of the first research question: "*What are benchmarks for relevant performance indicators and how are these indicators related?".* The correlations are shown in Table 27. Some remarkable values are discussed next. First of all the significant relations are covered and then the lack of correlations where it is expected, or counter-intuitive. Given that eleven regions are scored, the correlations above 0.6 are significant for an α of 0.05. The significant relations are bold in the table. No multiple regression was performed, because for data of only eleven regions, no significant relations were expected for any combination of factors.

9.4.1 SIGNIFICANT RELATIONS

A strong correlation (0.77) is observed between travel time satisfaction and perception. This may suggest that even though actual travel time satisfaction does not relate to travel time performance, it is important to somehow improve travel time satisfaction in order to improve overall satisfaction.

Perceived safety correlates strongly with perceived comfort (0.81). This is as expected, as it is unlikely for one to be positive without the other, i.e. for people to be comfortable while feeling unsafe or, to a lesser extent, for people to feel safe while not being comfortable. In one of the expert interviews it was pointed out that asking for satisfaction about safety is not very useful, since travelers do not have a decent grasp on how safe they actually are. The relation between safety and comfort possibly supports this; the travelers might actually be reporting satisfaction about safety based on their comfort level. For the design of the system that could mean that perception of safety can be improved by improving comfort. It would also mean that satisfaction about safety is not very useful as indicator to actual safety.

The next remarkable significant relation is between punctuality and perceived service (0.67). The perceived service objective is for the most part based on interaction with personnel. Therefore it is possible that having a reliable system goes a long way for the traveler towards feeling they are being treated and helped properly.

Travel time performance and perceived comfort also have a significant relation and the relation between travel time performance and safety is nearly significant. All the indicators for the objective travel time perform worse when the trip takes longer. Therefore this correlation suggests that increasingly long trips lead to lower comfort, which is not illogical.

Looking at the relation with system characteristics, cost cover has a negative correlation with average trip distance (-0.70). This is interesting for setting prices; if the call out fee is supposed to cover the costs of getting a vehicle at the origin and the kilometer charge has to cover the passenger kilometers, it seems that the call out fee is high and the kilometer charge is low.

Other significant relations are between:

- Total score and performance
- Total score and travel time performance
- Overall satisfaction and perception
- Punctuality and safety

These objectives do not have easily defined relations. All the higher hierarchy objectives (perception, performance, economic durability) are similarly correlated to total score, which is just a consequence of the total score being a result of those objectives. The relation between travel time and total score is also not clear.

Table 27 Correlation between objectives

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Total Score	1.0000												
2. Perception	0.5666	1.0000											
3. Performance	**0.6532**	0.1625	1.0000										
4. Economic Durability	0.5809	-0.2283	0.2264	1.0000									
5. Comfort	0.2927	0.3045	-0.0141	0.1594	1.0000								
6. Safety	0.4163	0.2396	-0.0549	0.4453	**0.8116**	1.0000							
7. Service	0.2801	0.2932	0.2506	-0.0166	-0.0914	-0.3200	1.0000						
8. Travel time satisfaction	0.5526	**0.7747**	0.2480	-0.0758	0.3993	0.5366	-0.1151	1.0000					
9. Punctuality	-0.0487	0.3815	0.1821	-0.5790	-0.3134	**-0.6465**	**0.6691**	0.0008	1.0000				
10. Availability	0.2415	0.0446	0.2145	0.1979	-0.3493	0.1465	-0.4001	0.4899	-0.2238	1.0000			
11. Travel time	**0.6488**	0.1667	0.5535	0.4985	**0.6208**	0.6027	-0.0682	0.3019	-0.4790	-0.1104	1.0000		
12. Tariff	0.2593	-0.3263	**0.6760**	0.3096	0.1151	-0.0992	0.1760	-0.3763	-0.0917	-0.3617	0.5971	1.0000	
13. Overall Satisfaction	0.3891	**0.9126**	0.1158	-0.3886	-0.0568	-0.1427	0.3257	0.5547	0.5688	0.0272	-0.0962	-0.3442	1.0000

The only way that could be explained is if travel time performance has a great impact on perception. However, this is not at all the case, as those objectives are only very weakly correlated. Overall satisfaction and perception are logically strongly correlated as overall satisfaction makes up a large part of perception. There's also no obvious explanation for the negative correlation between punctuality and perceived safety.

9.4.2 UNEXPECTED INSIGNIFICANT OR COUNTER-INTUITIVE RELATIONS

The insignificant results are also of interest, in the case where a significant relation is expected. This is discussed in this paragraph as well as counter-intuitive relations.

First of all, economic durability has a very weak correlation with perception and performance (-0.23 and 0.23 respectively). This is a surprise, as this objective was included in the model to counter-balance perception and performance. After all, it is expected to be easier to score well on performance and perception, if there's more money available to run the system. A good score on economic durability means that relatively little subsidy is available to the system compared to the amount of transport that is provided. The weak correlation suggests that the amount of subsidy that is used in the system is not as closely related to performance as one would expect. Maybe it is not that unexpected though, because the lack of relation between performance and subsidy is also found in conventional public transport. For example studies in Indiana and California even found a negative correlation between subsidy and performance for public transport (Karlaftis & McCarthy, 1997) (Cervero, 1984). This is also visible in the relation between cost cover and subsidy per kilometer. There is a significant negative relation between these indicators, so systems that receive a higher subsidy per kilometer have a worse cost cover.

A second important observation is that the perception concerning travel time correlates weakly with the actual travel time. Therefore, it is possible that improving actual travel time does not automatically lead to higher satisfaction on travel time. One possible explanation could be that travelers also take other time-related indicators into account for reporting satisfaction on travel time, like punctuality. However, punctuality also does not correlate with travel time satisfaction.

Thirdly, tariff has a negative correlation with perception. This may indicate that the travelers' attitude is not greatly influenced by tariff. A lower tariff leads to a better tariff score, so the negative correlation suggests that systems with higher tariffs also have higher satisfaction among customers. This could indicate that Regiotaxi users are willing to pay for the service and do not value a low tariff highly.

9.5 SENSITIVITY ANALYSIS

The generated scores depend on the choice for the ideal and anti-ideal boundaries. In order to see what the influence is of the boundaries, a sensitivity analysis is presented in this paragraph and possible alternatives for picking the boundaries are discussed.

In the framework of Andrade (2008) the ideal and anti-ideal boundaries are supposed to be set by experts. In this study experts advised to use the available data, as that would be better suited than asking experts. Therefore, in order to be consistent, boundaries of two standard deviations above and below the mean were chosen. The disadvantage of this method is, that for indicators the regions score relatively equal on, small differences between regions will still have an impact on the total score and vice versa the indicators with greatly varying performances will have a relatively too small impact on the total score. Small variations on indicators can especially occur when the design of the services is not independent, i.e. regions look at how other regions are setting up their services and they mirror the design choices. This leads to less variation than if the regions had designed their services independently. The advantage is that arbitrariness is avoided. In the interviews it was indicated that it would be difficult to assign proper values to the ideal and anti-ideal boundaries. By using the available data, wrong estimates are avoided.

For setting boundaries for the 10-point satisfaction scale, it could be argued that all the indicators should have the same boundaries, because the 10-point scale is already standardized. The counter argument to this would be, that the standardization is not reliable for different types of questions, as for example the attitude towards dissatisfiers is generally worse than the attitude towards satisfiers. This is also reflected in the results of Table 22 where the satisfaction about processing complaints is much lower than for other indicators. So while an '8' might already be very good for a dissatisfier, it is not extremely high for a satisfier.

For two cases scores are generated to get a sense of the influence of the aforementioned design choices: in the first case wider boundaries are chosen for the indicator for availability, in the second case one set of boundaries is used for all the questions on the 10-point satisfaction scale.

Table 28 Scores for availability for different sets of boundaries

Design of boundaries:	mean +/- 2σ (this study)	10h or 24h service (alternative)
Noordwest	0.7611	0.6786
Salland	0.7611	0.6786
Vechtdal	0.7611	0.6786
Noord Veluwe	0.5218	0.6429
Stedendriehoek	0.5218	0.6429
Achterhoek	0.5218	0.6429
De Vallei	0.5218	0.6429
Rivierenland	0.5218	0.6429
Eemland-Heuvelrug	**0.0430**	**0.5714**
Veenweide	**0.0430**	**0.5714**
West-Brabant	0.5218	0.6429

In this study, 'operating hours' is an indicator for which using two standard deviations to set the boundaries is less appropriate. Using this method for setting boundaries the anti-ideal situation is 17.9 operating hours and the ideal situation is 20 operating hours. However, when selecting boundaries separately from the other indicators, a more realistic set of boundaries would intuitively be further apart. Therefore a new set of boundaries is arbitrarily chosen. As anti-ideal situation daily operating times from 8.00 to 18.00 are taken. The ideal situation is set at a 24 hour service.

The change of boundaries is very impactful for this particular indicator, as can be seen in Table 28. Especially the Regiotaxi services in Utrecht (Eemland-Heuvelrug & Veenweide) have much better scores. Next, the impact on the higher level scores is compared in Table 29, and it can be seen that the large influence to the availability score for the regions in Utrecht carries over to the performance and total score. The changes now better reflect the perceived performance on availability, because the difference between operating 18 hours versus operating 19.5 hours no longer reflect the difference between nearly anti-ideal and nearly ideal. The concern for the boundaries not being appropriate when using two standard deviations from the mean therefore seems justified and care should be taken when determining how boundaries are set.

All the performance indicators are from the rescaled 10-point customer satisfaction survey questions. Table 30 compares the scores of this study with the scores when a single set of boundaries is used for all the answers on the 10-point satisfaction scale. For the alternative, '6' is considered to be the anti-ideal and '9' is set as the ideal boundary. These values are chosen somewhat arbitrarily, but at least in a manner that they encompass most of the scores. Table 30 shows that the new boundaries improve the scores for low scoring regions and

reduce the scores for the highest scoring regions. This is a consequence of the boundaries of '6' and '9' being further apart than the boundaries for the mean +/- two standard deviations are. Therefore the differences between regions are reduced. This also carries over to the total scores, especially Stedendriehoek and West-Brabant benefit greatly from the new boundaries.

Table 29 Comparison of the performance and total scores from using two different sets of boundaries for availability

	Performance (this study)	Performance (alternative)	Total score (this study)	Total score (alternative)
Noordwest	0.1966	0.1897	0.4861	0.4792
Salland	0.1278	0.1209	0.4174	0.4105
Vechtdal	0.2233	0.2164	0.6184	0.6116
Noord Veluwe	0.2075	0.2176	0.6404	0.6505
Stedendriehoek	0.1661	0.1762	0.4278	0.4379
Achterhoek	0.1529	0.1629	0.4720	0.4821
De Vallei	0.1384	0.1485	0.4615	0.4716
Rivierenland	0.1466	0.1567	0.5966	0.6067
Eemland-Heuvelrug	**0.1869**	**0.2309**	**0.4762**	**0.5202**
Veenweide	**0.1349**	**0.1789**	**0.4364**	**0.4804**
West-Brabant	0.1584	0.1685	0.4899	0.4999

Table 30 Comparison of the perception and total scores from using a single set of boundaries for all 10-point satisfaction scales

	Perception (this study)	Perception (alternative)	Total score (this study)	Total score (alternative)
Noordwest	0.1949	0.2031	0.4861	0.4944
Salland	0.1878	0.2012	0.4174	0.4308
Vechtdal	0.1852	0.2005	0.6184	0.6337
Noord Veluwe	0.2152	0.2101	0.6404	0.6353
Stedendriehoek	**0.0709**	**0.1637**	**0.4278**	**0.5207**
Achterhoek	**0.1541**	**0.1917**	0.4720	0.5096
De Vallei	**0.1305**	**0.1844**	0.4615	0.5154
Rivierenland	0.2487	0.2218	0.5966	0.5697
Eemland-Heuvelrug	0.1697	0.1941	0.4762	0.5005
Veenweide	0.1861	0.1951	0.4364	0.4453
West-Brabant	**0.1011**	**0.1689**	**0.4899**	**0.5577**

The conclusion of this sensitivity analysis is that the way the boundaries are chosen can significantly affect the total scores and therefore the results should be handled with care. The method used in this study, using two standard deviations, avoids arbitrary boundaries, however, since in this study only 11 regions are evaluated, it suffers from a lack of data. The method would probably provide better results if the data of 25+ regions could be used to set the boundaries, but this was not possible within the scope of this study. It still leaves the issue of regions copying each other's design choice and therefore creating a lack of variation to base the boundaries on.

10 ANALYSIS OF THE ACCOMPLISHMENT OF POLICY GOALS

In order to see whether policy priorities influence the scores of the objectives, two analyses have been performed. Firstly, a t-test was performed to determine whether the weighted scores are higher than the unweighted scores. Secondly, it is tested whether there is a relationship between the relative priority the transport authorities give to policy objectives and the relative scores they get for these objectives.

For the t-test, the null hypothesis H_0 is that the unweighted score S_{uw} and weighted score S_w are the same and the alternative hypothesis H_a is that the weighted score is better:

$$H_0: S_{uw} - S_w = 0$$

$$H_a: S_{uw} - S_w < 0$$

For an α=0.05, the null hypothesis cannot be rejected and therefore it cannot be definitely said that the weighted scores differ from the unweighted scores on the same objectives.

So comparison of the unweighted and weighted scores suggests that policy is not significantly influencing policy outcomes. Next, a check is performed to see whether a relation can be found between the policy priorities and the scores on the objectives. This is done by taking the relative priority of the lower hierarchy level objectives compared to the other regions.

The hypothesis is that a higher priority will lead to a higher score. In order to put these values against each other, they have to be adjusted. The objective score has to be adjusted for the total score of the region. For example: Two regions have a total score of respectively 0.3 and 0.5 and they both have a score of 0.7 for perceived safety, which they both give the same high priority. It could then be argued that the region with a total score of 0.3 has its high priority for perceived safety better reflected in its score, because score for perceived safety is more distinct from the system average than the region with a total score of 0.5. Since both the score for perceived safety and the total score are standardized to a scale of 0 to 1 for anti-ideal to ideal, the indicators are compared based on the score relative to the total score, by subtracting the total score from the objective score for perceived safety. This relative score lies between -1 and 1.

Similarly, the actual priorities cannot be used for evaluating the policy effects. When a system has a low score relative to other systems, a low priority does not necessarily indicate anything. If other authorities assign an even lower priority to that objective, it still would have been expected for this region to have a relatively higher score. Therefore, instead of actual priority, the priority relative to the other regions is the parameter of interest.

The absolute priority is determined as the weight that has been given to the lower hierarchy level objective multiplied by the higher hierarchy level weight, including a correction for the amount of items in the lower hierarchy level. For example: The performance objective 'reliability' has a weight of 30% for Gelderland. Performance has a 25.8% weight. There are four performance indicators, which means if no distinction is made, reliability would have a weight of 25%. There has to be a correction for the amount of indicators, otherwise the absolute priorities for performance indicators would be much smaller than for economic durability, because it is split up in more items. Using this correction means the absolute priority for reliability for Gelderland becomes:

$$Priority_{abs} = \frac{0.30}{0.25} * 0.258 = 0.3096$$

The relative priority is then calculated by normalizing the absolute priorities. Because of the correction of amount of items, the absolute priority no longer represents the weight of the objective, i.e. in the example the weight of reliability for Gelderland for the total score would be 7.74%, instead of 30.96%.

The relative priority is compared to the relative score. The relative score reflects the way a system scores on an objective compared to how the system scores as a whole. Since all the scores are already normalized and reflect the distance from ideal and anti-ideal boundaries, this can simply be done by subtracting the total score from the objective score. A negative relative score means that on that particular objective the system scores worse than on the total score; a positive score means the system scores better. If there's a relation between priorities and scores, it is to be expected that as the relative priority is higher, the relative score is also better. Figure 13 is an example of what the expected relation should be. Figure 14 shows the results for every couple of relative weight and score for each of the 9 lower hierarchy objectives for all 11 systems. The figures show the lack of relation in the actual situation.

This analysis also does not provide any indication that there's a relation between policy priorities and scores. Table 31 shows the correlation values between relative priority and relative score. All the points put together show a significant but weak correlation of 0.2075. Overijssel and West-Brabant have moderate correlations between the relative scores and priorities, but it is only significant for Overijssel.

Table 31 Correlations between relative priorities and scores

	Data points	Correlation relative priorities and relative score	Significant
Total	99	0.2075	yes
Overijssel	27	0.4729	yes
Gelderland	45	0.1293	no
Utrecht	18	0.0432	no
West-Brabant	9	0.4866	no

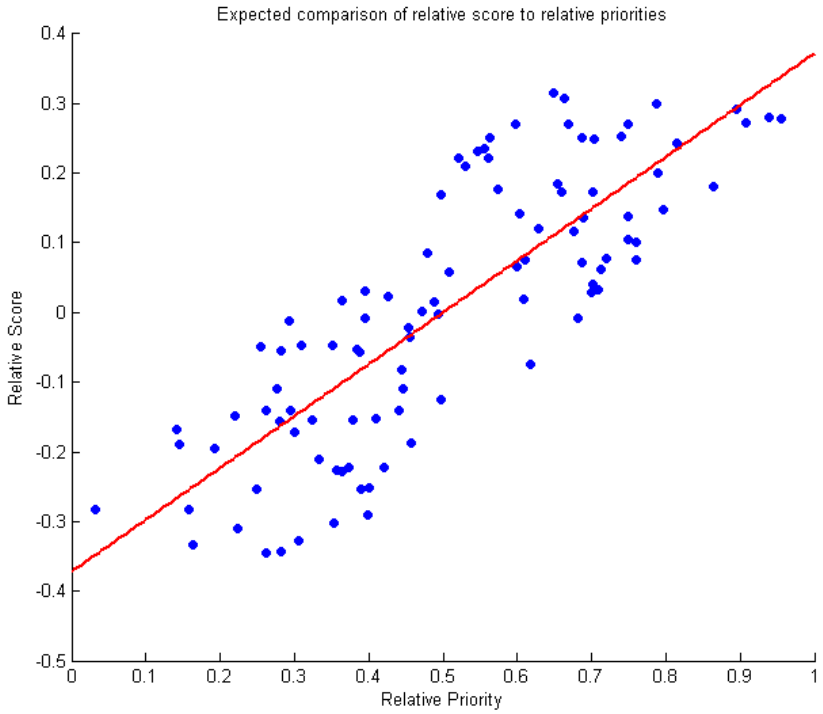

Figure 13 Example of the expected comparison of relative score to relative priorities

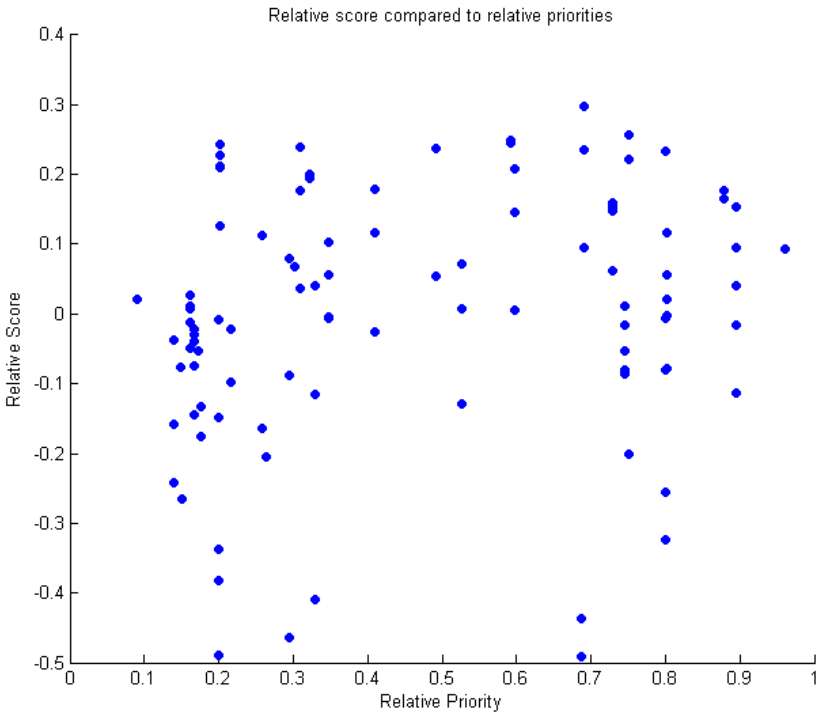

Figure 14 Relative score compared to relative priorities

11 DISCUSSION

This chapter concludes the thesis. First conclusions will be drawn based on the previous chapters. The answers to the research questions are provided and the implications of these results are discussed. Secondly, recommendations are given for the organization of Regiotaxi and for future research.

11.1 CONCLUSIONS

11.1.1 BENCHMARKS

The first research question was:

What are benchmarks for relevant performance indicators and how are these indicators related?

The relevant performance indicators were discussed in paragraph 9.1. Since the evaluation was performed with existing monitoring data and there was no possibility to gather additional data, the objectives that were used in the evaluation were in part determined by the data available to the study and the indicators were matched to the objectives. Besides the objectives from Figure 3, there are some other objectives that would have been interesting. First of all, almost any performance objective could be interesting as perception objective. Only travel time was available for both perception and performance, but the perception of punctuality, availability and tariff would also have been interesting to add. The lists of Pagano & McKnight (1983) and Knutsson (1999) in appendix A contain possible suggestions for additional objectives that can be used, or different groupings of indicators. In the interviews it was mentioned that information provision would be an interesting objective to be included. The most suitable indicator for this objective would probably be the satisfaction about information provision, as it is hard to operationalize the measuring of actual information provision.

The relation between the different indicator values and scores can be found in Table 27. Some of these relations stand out and require closer inspection. In the higher hierarchy of the evaluation framework, economic durability has a very weak correlation with both perception and performance. This effectively means that a variation in the amount of subsidy that is provided for a system, only very weakly relates to how the systems perform and how users perceive they perform. In the light of the budget cuts that are being made in the Netherlands right now, this is an interesting conclusion. It suggests that there are possibilities to reduce subsidy, without necessarily harming the system as a whole.

The second interesting relation is the weak correlation between travel time performance and travel time perception. It is not directly clear where this lack of correlation originates. It can be hypothesized that other time factors influence the travel time perception, for example if the pickup or drop off is not punctual, it is imaginable that this is projected on the travel time perception. However, the correlation between punctuality and travel time satisfaction is 0. The data give no satisfactory explanation. It is possible that it is difficult for people to correctly estimate how long a trip should take and how long it actually took.

The third relation which jumps out is the negative correlation between tariff and perception. This means that systems with higher tariffs score also scores higher on perception. This suggests that tariff has little influence on the perception travelers have of the system and that they are willing to pay the current prices for the service they are receiving. This is again good news for transport authorities, who are looking to cut funding for Regiotaxi, as these results suggest that at least for some users, there is room to increase prices.

Fourthly, perceived safety correlates strongly with perceived comfort. This relation could work both ways, as its imaginable travelers feel safer when they are more comfortable or feel more comfortable when they feel safer. In the interviews it was pointed out that people have a poor estimation of safety. The relation between these objectives therefore raises the question what asking how satisfied travelers are with safety in a customer

satisfaction survey exactly measures. It is possible that when asking about safety in a survey, it is actually comfort that is reported. This should be taken into account when designing surveys.

Another interesting correlation is between average trip distance and trips per inhabitant. The fewer trips that are made the longer the trips are. This is a curious phenomenon that could have different causes. A possibility could be that the most basic trips are longer trips and that the more freely the service becomes available to travelers, the more short trips are made. The service could for example be more freely available when it is cheaper or reservation times become shorter. The correlation could also work the other way, i.e. because Wmo-travelers have a kilometer budget, they can make more trips if they generally make shorter trips, provided that they want maximal use of the budget without exceeding it.

Finally, of the separate satisfaction indicators, travel time satisfaction correlates by some margin the strongest with perception. So in order to improve the overall satisfaction, it is prudent to focus on travel time satisfaction. However, as mentioned, travel time satisfaction correlates only weakly with travel time performance, so a better understanding is definitely needed into factors that influence travel time satisfaction.

11.1.2 EVALUATION

After looking at the benchmarks, the systems were evaluated using the evaluation framework from Figure 3 in order to answer the question:

How are different Regiotaxi systems performing overall and how do they perform relatively to each other?

In the setup of this study, an evaluation framework was developed, that was based on the work of Andrade (2008). One of the considerations in the development was that the framework should not just rank the Regiotaxi services, but also be able to judge them on their merits, allowing single services to be evaluated with the framework. To accomplish this, TOPSIS was chosen as a multi-criteria approach to generate an overall performance score. Since TOPSIS uses the distance from a hypothetical ideal and anti-ideal situation to give a score to a system, instead of a relative score, a single system can be evaluated. However, one condition for this is that the ideal and anti-ideal scores are determined independently from the systems' actual performance. The intention when choosing TOPSIS as a multi-criteria approach was to ask experts for their estimations of the ideal and anti-ideal situations. However, during the interviews, it was advised to take a different approach, since the experts felt it was not possible for them to do this, nor could they think of potential candidates who would be able to help. In the end it was therefore decided to use the data that were available to set ideal and anti-ideal values. This voided the advantage of the method being able to evaluate systems independently, but it did not introduce any negative effects. The use of the data, that were obtained to set ideal and anti-ideal values, resulted in a relative ranking of the systems. However, for future evaluation of Regiotaxi systems this approach may be very useful. It should be noted that the boundaries do have a large effect on the final scores as was shown in the sensitivity analysis (paragraph 9.5) and therefore the suitability of the boundaries should be checked before applying them. For some indicators, especially for indicators that have very little variation in the values, it might be wiser to look for a different method of setting boundaries.

The evaluation framework was set up in such a way that a score of '1' would correspond to the ideal situation and a score of '0' would correspond to the anti-ideal situation. Because of the way the boundaries were set, the overall scores do not approach either boundary. To get a score of '1' would indicate that the performance of the service is among the 2.5% best on all indicators. Because of these limitations, it is hard to say how the overall performance is, since only the relative overall performance is measured. What can be said is that the lowest overall score of 0.4277 means that none of the systems are scoring poorly across the board. The fact that all systems hover around 0.5 means that what they do worse in one area compared to the other services, they compensate for in a different area. That also means that all services have room for improvement and effort should be put in sharing knowledge and trying to learn from each other.

The perception of the system is good: of the indicators which were used, almost every system scores at least a '7' and in a lot of cases an '8' on a scale of 10 for all indicators except for processing complaints. This is also in line with the results from the interviews. The interviewees all saw problems with Regiotaxi and areas for potential improvement. However, overall they all felt that Regiotaxi was already doing a good job providing service. Processing complaints was mentioned in one of the interviews as a problem area. The low score across the board seems to support that. It should be noted though, that the satisfaction with processing a complaint, is only surveyed among users who have filed a complaint. These users will already have a negative attitude towards the service, which could explain the lower scores on this indicator. It is not known to which extent this is contributing to the lower score and to what extent the way complaints are processed is a legitimate concern for the transport authorities.

11.1.3 THE INFLUENCE OF POLICY

The third research question that was brought up in this study was:

To what extent are policy objectives of the transport authorities reflected in the performance of Regiotaxi?

This question received a concerning answer in this study, namely that the influence of policy on achieving policy objectives is negligible in most cases. Only Overijssel shows some signs that their policy is leading to a better performance in the areas they give priority to. Two possible explanations are: (1) a poor translation from policy objectives to requirements for the transport operator or (2) a lack of monitoring of the requirements for the transport operator.

When a tender is designed, the transport authorities have to consider what they want to achieve by awarding the tender. In the interviews the tendering process was discussed and it was mentioned by several experts that they felt that the transport authorities put serious effort in setting up the requirements for the tender and getting straight what is important for their Regiotaxi service. However, it was also remarked that this is a difficult process and that it is difficult to fine tune the requirements in such a way that they accomplish the objectives the transport authorities set for the service. It is therefore plausible that the lack of relation between the evaluation scores and the priorities set by the transport authorities stems from a poor translation of policy objective to policy implementation.

In the interviews it was also pointed out that the effort put into monitoring and evaluating after the tendering process is often not as good as the effort put into setting up and awarding the tender. This could lead to a situation where the transport authority sets the right requirements to obtain its policy objectives. However, the execution does not always conform to the requirements and no action is taken because this is not caught in the monitoring of the system.

11.2 RECOMMENDATIONS

The process and outcomes of the study lead to several recommendations, both for the practical organization of Regiotaxi as for future research.

11.2.1 RECOMMENDATIONS FOR PRACTICE

In the previous paragraph it was mentioned that no system scored good or bad across the board. Generally they had some indicators that performed good and some that performed poorly. This indicates that there's room for transport authorities to improve. This could be achieved by sharing knowledge and best practices. Some of the transport authorities that were interviewed indicated that they already collaborated with other transport authorities, but others indicated that one of the reasons they were interested in this study and willing to participate was a lack of knowledge about how other regions designed their Regiotaxi service and

how they perform. This study provides an example of the type of data that could be shared among regions and offers a framework for interregional evaluation of services and the sharing of knowledge.

Related to the sharing of knowledge is the gathering of monitoring data. This study originated from a request of Goudappel Coffeng to take a look at different customer satisfaction surveys for Regiotaxi they had available. As discussed in paragraph 7.2, these surveys lacked the uniformity for a direct comparison. Furthermore, the trip data provided by the transport authorities also varied greatly in the indicators they used for monitoring. Where the indicators were the same, they were sometimes gathered in a different unit. One of the largest challenges was to make the data of the different transport authorities comparable. In order to share knowledge and compare what the different regions are doing and how they are performing, it would be very useful if the data were more uniform. A guideline similar to what MIPOV (TransTec adviseurs BV, 2008) is for conventional public transport might be useful to accomplish this. As with MIPOV, it is recommended that KpVV develops such a guideline with the transport authorities. Furthermore, it is recommended to the transport authorities to actively share data and experiences on a planned and regular basis. This could be accomplished by creating a national platform for Regiotaxi. In time, this may increase efficiency for all Regiotaxi services.

The different transport authorities used two types of surveys: telephone surveys and paper surveys. For the purpose of comparing the different systems and performing a comprehensive evaluation of the systems, the telephone survey did not give complete enough data. This type of survey allows for fewer responses, as calling travelers is more time-intensive, and for the survey itself, fewer questions can be posed. The telephone surveys are useful for detecting emerging problems in a timely fashion. It is therefore advised that for evaluation purposes a paper survey is used and that call surveys are reserved for keeping a finger on the pulse of the system and not for a full-fledged yearly evaluation.

One of the conclusions is that policy execution does not seem to lead to the achievement of the policy objectives. It is therefore recommended that transport authorities take another look at how their tendering process and system monitoring is being done. Specifically the following questions should be asked:

- How are policy objectives determined?
- How are system requirements developed for the tender from the policy objectives? Does the fulfillment of these requirements also lead to the desired result?
- What data are gathered to monitor the system? Do these data give a complete picture of the performance of the system? What is done when the performance does not meet the standards that were set?

For the tender in a broader perspective, a remark must be made about the budget cuts many transport authorities are experiencing. In the interviews, concerns about different types of fall out were mentioned. When considering budget cuts, the costs induced in other systems must also be taken in to consideration: such as a strain on the timetables and costs of the conventional public transport system from Regiotaxi users switching modes. Another externality is the reduced mobility and increased social exclusion of people in the grey-area of poor mobility who do not qualify for a Wmo indication. Cuts also have an effect on the potential to combine trips. Therefore, when considering budget cuts, the transport authority should cast a wide net on the effects it has, to make sure budget cuts do not lead to a negative net outcome.

A different solution to the budget cuts may lie in reorganizing DRT in the Netherlands. It seems that currently the organization of DRT systems is determined by the funding of the systems. It is possible that by reassigning DRT functions like school transport, work transport and transportation for disabled to new systems, savings can be made by organizing the new systems based on the technical requirements instead of on the source of funding. Since this reshuffling of DRT systems requires laws to be changed, ideally, the National Government

should research whether the efficiency gain from the DRT systems outweights the increased bureaucracy of redirecting funding streams.

Pagano & McKnight (1983) advised to divide an evaluation into user groups for a satisfaction evaluation. It would be interesting to split the user groups of Wmo-travelers and public transport travelers and also to split different age groups (for example 65-, 65-80 and 80+). In this study that advice was not applied, as it is better suited for satisfaction evaluation than the overall evaluation in this study. However, it would also not have been possible since not all the customer satisfaction surveys have data about the user groups. It is therefore very much recommended for future customer satisfaction survey designs to include more information about the traveler. This will also allow for more focused policy implementation.

The evaluation model that was used in this study can be applied to other Regiotaxi systems in the Netherlands, although some remarks about the application have to be taken into consideration. The objectives that were chosen for the model were limited by the data that were available. If data would be gathered – if a national evaluation of Regiotaxi services would ever be done – the objectives and indicators would have to be expanded to give a more complete overview of the performance of the system.

11.2.2 RECOMMENDATIONS FOR FUTURE RESEARCH

As for future research, it would be recommended to delve further into the performance of Regiotaxi and DRT. There's already some research on the planning of DRT systems, the function of Regiotaxi in the total public transport system and best practices. These best practices are often based on experiences without a deeper understanding of the underlying causes. This study found some interesting correlations, but was not able to identify causes. For example, travel time satisfaction has a weak relation to actual travel time performance. It would be interesting to find out why this is the case, as they obviously should be closely related. Also, relatively more trips lead to a shorter average trip. Possibly this is linked to the travel budget of travelers, so if they have a small budget, they save it for longer trips and if they have a larger budget they start using it for shorter trips for which they may also have alternatives available. Another explanation could be that if the Wmo users are trying to maximize the use of their travel budget, they simply can make more trips if they make shorter trips on average. It would be interesting to find out if such relations indeed exist, as it would for example aid the setting of travel budgets for Wmo-indicated.

When researching how to compare the customer satisfaction scores from different scales, surprisingly little literature could be found on the subject. Literature on survey scales mostly focuses on using the correct scale in survey design. No valid rescaling methods were found. In this study several rescaling methods were compared and a rescaling method was applied that used scale profiles, based on the answers given to different questions. This method was possible in this study because a large amount of data were available. Sadly though, the scale profiles cannot be used for different studies. For future research, it would be useful to develop a valid rescaling method that can be widely applied, and that takes the different reactions of people to different scales into account.

In this study it was chosen to evaluate the service from the perspective of the transport authority, because policy was a topic of interest. This means that weights were matched to policy priorities in order to generate a score. This could also be done for the perspectives of travelers and transport companies in order to evaluate the performance of the system from their perspective. The policy priorities survey could also be filled out by them, in order to discover how they rate the performance of the service.

All in all, Regiotaxi seems to be performing well. In the future, improvements have to be sought in a cross-regional collaboration and exchange of information. The scope of transport authorities has to be extended beyond their own systems.

12 BIBLIOGRAPHY

Andrade, M. (2008). *A multi-criteria approach for the design and evaluation of Demand Responsive Transport (DRT) Services.* Porto: Escola de Gestão do Porto.

ANIMATE. (1999). *CORDIS Archive: CORDIS: TAP for Transport: Samplus.* Opgeroepen op November 26, 2013, van Transport sector of the Telematics Applications Programme: http://cordis.europa.eu/telematics/tap_transport/research/projects/samplus.html

ANIMATE. (1997). *CORDIS Archive: CORDIS: TAP for Transport: Sampo.* Opgeroepen op November 26, 2013, van Transport sector of the Telematics Applications Programme: http://cordis.europa.eu/telematics/tap_transport/research/projects/sampo.html

Bergkvist, L., & Rossiter, J. (2007). The predictive validity of multiple-item versus single-item measures of the same constructs. *Journal of Marketing Research* , 175-184.

Brake, J., Mulley, C., Nelson, J., & Wright, S. (2007). Key lessons learned from recent experience with Flexible Transport Services. *Transport Policy* , 458-466.

Brake, J., Nelson, J., & Wright, S. (2004). Demand responsive transport: towards the emergence of a new market segment. *Journal of transport geography* , 323-337.

Cervero, R. (1984). Examining the performance impacts of transit operating subsidies. *Journal of Transportation Engineering , 110* (5), 467-480.

Crainic, T., Errico, F., Malucelli, F., & Nonato, M. (2008). *A proposal for the evaluation of Demand-Adaptive Transit Systems.* Quebec: CIRRELT.

Crosby, P. (1979). *Quality is free: The art of making quality certain.* New York: McGraw Hill Custom Publishing.

Dawes, J. (2002). Five Point vs. Eleven Point Scales: Does It Make A Difference To Data Characteristics? *Australian Journal of Market Research , 10* (1).

Derek Halden Consultancy. (2006). *How to plan and run flexible and demand responsive transport.* UK: Scottish Executive.

Diana, M., Quadrifoglio, L., & Pronello, C. (2007). Emissions of demand responsive services as an alternative to conventional transit services. *Transportation Research Part D* , 183-188.

Enoch, M., Ison, S., Laws, R., & Zhang, L. (2006). *Evaluation Study of Demand Responsive Transport Services in Wiltshire.* Loughborough: Loughborough University.

Enoch, M., Potter, S., Parkhurst, G., & Smith, M. (2004). *INTERMODE: Innovations in demand responsive transport.* UK: Department for Transport and Greater Manchester Passenger Transport Executive.

Falcocchio, J. (1979). *A methodology for evaluating the effectiveness of transportation improvements for the elderly and handicapped.* Washington, D.C.: Department of Transportation.

Ferreira, L., Charles, P., & Tether, C. (2007). Evaluating Flexible Transport Solutions. *Transport Planning and Technology , 30* (2-3), 249-269.

Grönroos, C. (1990). *Service management and marketing: Managing the moments of truth in service competition.* Lexington, MA: Lexington Books.

Grosso, S., Higgins, J., Mageean, J., & Nelson, J. (2002). *Demand Responsive Transport: Towards best practice in rural applications.* Association for European Transport.

Karlaftis, M., & McCarthy, P. (1997). Subsidy and public transit performance: A factor analytic approach. *Transportation , 24* (3), 253-270.

Kennedy, R., Riquier, C., & Sharp, B. (1996). Practical applications of correspondence analysis to categorical data in market research. *Journal of Targeting, Measurement and Analysis for Marketing , 5*, 56-70.

Kennisplatform Verkeer en Vervoer. (2009). *Handboek: Professioneel aanbesteden Regiotaxi, WMO-vervoer en Valys.* Schiedam: TDS Printmaildata.

Kennisplatform Verkeer en Vervoer. (2013). *OV-klantenbarometer 2012.* Ede: Kennisplatform Verkeer en Vervoer.

Kennisplatform Verkeer en Vervoer. (2013). Wmo-vervoerregelingen en Regiotaxicontracten per 1-1-2013. Ede: Kennisplatform Verkeer en Vervoer.

Knutsson, S. (1999). *Valuing rider quality attributes in the Swedish special transport services.* Stockholm: Royal institute of Technology.

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review , 2*, 81-97.

Ministerie van Volkshuisvesting, Ruimtelijke Ordening en Milieubeheer. (2004). *Begrenzing Bebouwd Gebied 2000.* Den Haag: Ministerie van Volkshuisvesting, Ruimtelijke Ordening en Milieubeheer.

Nelson, J., & Phonphitakchai, T. (2012). An evaluation of the user characteristics of an open access DRT service. *Research in Transportation Economics* , 54-65.

Nelson, J., Wright, S., Masson, B., Ambrosino, G., & Naniopoulos, A. (2010). Recent developments in Flexible Transport Services. *Research in Transportation Economics* , 243-248.

Pagano, A., & McKnight, C. (1983). Quality of service in special service paratransit: The users' perspective. *Transportation Research Record, 934* , 14-23.

Paquette, J., Cordeau, J., & Laporte, G. (2009). Quality of service in dial-a-ride operations. *Computers & Industrial Engineering , 56* (4), 1721-1734.

Preston, C., & Colman, A. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica , 104*, 1-15.

Round, A., & Cervero, R. (1996). *Future Ride: Adapting New Technologies to Paratransit in the United States.* Berkeley: The University of California Transportation Center.

Saaty, T. (2003). Decision-making with the AHP: Why is the principal eigenvector necessary. *Journal of Operation Research , 145*, 85-91.

SEO economisch onderzoek. (2011). *Contractvervoer op de rit: Consequenties van marktwerking in het contractvervoer.* Amsterdam: SEO.

Slevitch, L., & Oh, H. (2010). Asymmetric relationship between attribute performance and customer satisfaction: A new perspective. *Internation Journal of Hospitality Management* , 559-569.

Transit Cooperative Research Program. (2008). *Guidebook for Measuring, Assessing, and Improving Performance of Demand-Response Transportation.* Washington, D.C.: Transportation Research Board.

Transit Cooperative Research Program. (2009). *Guidebook for Rural Demand-Response Transportation: Measuring, Assessing, and Improving Perfromance.* Washington, D.C.: Transportation Research Board.

TransTec adviseurs BV. (2006). *Benchmark ov Vlaanderen.* Amsterdam: TransTec adviseurs BV.

TransTec adviseurs BV. (2008). *Model Informatieprofiel Openbaar Vervoer (MIPOV) 2008.* Rotterdam: Kennisplatform Verkeer en Vervoer.

13 APPENDIX A DIMENSIONS AND ATTRIBUTES FOR SURVEYING SATISFACTION IN DRT SERVICES

Table 32 List of dimensions and attributes for surveying satisfaction by Pagano & McKnight (1983)

Dimensions	Attributes
Reliability	**1. Notification of delays or cancellation of service**
	2. Wait time for pickup at home
	3. Wait time for pickup away from home
	4. Arriving at destination on time
	5. Few delays while on the vehicle
Comfort	**6. Guaranteed seat or location for wheelchair**
	7. Condition and cleanliness of the vehicle
	8. Smoothness of the ride
	9. Air conditioning and good ventilation
	10. Sheltered waiting areas for pickups away from home
	11. Seats at waiting areas for pickups away from home
Convenience of making reservations	12. Accommodation to changes in reservations
	13. Being picked up at time selected by traveler
	14. Shortness of reservation time
	15. Convenience of return reservation procedure
Extent of service	**16. Total number of hours of service**
	17. No or few restrictions on where vehicle will go
	18. Service on evenings
	19. Service on weekends
	20. Low rate of turning down reservations
Vehicle access	21. Width of aisle
	22. Height of first step
	23. Number of steps
	24. Presence of wheelchair lift or ramp
	25. Assistance in getting from vehicle to destination
	26. Assistance in carrying packages
	27. Short distance from house or destination to vehicle
Safety	28. Low probability of personal assault
	29. Low probability of falling
	30. Type of tie down
	31. Position of the wheelchair in the vehicle
	32. Low probability of a traffic accident
Driver characteristics	33. Ability to handle medical emergencies
	34. Courtesy and friendliness
	35. Knowledge of general needs
	36. Familiarity with habits and needs of individual user
	37. Neatness and professionalism
Responsiveness	**38. Courtesy and friendliness of telephone operators**
	39. Ease of getting clear information on service
	40. Receptiveness to complaints and user suggestions
	41. Procedure for follow-up on complaints

Table 33 List of dimensions and attributes for surveying satisfaction by Knutsson (1999)

Dimensions	Attributes
Information	1. Information access
	2. Understandable information
	3. Faultless and complete information
	4. Unambiguous information
Dignity	5. Being taken seriously as a traveler
	6. Confidence with respect to what to do and where to go
	7. Personal privacy
	8. Reliability of service
	9. Safety day and night time
	10. Medical emergency capability
	11. Suitable and motivated driver
	12. Courtesy and friendliness
	13. Familiarity with personal needs
Comfort	14. Service on weekdays
	15. Service on weekend
	16. Punctuality, departure
	17. Punctuality, arrival
	18. Freedom of crowding
	19. Booking
	20. Follow-up to complaints
	21. Few restrictions
	22. Prebooking of return
	23. Smoothness of ride
	24. Vehicle inside design
	25. Number of steps
	26. Space and seating
	27. Lift (or ramp)
	28. Distance to vehicle
	29. Driver assistance
	30. Ease of complaining
	31. Possibility to choose departure time
Travel Time	**32. Reasonable in-vehicle time**
	33. Waiting time away from home
	34. Waiting time in the telephone switchboard
	35. Total trips time
	36. Delays on vehicle
	37. Prebooking time
	38. **Punctuality, pickup time**
Fare	39. Worth its price compared to public transport
	40. Fare

14 APPENDIX B: SCHEME FOR DEFINING BUILT-UP AREA

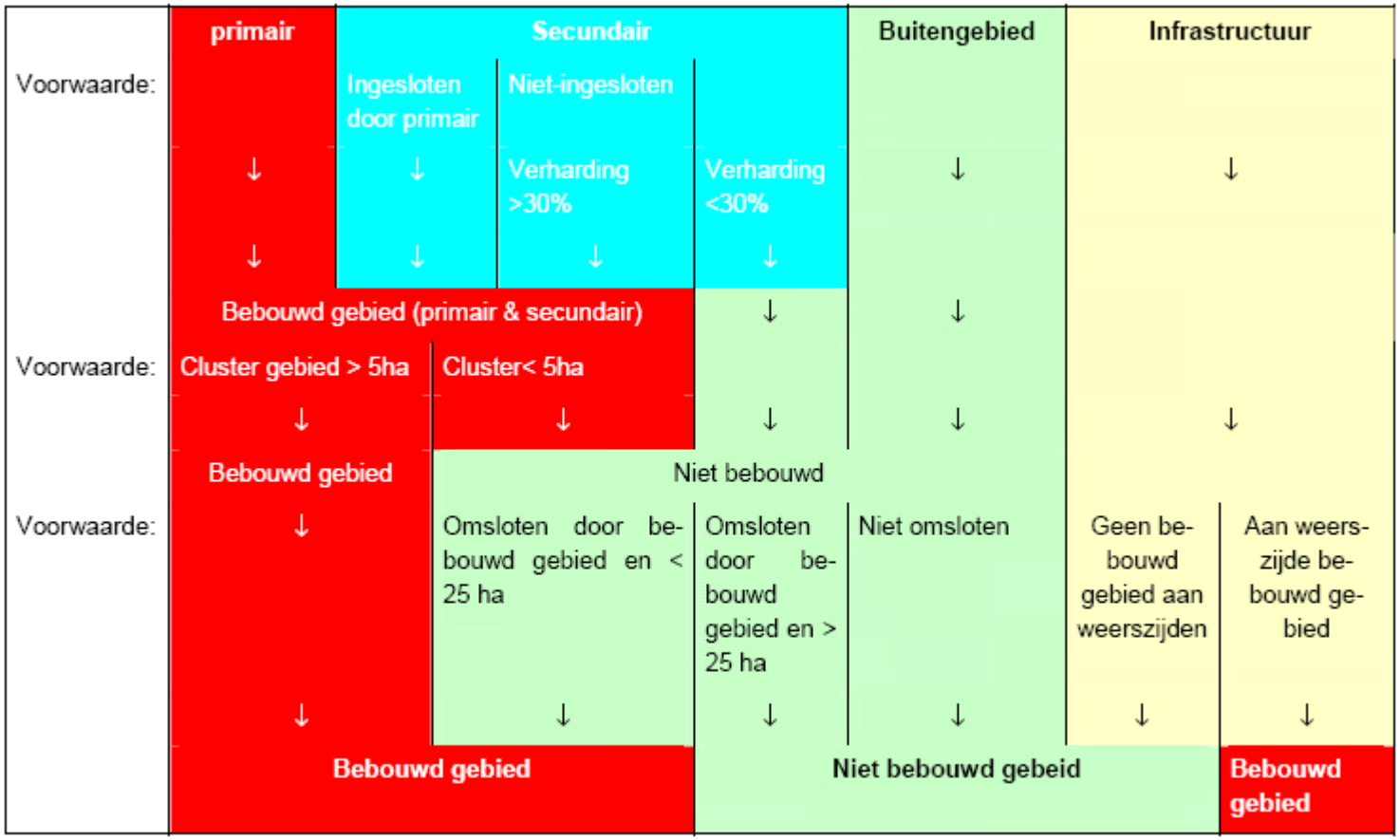

Figure 15 Scheme for determining whether an area is considered built up area, used in the policy processes for Nota Ruimte 2004, Intenvsiveringsbudget Stedelijke Vernieuwing 2005-2009 and Bebouwd Gebied voor monitoringsdoeleinden (Ministerie van Volkshuisvesting, Ruimtelijke Ordening en Milieubeheer, 2004)

15 APPENDIX C: QUESTIONS FROM CSS FOR EVALUATION

Scale	
4 5	Zeer ontevreden - Zeer Tevreden
10	Grade for performance on a scale of 10
4-fq	Altijd - Meestal - Soms - Nooit
5-perf	Uitstekend - Goed - Voldoende - Onvoldoende - Slecht
x	not available

Objective	Indicator	Region	Scale	Question
Comfort	*Comfort CSS*	Gelderland	x	
		Overijssel	x	
		West-Brabant	4	Bent u tevreden of ontevreden over het voertuig als het gaat om: comfort?
		Eemland-H	10	Hoe beoordeelt u: Comfort van het voertuig?
		Veenweide	10	Hoe beoordeelt u: Comfort van het voertuig?
	Cleanliness	Gelderland	10	Hoe beoordeelt u: De mate waarin het voertuig schoon is?
		Overijssel	5	In welke mate bent u tevreden, dan wel ontevreden over: Reinheid / Schoonheid van het materieel?
		West-Brabant	4	Bent u tevreden of ontevreden over het voertuig als het gaat om: netheid?
		Eemland-H	10	Hoe beoordeelt u: Reinheid van het voertuig?
		Veenweide	10	Hoe beoordeelt u: Reinheid van het voertuig?
Safety	*Driving style*	Gelderland	10	Hoe tevreden bent u over de rijvaardigheid van de chauffeur?
		Overijssel	5	In welke mate bent u tevreden, dan wel ontevreden over: Rijstijl van de chauffeur?
		West-Brabant	4	Bent u tevreden of ontevreden over het rijgedrag van de chauffeur?
		Eemland-H	5	Hoe beoordeelt u: Het rijgedrag van de chauffeur?
		Veenweide	5	Hoe beoordeelt u: Het rijgedrag van de chauffeur?
	Safety CSS	Gelderland	10	Hoe tevreden bent u over het gevoel veilig te worden vervoerd?
		Overijssel	x	
		West-Brabant	x	
		Eemland-H	4-fq	Heeft u een veilig gevoel in de taxi?

Objective	Indicator	Region	Scale	Question
			5-perf	Wat is uw algemene indruk van de veiligheid tijden een rit met de Regiotaxi?
		Veenweide	5-perf	Wat is uw algemene indruk van de veiligheid tijden een rit met de Regiotaxi?
Service	*Friendliness*	Gelderland	10	Hoe tevreden bent u over: De vriendelijkheid van de telefoniste?
		Overijssel	5	In welke mate bent u tevreden, dan wel ontevreden over: De vriendelijkheid van de telefoniste?
			5	In welke mate bent u tevreden, dan wel ontevreden over: De klantvriendelijkheid van de chauffeur?
		West-Brabant	4	Bent u tevreden of ontevreden over hoe u door de chauffeur wordt benaderd?
		Eemland-H	10	Hoe beoordeelt u: de klantvriendelijkheid van de telefoniste?
			4-fq	Is de chauffeur vriendelijk?
		Veenweide	10	Hoe beoordeelt u: de klantvriendelijkheid van de chauffeur?
	Helpfulness	Gelderland	10	Hoe tevreden bent u over: De houding en de behulpzaamheid van de chauffeur?
			10	Hoe tevreden bent u over: De hulpvaardigheid van de telefonist bij de ritreservering?
			10	Hoe tevreden bent u over: De hulpvaardigheid van de telefonist bij problemen?
		Overijssel	x	
		West-Brabant	4	Bent u tevreden of ontevreden over de hulp en begeleiding van de chauffeur?
		Eemland-H	10	Hoe beoordeelt u: de behulpzaamheid en begeleiding door de chauffeur?
			4-fq	Is de chauffeur behulpzaam bij het instappen?
		Veenweide	10	Hoe beoordeelt u: de behulpzaamheid en begeleiding door de chauffeur?
	Availability call centre	Gelderland	10	Hoe tevreden bent u over: De wachttijd voordat u de telefonist aan de lijn krijgt (het callcenter dient u binnen 25 seconden te woord te staan)?
		Overijssel	5	In welke mate bent u tevreden, dan wel ontevreden over: Wachttijd bij de telefooncentrale?
		West-Brabant	4	Bent u tevreden of ontevreden over de wachttijd voordat u de telefonist aan de lijn krijgt?
		Eemland-H	10	Hoe beoordeelt u: Telefonische bereikbaarheid/wachttijd reservering (wachttijd hoort korter te zijn dan 2 minuten)?
		Veenweide	10	Hoe beoordeelt u: Telefonische bereikbaarheid/wachttijd reservering (wachttijd hoort korter te zijn dan 2 minuten)?
	Processing complaints	Gelderland	5	Hoe tevreden bent u over de manier waarop uw klacht is afgehandeld?
		Overijssel	x	

Objective	Indicator	Region	Scale	Question
	Processing complaints	West-Brabant	x	
		Eemland-H	4	Bent u tevreden of ontevreden over de wijze waarop uw klacht uiteindelijk is afgehandeld?
		Veenweide	4	Bent u tevreden of ontevreden over de wijze waarop uw klacht uiteindelijk is afgehandeld?
Travel time	*Travel time CSS*	Gelderland	10	Hoe tevreden bent u over: De tijd die u met de Regiotaxi onderweg bent?
		Overijssel	5	In welke mate bent u tevreden, dan wel ontevreden over de gemiddelde reistijd?
		West-Brabant	x	
		Eemland-H	10	Hoe beoordeelt u: De reistijd?
		Veenweide	10	Hoe beoordeelt u: De reistijd?

16 APPENDIX D: METHOD OF DEDUCING SUBSIDY IN EURO/KM FOR ALL REGIOTAXI SYSTEMS

Region		Available	Calculating Subsidy [euro/km]
Overijssel		Total subsidy, total claimed kilometers, kilometers per trip, claimed kilometers per trip	$S = \frac{S_t}{D_{ct}} * \frac{D_c}{D}$
Gelderland		Zones per trip, kilometer per trip, total costs per zone, customer contribution	$S = \frac{D_z * C_z - P}{D_t}$
Utrecht	*Eemland-Heuvelrug*	Total subsidy, total zones, zones per trip, kilometer per trip	$S = \frac{S_t}{D_z} * \frac{D_{zt}}{D_t}$
	Veenweide	Claimable zones, total zones, km per trip, zones per trip, subsidy per claimable zone	$S = S_z * \frac{D_{ct}}{D_z} * \frac{D_z}{D}$
West-Brabant		Total zones, total kilometers, subsidy per zone	$S = S_z * \frac{D_z}{D}$

a

D_c = Average claimable trip distance [km]
D_{ct} = Total claimable distance [km]
D_t = Average trip distance [km]
D = Total trip distance [km]
D_{zt} = Average trip distance [zones]
D_z = Total distance [zone]

S = Subsidy [euro/km]
S_z = Subsidy [euro/zone]
S_t = Total subsidy [euro]
P = Trip price [euro]
C_z = Cost per zone [euro/zone]d

17 APPENDIX E QUESTIONNAIRE FOR THE POLICY PRIORITIES OF THE TRANSPORT AUTHORITIES

Doelstellingen binnen de Regiotaxi

Het invullen van de enquête duurt ongeveer 5 minuten. Bedankt voor uw medewerking met mijn onderzoek.

Doelstellingen binnen de Regiotaxi

In de enquête worden uw prioriteiten gevraagd tussen verschillende doelstellingen binnen de Regiotaxi. U krijgt telkens twee doelstellingen voorgelegd, waarbij u aan kunt geven wat u, in welke mate, belangrijk vindt binnen uw beleid. Eerst worden de verschillende onderwerpen die worden gebruikt gedefinieerd. Dit zijn de definities die binnen het onderzoek worden gebruikt en in sommige gevallen is de definitie smal, bijvoorbeeld veiligheid, wat slechts wordt beschreven door de tevredenheid van de reiziger over de veiligheid (en dus niet door het aantal ongevallen per 100.000 ritten of iets dergelijks). Het is belangrijk dat u bij het invullen van de enquête de onderwerpen beoordeeld volgens de definities zoals ze gegeven worden. De onderstaande definities worden bij de relevante vragen herhaald.

Perceptie	De mate waarin de reiziger het Regiotaxi systeem waardeert
Prestatie	De meetbare objectieve prestatie van de Regiotaxi
Economische duurzaamheid	De hoeveelheid passagierskilometers die gereisd kunnen worden voor een bepaald budget. De mate van onafhankelijkheid van subsidie.
Comfort	De tevredenheid van de reiziger ten aanzien van het comfort van het voertuig
Veiligheid	De tevredenheid van de reiziger over de veiligheid
Reistijdservaring	De tevredenheid van de reiziger over de reistijd
Service	De zorg waarmee de vervoerder de reiziger van dienst is: de tevredenheid van de reiziger over de vriendelijkheid en behulpzaamheid van het personeel en de tevredenheid over de beschikbaarheid van het call centre.
Informatievoorziening	De tevredenheid van de reiziger over de informatie voorziening
Betrouwbaarheid	Het op tijd ophalen en afzetten van de reiziger
Beschikbaarheid	Het aantal uren in de week dat het mogelijk is met de Regiotaxi te reizen
Ritprijs	Het tarief voor de reiziger
Reistijd	De reistijd ten opzichte van de berekende snelste reistijd

Bij het aangeven in welke mate het ene onderwerp prioriteit heeft vergeleken met het andere, varieert de schaal van 'gelijke' prioriteit tot 'extreem hogere' prioriteit voor een van de twee onderwerpen.

Klik op de pijl om naar de vragen te gaan.

Doelstellingen binnen de Regiotaxi

2. **Voor welke OV-autoriteit of Regiotaxi werkt u?**

Provincie Utrecht

--Maak uw keuze--
Provincie Overijssel
Provincie Gelderland
Provincie Utrecht
Regio West-Brabant

Doelstellingen binnen de Regiotaxi

Perceptie	De mate waarin de reiziger het Regiotaxi systeem waardeert
Prestatie	De meetbare objectieve prestatie van de Regiotaxi
Economische duurzaamheid	De hoeveelheid passagierskilometers die gemaakt kunnen worden voor een bepaald budget. De mate van onafhankelijkheid van subsidie.

In welke mate heeft het linker, danwel het rechter onderwerp een hogere prioriteit?

	Extreem hogere		*Hogere*		*Gelijke*		*Hogere*		*Extreem hogere*	
Perceptie	○	○	○	○	○	○	○	○	○	Prestatie
Prestatie	○	○	○	○	○	○	○	○	○	Economische duurzaamheid
Economische duurzaamheid	○	○	○	○	○	○	○	○	○	Perceptie

Doelstellingen binnen de Regiotaxi

Comfort	De tevredenheid van de reiziger ten aanzien van het comfort van het voertuig
Veiligheid	De tevredenheid van de reiziger over de veiligheid
Reistijdervaring	De tevredenheid van de reiziger over de reistijd
Service	De zorg waarmee de vervoerder de reiziger van dienst is: de tevredenheid van de reiziger over de vriendelijkheid en behulpzaamheid van het personeel en de tevredenheid over de beschikbaarheid van het call centre.
Informatievoorziening	De tevredenheid van de reiziger over de informatievoorziening

Perceptie - de volgende onderwerpen zijn doelstellingen om *in de perceptie van de reiziger* een zo goed mogelijke systeem te maken. In welke mate heeft het linker, dan wel het rechter onderwerp prioriteit?

	Extreem Hogere		*Hogere*		*Gelijke*		*Hogere*		*Extreem Hogere*	
Comfort	○	○	○	○	○	○	○	○	○	Veiligheid
Informatie-voorziening	○	○	○	○	○	○	○	○	○	Comfort
Reistijdervaring	○	○	○	○	○	○	○	○	○	Service
Veiligheid	○	○	○	○	○	○	○	○	○	Informatie-voorziening
Comfort	○	○	○	○	○	○	○	○	○	Reistijd-ervaring
Informatie-voorziening	○	○	○	○	○	○	○	○	○	Service
Service	○	○	○	○	○	○	○	○	○	Veiligheid
Veiligheid	○	○	○	○	○	○	○	○	○	Reistijd-ervaring
Reistijdervaring	○	○	○	○	○	○	○	○	○	Informatie-voorziening
Service	○	○	○	○	○	○	○	○	○	Comfort

Doelstellingen binnen de Regiotaxi

Betrouwbaarheid	Het op tijd ophalen en afzetten van de reiziger
Beschikbaarheid	Het aantal uren in de week dat het mogelijk is met de Regiotaxi te reizen
Ritprijs	Het tarief voor de reiziger
Reistijd	De reistijd ten opzichte van de berekende snelste reistijd

Prestatie - de volgende onderwerpen zijn doelstellingen om *voor de reiziger* een zo goed mogelijke systeem te maken, dat wil bijvoorbeeld zeggen dat bij ritprijs de doelstelling is om het vervoer voor een zo laag mogelijke eigen bijdrage van de reiziger aan te bieden. In welke mate heeft het linker, dan wel het rechter onderwerp een hogere prioriteit?

	Extreem hogere		*Hogere*		*Gelijke*		*Hogere*		*Extreem hogere*	
Betrouw-baarheid	○	○	○	○	○	○	○	○	○	Beschik-baarheid
Reistijd	○	○	○	○	○	○	○	○	○	Ritprijs
Beschik-baarheid	○	○	○	○	○	○	○	○	○	Reistijd
Betrouw-baarheid	○	○	○	○	○	○	○	○	○	Ritprijs
Reistijd	○	○	○	○	○	○	○	○	○	Betrouw-baarheid
Ritprijs	○	○	○	○	○	○	○	○	○	Beschik-baarheid

Hartelijk dank, u kunt uw antwoorden insturen door te klikken op de knop voor de volgende pagina.