Benchmarking municipal sport policy

The design and execution of a benchmark for municipal policy-makers to evaluate the portfolio sport facilities.

Master thesis submitted to Delft University of Technology
in partial fulfilment of the requirements for the degree of
MASTER OF SCIENCE
In ***Construction Management and Engineering***
Faculty of Civil Engineering

By

Daan Spijkers
Student number: 1392875

To be defended in public on October 23 2017

Thesis committee

Chair person:	Prof. dr. ir. M.F.W.H.A. Janssen	TU Delft, Section ICT
First supervisor:	Dr. ir. B. van Loenen,	TU Delft, Faculty A+BE
Second supervisor:	Dr.ir. B.M. Steenhuisen,	TU Delft, Multi-Actor systems
Supervisor:	Dr. A.M.G. Zuiderwijk-van Eijk,	TU Delft, Section ICT
Daily supervisor:	mr. drs. A.F. de Clerck,	Republiq

"The only true wisdom is knowing that you know nothing"
Socrates

Abstract

This thesis is focused on benchmarking municipal sport policy. Physical activity is used by the Dutch government as a policy tool to create a healthy society. The Ministry of Health, Welfare and Sport stimulate people to adopt active life styles in order to combat the prosperity diseases: overweight, diabetes and depression. In benchmarking municipal sport policy there are the following complexities. Firstly complexities related to the governance of municipal sport policy, there are many actors involved with contradictory interests. In addition municipal sport policy is focussed within municipal boundaries and evolves on previously pursued policies. Secondly complexities related to benchmarking. Benchmarking is from origin a method developed for the industrial industry and now is also applied in the public sector, but the applicability of the method is contested. Moreover benchmarking is not straightforward and is dependent on its design, data reliability and the way of processing data.

An essential actor in sport policy are the municipalities: they conduct sport policy in order to heave healthy citizens and focus their policy to a large extent on sport facilities. Currently municipalities lack insight in whether the output of the sport policy (namely the quantity, type, and location of sport facilities) leads to the desired effects (namely healthy citizens), which we define as a lack of insight in the effectiveness of the sport policy. This insight is crucial for municipalities, since it can be used in the policy evaluation to assess the performance of their policy.

A method that has been proven useful for measuring and comparing of performance is benchmarking. Benchmarking is the research in which the output between organisations can be compared. However, there is a lack of literature describing how sport policy by municipalities can be compared, whereas comparison with other municipalities can provide additional insights and stimulate learning. In this study a benchmarking method for municipal sport policy aspects is developed and conducted.

The benchmarking method is designed based on a benchmarking literature review. Based on the review it was established that the design of the benchmarking method consists of phases, process steps and criteria. Subsequently with the literature review a benchmarking model is designed consisting of phases, steps, and criteria directed at assessing the performance of municipal sport policy. In the benchmarking criteria common elements are: the usage of indicators, transparency in performance measurement method and self-assessment of performance. In the benchmarking models common process steps are: set objectives, define indicators, select benchmark groups, collecting data, preparing data, analyzing data, determining significant different findings and reporting findings.

With the designed benchmarking model the benchmark is conducted, resulting in the following findings. The benchmarking results found no significant correlation between the dimensions, the output and the outcome of municipal sport policy. This study found that for benchmarking both the benchmarking design as the benchmarking process are crucial. Currently literature focuses on the design of the benchmarking method. This study found that the benchmarking process confronts challenges mainly related to data analytics.

Key words: benchmark, municipal policy, sport facilities, performance measurement, open data, public data

Table of Contents

1. Introduction

This thesis focuses on benchmarking municipal sport policy. Physical activity is used by the government as a policy tool for to realize a healthy society. The Ministry of Health, Welfare and Sport (VWS) stimulate people to adopt active life styles (VWS, 2017). In order to combat the prosperity diseases: overweight, diabetes and depression. In the governmental health policy, physical activity, is used as a tool since it promotes physical fitness, gives social contacts and contribute to self-development (VWS, 2017). However policy-makers lack insight in the effectiveness of their sport policy therefore this study develops and conducts a municipal sport policy benchmark. The main research question that is answered is: How to design a sport policy benchmark that increases the effectiveness of a sport policy in municipalities in the Netherlands?

1.1. Sport policy system

The Dutch government conducts sport policy because it aims to have healthy citizens. Therefore sport is used as a tool. Sport policy is defined as a course of action proposed by the government for the field of sport. The output of sport policy are decisions regarding budgets, subsidies, and permits for sport facilities and sport participation programs. The main instrument within sport policy are sport facilities. The main actors in sport policy directed at realising sport facilities are the municipalities. Each municipality develops their own municipal sport policy in which sport facilities policy takes the most prominent place.

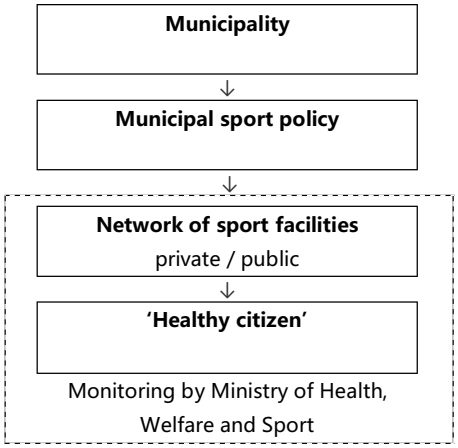


Figure 1 system of sport policy (own illustration)

The system of municipal sport policy is depicted in Figure 1. The municipalities in the Netherlands conduct sport policy because they want healthy citizens (van der Poel et al., 2016, p. 253). Another reason is because sport is used for political ambitions, for identity and pride (van der Poel et al., 2016, p. 253). To reach the goal of healthy citizen the output of municipal sport policy is primarily directed at sport facilities. For the management of the sport facilities municipalities depend on sport clubs and private sport companies. The ministry of Health, Welfare and Sport plays an indirect role in municipal sport policy and tries to make the performance of sport policy transparent with monitoring of the sport facilities and the health of citizens.

Decades of sport policy has resulted in an extensive network of sport facilities, including hockey courts, soccer courts, swimming pools, and indoor halls. And this network of sport facilities is used by citizens for physical activity.

1.2. Actors

The following actors are involved in the municipal sport policy system: The governmental actors are, the ministry of Health, Welfare and Sport and the municipalities. The civil society is represented by the public sport clubs. The business entity by private sport companies and finally the citizen are the users of sport facilities.

Ministry of Health, Welfare and Sport

The role of the ministry of Health, Welfare and Sport in sport policy is decision- and policy maker. Their sport policy activities concern four categories:

- Sport participation programs, examples are sport in the neighbourhood, sport and health, and sport with disabilities.
- Regulations for sport facilities, regarding sustainability, safety and accessibility.
- Elite sport subsidizing, examples are top sport programs, elite sport events, and honest competition programs such as anti-doping and anti-match fixing.
- Subsidizing research in sport performance. This research is both innovation for elite sport as making the performance of sport policy transparent.

For the latter, making the performance of sport policy transparent, the ministry of Health Welfare and Sport commission organisations to monitor a certain aspect of sport policy performance. Four examples of these monitors and organisations are:

- Netherlands Institute for Social Research (SCP) is asked to research sports-related trends and developments.
- Mulier institute is commissioned to gather a national database of all sport facilities in the Netherlands and researches sport policies in all aspects.
- Municipal Health Services (GGD) are commissioned to monitor the health of the citizens in the Health-monitor.
- CBS is tasked with collecting, processing and publishing demographic data.

The ministry of Health, Welfare and Sport develop sport policy and expect that this results in a healthy society and depend for this on two actors. It depends on municipalities for conducting sport policy on municipal level, and on sport providers to accommodate sport programs.

Municipalities

The role of the municipalities in sport policy is one of policy maker. Their main interest is to ensure sport policy implementation effectively, in other words to provide an adequate sport service level with the objective to have healthy citizens (van der Poel et al., 2016, p. 253). The municipal sport policy concern two categories sport facilities and sport participation programs. In terms of money, sports facilities are by far the most important instrument used in municipal sport policy (Tiessen-Raaphorst, Verbeek, de Haan, & Breedveld, 2010; van der Poel et al., 2016, p. 247).

To develop municipal sport policy the organisation of a municipality is divided in the administrative- and the civil service and can be seen as a hierarchical structure. The Administrative services consist of city council and the Mayer and alderpersons. The city council, is elected every four years, and their main tasks are representing the citizens, set budget, establishing the outline of the policy in guidelines, and monitoring the implementation of the policy. The Mayer and alderpersons, are responsible for the daily management and the implementation of the policy, by transforming the guidelines in to policy plans. The civil services are supportive of the administrative services and their task is to implement the municipal (sport) policy determined by Mayer and alderpersons. Depending on the municipality sport policy regarding sport facilities fall within a department sport or department real estate (Beemt & Veuger, 2016).

Municipalities conduct sport policy for two objectives, healthy citizens and identity and pride. Nevertheless the effectiveness of sport policy for healthy citizen is difficult to determine. Due to the

following complexities, they lack insight in the effectiveness of sport policy. Firstly within the municipality the effects off sport facilities on citizens is unclear. Additionally it is unclear what criteria to use to measure the effectiveness. Secondly health and sport facilities are managed by different departments because the real estate department is centralised in the majority of the municipalities (Beemt & Veuger, 2016). Thirdly sport policy evolves on previously pursued policies in which the existing sport facilities largely depend the supply of sport facilities making the effectiveness of changes difficult to perceive. These complexities could lead to the following conflicting interests. The focus could be more on the easy to measure political objectives identity and pride. This may result in developing sport policy that connects to the political objectives, for example allocating budget for a state-of-the-art elite sport facility that achieves the political ambitions but not the desired effect of healthy citizens. A second conflict is that municipalities play a decisive role and depend on sport clubs and private sport companies for the implementation of sport policy.

Sport clubs

Within the sport policy system sport clubs' role is one of executing policy. They receive subsidies to enable citizen to practice their favourite sport in an organised form and in a secure sport facility. In addition to the subsidies sport clubs will receive a membership fee from their members in exchange for their services. The sport clubs in the Netherlands are non-profit organisations.

Sport clubs depend on municipal sport policy for permits and subsidies that allow them to rent a sport facility and subsequently to organise sport. If they have issues related to sport policy they have to follow the political structure within the municipality. They have to inform the alderperson sport and stress the urgency of the issue in the hope he or she puts it on the policy agenda.

Private sport companies

In sport policy private sport companies execute the sport policy. They receive a permit to provide a commercial sport facility in a certain place. They have a commercial interest and focus their business model on gaps in the market. Examples of private sport companies are fitness and yoga studios.

In order to conduct their business, private sport companies depend on permits that derive from the municipal sport policy. Their focus is not in accordance as with the municipality. Their focus is on commercial interest, provide sport facility for profit not for healthy citizens. This does not result in a conflict of interest because the private sport companies can conduct their business and by doing so hopefully make a profit. And for the municipality it means a wider spectrum of sport facilities and a higher sport service level possibly leading to healthy citizens.

Citizen

Citizen in the sport policy system use the output of the sport policy, they use the services offered by government and private sport companies. Consequently they have choice in a variety of sport facilities. The conducted sport policy results for citizens in the ability to conduct physical activity in their municipality. Their main interest is easy access to a variety of sport facilities within a certain proximity but depend for this on the municipal sport policy.

The main actor in sport policy directed at sport facilities are the municipalities, with their municipal sport policy they a decisive role in the field of sport. The output of the municipal sport policy has led to an extensive network of sport facilities that are used by citizens for physical activity. This study is directed at

municipal sport policy, the following paragraph will elaborate on municipal sport policy and how it is developed.

1.3. Sport policy cycle

The policy cycle within a municipality can be seen as a policy process that consist of sub-processes which together lead to decision-making about the policy and concrete implementation. The policy cycle is a heuristic model and is originally developed by Lasswell (1951) Subsequently Hoogerwerf and Herweijer (2014) expanded the theory for the municipal context in the Netherlands.
A policy cycle commonly consists of six main phases, namely, policy agenda, policy design, decision-making, policy implementation, policy evaluation and policy feedback (Hoogerwerf & Herweijer, 2014). In the sport policy cycle municipalities conduct sport policy to keep citizens healthy. Below the process is described within the municipal context (Hoogerwerf & Herweijer, 2014).

In order to develop new policy or adjust existing policy the first process step is that societal problems need to be identified, recognised and put on the *policy agenda* by the policy makers. Subsequently the *policy design* can start. This is the process of collecting, analysing, and formulating advice regarding the policy to be implemented. The advice consists of different ways to solve the societal problem, documented in policy alternatives. Based on the policy alternatives a *decision* is made concerning the contents of the policy, specifying the purposes, resources, and timeframe. Accordingly, the chosen policy can be implemented and executed and once it is finished the policy evaluation starts.

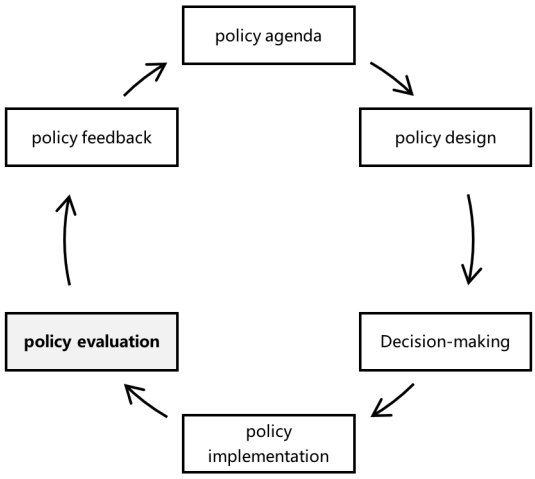


Figure 2 Policy cycle Dutch municipalities (Hoogerwerf & Herweijer, 2014).

The policy evaluation, is the process that tries to assess the effectiveness of a policy, the extent to which the output leads to the envisioned effects, the outcome (Hoogerwerf & Herweijer, 2014). These effects could be both positive as negative. The policy evaluation consists of monitoring and evaluation (Cobussen, Puyt, & Ven, 2015). Monitoring is observing the progress of the implemented policy. Evaluation is assessing the effectiveness of the policy with the information that is obtained from monitoring the policy output and outcome. With the monitoring information, possibly insights in the causes and effects of the policy can be found and used for recommendations for policy adaptations. Finally, this evaluation information needs to be fed back to the policy cycle in order to make adjustments to the policy.

Because limited information is available for municipal policy makers, the evaluation of sport policy is a complex step. Consequently learning from previous decisions is difficult and the possibly insights in the causes and effects of the policy cannot be found and used for recommendations for policy adaptations.

1.4. Problem statement

Decades of municipal sport policy have resulted in a network of sport facilities, i.e. a number of sport facilities to play soccer, tennis, swimming or other sports. Currently there are trends asking for a new position of physical activity in the city: aging of society, population in- and decline, and change in the way we exercise (van der Poel et al., 2016, p. 16). The current Dutch society is expected to grow due to longer life expectancies and immigration (van der Poel et al., 2016, p. 16). In urban areas in the west of the Netherlands, an increase in demand for specific sport facilities is expected of 10 to 20% (van der Poel et al., 2016, p. 16). Contrary in deprived areas a decrease in demand with the same percentages is expected (van der Poel et al., 2016, p. 16). Moreover the manner of exercising is changing. This is illustrated by an increase in the use of public space for physical activity and a decrease in the use of swimming pools (van der Poel et al., 2016, p. 16). Due to the aforementioned trends sport policy and its effectiveness are being reconsidered and debated.

1.4.1. Practical problem: Lack of insight in the effectiveness of sport policy

The evaluation of sport policy is complex. Municipalities lack insight in the effectiveness of sport policy (van der Poel et al., 2016, p. 253). They do not know whether their budget for sport policy is spent effectively. Currently municipalities use methods that focus on real estate criteria to assess the performance of sport facilities, examples of these criteria are: rentable surface, energy consumption, depreciation, and maintenance costs. These criteria focus on the performance of the real estate while the lack of insight is in whether the output of the sport policy (namely the quantity, type, and location of sport facilities) leads to the desired effects (namely healthy citizens) (van der Poel et al., 2016), which we define as a lack of insight in the effectiveness of the sport policy.

Insight in the effectiveness of a sport policy is crucial for municipalities, since it can be used in the policy evaluation step to assess the performance of the policy. It gives municipalities insight in whether they spend their sport budget in the way that leads to the desired effects. To increase the impact of the policy evaluation comparison with other municipalities can provide additional insights and stimulate learning (de Bruijn, 2007; Neely, Gregory, & Platts, 1995). It can help organizations understand their own position and identify growth opportunities (Maheshwari, 2013). However, this is not a straight forward process. In addition, limited information is available making the evaluation a complex step.

1.4.2. Theoretical problem: Benchmarking design and lack of empirical base

A method that has been proven useful for measuring and comparing of performance is benchmarking. Benchmarking is the research in which the output between organisations can be compared. It uses monitoring and evaluation information to compare results with the aim of creating transparency and enable learning. For now benchmarking is defined as measuring elements and comparing the outcomes with comparable organisations.

Benchmarking from origin is a method developed for the manufacturing industry and currently it is also applied in the public sector, although the applicability of the method in the public sector is contested. Benchmarking can be an effective method in the public sector, but there is also strong criticism from de Bruijn (2002) and Bannister (2007). According to de Bruijn (2002) the design can lead to perverse effects, namely it can considerably widen the gap between political decision-making and implementation (de Bruijn, 2002, p. 5). Therefore, special attention has to be paid to the design of the

benchmarking study and this has resulted in a plethora of benchmarking methods (Anand & Kodali, 2008) in which each benchmarking method has its strengths and weaknesses and is applicable in specific situations (Susha, Grönlund, & Janssen, 2015, p. 13). Even though a plethora of benchmarking methods exits, benchmarking methods often lack an empirical and theoretical base (Bannister, 2007; Maheshwari, 2013). Hence the problem with benchmarking is the lack of theoretical foundation, this is relevant since it can lead to wrong conclusions and perverse effects.

1.5. Research questions

In the problem statement two problems were identified. Firstly it was stated that municipalities lack insight in the effectiveness of sport policy. This insight is crucial for municipalities, since it can be used in the policy evaluation to assess the performance of their policy. The second problem is related to the solution, and states that there is a lack of literature describing how sport policy by municipalities can be compared, whereas comparison with other municipalities can provide additional insights and stimulate learning. This research takes into account the two aforementioned problems and uses it to develop a benchmarking method to evaluate the effectiveness of sport policy and its limitations. Currently this method does not exist and it is important because years of sport policies has led to a dense network of sport facilities. However over the years the way we exercise has changed and municipal policy makers are reconsidering and debating the effectiveness of their sport policy.

This research develops an approach to design a sport policy benchmark and its limitations. The aforementioned lead to the following research question and sub-questions:

Main research question:

> *How to design a sport policy benchmark that increases the effectiveness of a sport policy in municipalities in the Netherlands?*

Sub questions:

The sub questions are structured in such a way that they answer in a systematic way the main research question.

Research question 1: *What are the process steps and requirements for designing a benchmark?*

The first question explores benchmarking, its challenges, and designs. Since the design of benchmarking can lead to perverse effects and the fundamental cause for this is that benchmarking methods often lack theoretical foundation. A literature study regarding benchmarking is conducted defining benchmarking and its challenges. Subsequently based on the established challenges, specific benchmarking criteria scholars are identified and depicted in an overview of benchmarking criteria. In addition an overview is given from different benchmarking designs and the corresponding process steps.

Research question 2: *How can we design a sport policy benchmark?*

The second research question designs method of the benchmarking design. It uses the identified benchmarking criteria and process steps from research question 1. The criteria and process steps are analysed and common elements are identified. After which the prevailing elements are used for the design of the municipal sport policy benchmark. Resulting in a method for municipal sport policy benchmarking, consisting of 8-process steps and corresponding criteria

Research question 3: *What does the sport policy benchmark look like?*
The third research question builds upon the designed benchmarking process and utilizes it to address the practical problem. To solve the practical problem a benchmarking study of sport policy is developed. This is expected to enable learning and increase transparency. To conduct the benchmarking of municipal sport policy the benchmarking design needs to be applied in the specific context. The identified benchmarking criteria from research question two are elaborated in the context of the municipal sport policy benchmark. This results in the explanation of the benchmarking objectives, indicators, and group. Consequently the benchmarking of sport policy is made ready for the next phase, the application.

Research question 4: *How can the developed sport policy benchmark be used?*
With the benchmarking of municipal sport policy made ready for use, now it can be applied. Research question three focusses on the application of the municipal sport policy benchmark. It explains in a systematic manner the benchmarking process, the series of steps that are performed in order to apply the benchmark.

Research question 5: *What are the effects of the developed municipal sport policy benchmark?*
The aforementioned research questions are aimed at developing, designing, and applying the benchmarking of municipal sport policy. Apart from the beneficial effect of benchmarking it can also lead to perverse effects. Therefore to design a sport policy benchmark that increases the effectiveness of a sport policy in municipalities in the Netherlands it is important to assess whether the possible perverse effects are minimised. Research question five is aimed at evaluating the developed and applied benchmarking method. It evaluates if the design and if the benchmarking study improves the situation. The design of the developed method is evaluated by assessing whether the method leads to perverse effects. Subsequently the question is answered if the conducted benchmarking study improves the current situation, meaning that by conducting the benchmarking study more insight in the effectiveness of the sport policy is obtained.

In this study, we develop and conduct a benchmarking study of sport policy. We focus on municipal policy makers, since they create the sport policy and because domain knowledge is needed to make sense of the benchmarking results. In addition the output of sport policy to a large extent sport facilities, what this exactly is will be discussed in chapter 4.

1.6. Research approach

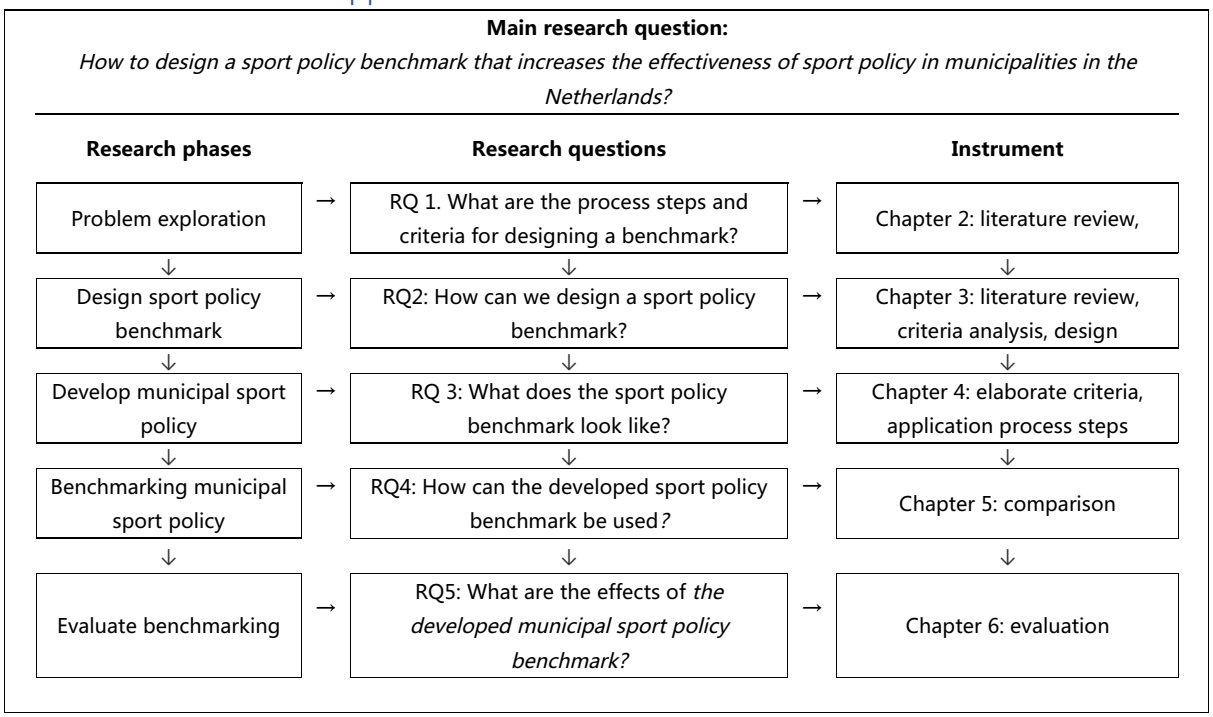


Figure 3 research approach of this study

1.7. Research contributions

This research contributes practically by developing a benchmarking method for municipal policy-makers and subsequently conducting it. Currently a similar benchmark does not exist yet while it is needed to make more informed decision on effective sport policy. By developing and conducting the benchmarking method the effectiveness of municipal sport policy is made transparent. Comparison with other municipalities may provide additional insights and stimulate learning.

This research contributes to the literature by conducting and developing a benchmarking method for municipal sport policy. Currently there is a lack of theoretical methods and literature to guide benchmarking. This research gives an approach to reaching this municipal sport policy benchmark and its limitations.

1.8. Research overview

Figure 4 provides a flow diagram of this thesis and shows the activities.

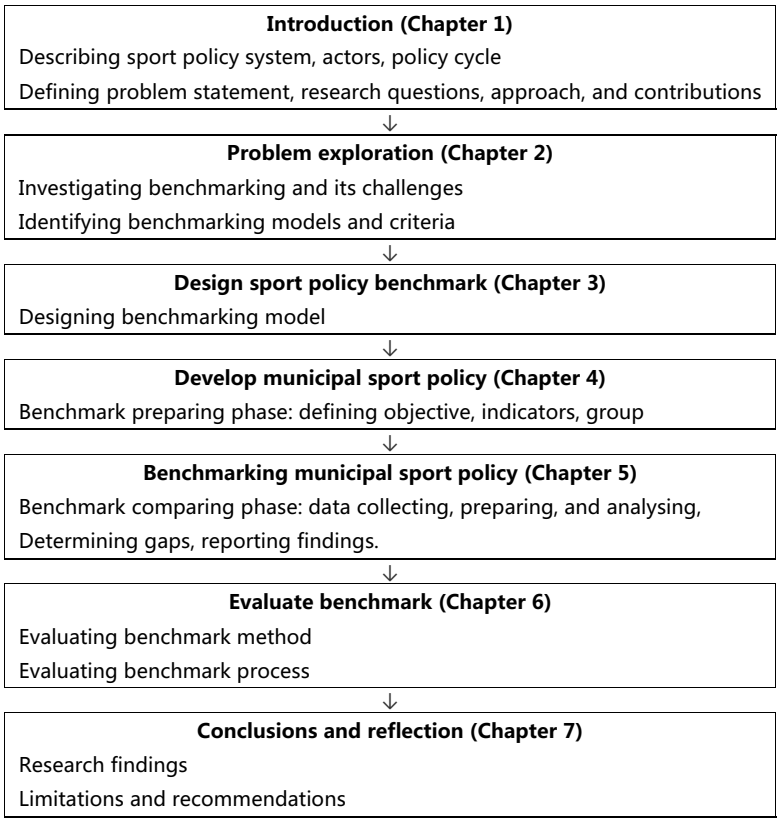


Figure 4 Flow diagram research activities.

2. Research background

As mentioned in the problem statement policy-makers lack insight in the effectiveness of sport policy. A method that has been proven useful for measuring and comparing of performance is benchmarking. However benchmarking methods often lack theoretical and empirical foundation, this can lead to a wrong benchmarking design and perverse effects. This chapter reviews literature regarding benchmarking this will finally lead to an overview of benchmarking criteria and benchmarking models and the corresponding process steps. This overview will eventually form the base for the design of the benchmarking model.

Firstly in paragraph 2.1. the literature review approach is depicted after this the literature review starts. Paragraph 2.2. gives insight in what is benchmarking and elaborates on the reasons for benchmarking, the types of benchmarking and give definitions of benchmarking. Subsequently paragraph 2.3. elaborates on the possible challenging effects that come from benchmarking. Paragraph 2.4. will continue on the challenging effects and describes the criteria scholars have pointed out in order to combat the challenging effects of benchmarking. Subsequently paragraph 2.5 describes different benchmarking models and the corresponding process steps. Finally paragraph 2.6. sums the review part by giving an overview of the identified benchmarking criteria and an overview of the benchmarking models and the corresponding process steps.

2.1. Literature review approach

A methodological review of past literature is a crucial endeavour for any academic research and an effective review creates a firm foundation for advancing knowledge (Webster & Watson, 2002, pp. 48–49). Many scholars have made contributions on how to perform an effective literature review, for this research the theories of Levy and Ellis (2006) and Webster and Watson (2002) are used.

Levy and Ellis (2006) introduce a framework for an effective literature review, they use a systematic data processing approach that is comprised of the following three stages (2006, p. 182):

1. Input (literature collection and scanning),
2. Processing (understand information, apply, analysis, synthesize and evaluate),
3. Output (writing the literature review).

Webster and Watson (2002) defined what an ideal article should consists of. Firstly it should include the focus of the research with a motivation for the research topic and the explanation of the contribution. Secondly is should consist a description of the key concepts, followed by an explanation of the determination of the boundaries of the research. After this the selection criteria and the process to determine the relevance of the articles need to be determined. With the defined focus, concepts, boundaries, selection criteria and the process the literature review can start. When the articles are chosen, the research can start the synthesis part and try to find associations between the studies. Finally, conclusions derived from the literature review can be drawn and presented with the implications for the study.

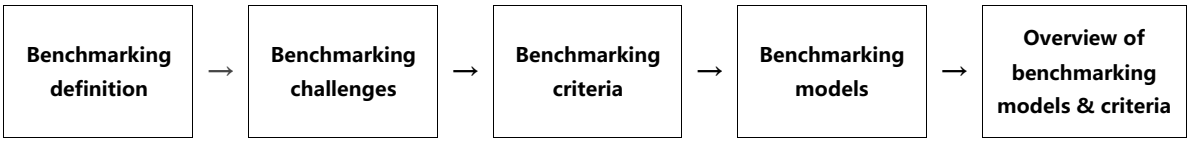
When combining the approach and points of Levy and Ellis (2006) and Webster and Watson (2002) the following model for an effective literature review process can be constructed.

1. Input	2. Processing	3. Output
Focus: identify: benchmarking, benchmarking criteria, benchmarking models. **Key concepts:** benchmarking, performance measurement, data analysis / performance analysis. **Boundaries**: articles found with keyword searching in Sciencedirect, Scopus, JSTOR, ACM, and repository TU Delft and backward– and forward searching.	**Select criteria**: topic relevance, establishment of authors. **Determine applicability:** scanning titles, reading abstracts.	**Analysis:** description of the studies. **Synthesis:** identification of the associations between the studies. **Conclusions:** Draw conclusions from literature review and explain its implications for this study.

Table 1 literature review process based on Levy and Ellis (2006) *and Webster and Watson* (2002).

When the benchmarking literature review is performed, it is concluded with an overview of benchmarking criteria and models. This overview is aimed for the design of the municipal sport policy benchmarking method. The first part of the literature review process is the input phase this is composed of three steps: focus, key concepts and boundaries. The literature review is aimed to answer the following research question: *What are the process steps and criteria for designing a benchmark?* This research question can be decomposed into the following four sub questions:

- What is benchmarking?
- What are benchmarking challenges?
- What benchmarking criteria have scholars formed to combat challenges?
- Wat benchmarking models have scholars developed to combat challenges?



The benchmarking criteria and the common elements in the benchmarking models will in chapter 3 define the conditions for the benchmarking method, that measures the performance of the sport facilities portfolio of municipalities in the Netherlands. Subsequently the key concepts of the research are defined in order to obtain the literature review goals, for this research the following keyword searches used: benchmarking, criteria, performance measurement, data analytics, and data. Finally the last step of the input phase is the demarcation of the boundaries for the search, this is done by confining the databases for the keyword searches. The following databases have been used: Sciencedirect, Scopus, JSTOR, ACM, and repository TU Delft.

In the second part of the literature review process the search results are processed. The way of processing is determined with the selection criteria and the applicability of the literature. The selection criteria for the literature is based on topic relevance and establishment of the authors. The applicability of the articles is determined based on scanning titles and reading abstracts. It is conceivable that due to the novelty of the research topics, the combination of the applicability and selection criteria that not all needed information can be obtained with acknowledged scholars. In such a case the selection criteria will be relieved and also thesis dissertations will be used. The used literature consists of books, journal articles, conference proceedings, promotion and thesis reports.

The third and last part of the literature review processes the output, in other words the literature review is written. According to Levy and Ellis (2006) the output of the literature review should contribute to the overall body of knowledge, this is achieved by building on other research results, and is depicted in chapter 3. The individual studies are analysed, subsequently associations between the different studies are identified and finally conclusions are drawn from the literature review.

For the literature review of relevant benchmarking related topics the following literature and scholars are used. To identify the benchmarking challenges, research from Bannister (2007), de Bruijn (2002), and Maheshwari (2013) are used. To describe the benchmarking criteria, the studies of de Bruijn (2002), Janssen (2010), and Maheshwari (2013) are used. Finally for the benchmarking methods the research of Anand and Kodali (2008), Bhutta and Huq (2008), and Camp (1989) are used. In addition the research from the European Committee for Standardization (2012) is used, they have established a benchmarking standard by scientific research, and consensus.

2.2. What is Benchmarking

The basic idea of benchmarking is the following it is a method for improvement achieved through measuring and comparing the output of organizations. Benchmarking is used in an organizations evaluation phase and uses monitoring and evaluation information to compare results, with the aim of creating transparency and enable learning. Accordingly the following two terms are relevant: measuring and benchmarking. Measuring is assessing the performance of an organization and benchmarking is the activity of comparing the performance scores (Maheshwari, 2015). Therefore the premise of benchmarking is that organizations provide products and services and that their performance – their output – can be measured (de Bruijn, 2007, p. 57). Hence benchmarking is restricted to the output and effects that are measurable (de Bruijn, 2007).

Benchmarking can be grouped under the broader field of performance measurement which is a part of performance analytics. This is defined as the extensive use of data and statistical and quantitative analysis to drive decisions and actions (Davenport & Harris, 2007, p.7). Subsequently performance measurement can be defined as the process of quantifying the efficiency and effectiveness of action (Neely et al., 1995). Hence, benchmarking is performance measurement and requires the resource data. Regardless benchmarking definitions lack the connection with data.

The definition of benchmarking according to the Oxford dictionary is: "The process of comparing against a point of reference". There are plenty of varying definitions for benchmarking. Anand and Kodali (2008) have reviewed these and concluded that the key benchmarking themes are: measurement, comparison, identification of best practices, implementation and improvement (Anand & Kodali, 2008, p. 258). The most commonly cited definition is from Camp (1989) since he developed the first benchmarking method: the ten-step Xerox benchmarking method. He defines benchmarking as: "the search for the best industry practices which will lead to exceptional performance through the implementation of these best practice" (Camp, 1989). Maheshwari (2013) defines benchmarking as: the measurement of some elements and the comparison of the results to a certain norm, the benchmark (Maheshwari, 2013, p. S83).

This study follows the definition of Maheshwari (2013) and adapts it towards this study's scope. Benchmarking is defined as: measuring some elements and comparing the results with comparable organisations.

By using monitoring and evaluation information and comparing this performance with others benchmarking can be beneficial. Accordingly the comparison is made on the following different levels (Anand & Kodali, 2008; Maheshwari, 2013):

- Organisational level, comparison inside one organization regarding the performance of related business parts.
- Industry level, comparison with companies operating in the same industry.
- Generic level, comparison with all other organizations.

Depending on the purpose of the benchmark a choice regarding the comparison level needs to be made.

Benchmarking purposes

Once an organisation has developed its policy, it can be implemented. This is done within the specified timeframe and at the end of this the achieved outcome is assessed. As a result an organisation paces through an planning cycle of policy, in which policy is developed, implemented, measured and evaluated (de Bruijn, 2002, p. 8). In the last step of the policy cycle, the policy is evaluated, in this step the effectiveness of the policy is assessed (Hoogerwerf & Herweijer, 2014). This is defined as the extent to which the output of the policy leads to the envisioned effects, the outcome.

Within the policy cycle benchmarking can fulfil a number of purposes, especially in the policy evaluation process step (de Bruijn, 2002, p. 8). The most frequent mentioned benchmarking purposes, although not always in the exact same terminology, are the following (de Bruijn, 2002; de Goede, Enserink, Worm, & van der Hoek, 2016):

- Transparency, makes the output an organization provides clear.
- Learning, as a result of the created transparency, an organization can learn by comparison of output and outcome.
- Appraising, with the comparison an appraisal of the performance may now be given (by the management of the organization or by third parties) about an organizations performance and effectiveness creating virtual competition.
- Sanctioning, finally after appraising the performance can be sanctioned. Resulting in incentives or a ranking.

De Bruijn (2002) uses the aforementioned general terms for benchmarking purposes. Maheshwari & Janssen (2013) specify the purpose of benchmarking more: Benchmarking is necessary for comprehend an organization's position and indicating opportunities to develop. Benchmarking models can be used by giving organizations an idea about their position (Maheshwari, 2013).

This study focuses on the designing and conducting a specific benchmarking study. Here the objective is to compare performance of policy between a multitude of organisations. Building on the definition of Maheshwari (2013), we define benchmarking as: *measuring some elements and comparing the results with comparable organisations.*

2.3. Benchmarking challenges.

In the aforementioned paragraph benchmarking is explained and its purposes, it can create transparency and enable learning. However apart from the beneficial effects, there are also challenges as indicated by de Bruijn (2002), Bannister (2007), Maheshwari (2013), Anand and Kodali (2008) and Susha (2015). They all stress the importance of the design of benchmarking. As stated in the problem statement: "The problem with benchmarking is that the wrong benchmarking design can lead to perverse effects and the fundamental cause for this that benchmarking methods often lack theoretical foundation".

The main challenge related to benchmarking is the design of the benchmarking method. Since the design of the benchmarking method can lead to perverse effects (de Bruijn, 2002, p. 5). De Bruijn (2002) describes seven perverse effects that apply for benchmarking in the public sector. He stresses that the benchmarking method should be designed in such a way that it minimizes the possible perverse effects and provides the beneficial effects (de Bruijn, 2002, p. 5). The seven perverse effects of performance measurement (de Bruijn, 2002, pp. 17–34):

1. *Performance measurement is an encouragement for strategic behaviour.* This occurs when an organization increases its output according to the criteria of the measurement system. This increase in output has no significance from a professional perspective. It is 'gaming the numbers' in order to obtain high scores and look good in the performance measurement rather than achieving the organizations objectives.
2. *Performance measurement blocks innovations.* An organisation will focus on output that is measured and choses for the save option. This is output that is well-known and easy to make. Because wanting to innovate means exploring the not known and taking the chance that the results are not what is anticipated.
3. *Performance measurement blocks ambitions.* Selection criteria for the input can result in strategically selecting the inputs in order to achieve the desired output with minimal effort. In other words, cherry picking, manipulating the quality or quantity of the input in order to achieve the wanted output.
4. *Performance measurement veils actual performance.* The measurement results need to be placed in context in order to assess the performance. During the measurement process, causal connections can be lost which give meaning to the results. This can result in a poor assessment of output while actually the effectiveness is immense.
5. *Performance measurement drives out the professional approach: no system responsibility, no quality, more bureaucracy.* The output of an (public) organisation is consistently a trade-off between different values. The performance measurement focusses on measurable output and when only one value is measured the organization will concentrate on this value and relinquish the other values. For example a museum that builds up a collection has a variety of values: the collections cultural value, an educational purpose, and serve the public. Acquiring a new piece for the collection benefits the cultural value and the educational purpose but might not attract many visitors. When only visitor numbers are measured possibly the cultural and educational values may suffer and therefor the museum will not comply to its system responsibility.
6. *Performance measurement leads to copying, not learning.* Benchmarking may degenerate into duplicating best practices. This is risky because in measuring the performance the complexity is

reduced making it a poor copy of reality. In addition success cannot always be taken over since each organization differs and have a specific target group.

7. *Performance measurement punishes good performance.* Due to the created transparency and good performance an organisation can receive higher targets. Or due to performance measurement an excellent output is achieved resulting in a surplus.

These possible perverse effects can considerably widen the gap between decision-making and implementation and therefore rendering the benchmarking ineffective (de Bruijn, 2002). The main cause according to de Bruijn (2002a) is because the design of the benchmarking method takes the complexity of the extent of the benchmark subject insufficiently into account.

Benchmarking methods often cannot deal with content variety and therefore it degenerates benchmarking into a quantitative activity (de Bruijn, 2002; Maheshwari, 2013). Herein the challenges are related to the lack of insight in the way of measuring and presenting the results.
How the results are obtained is challenging because only a limited number (Bannister, 2007) of easy to measure aspects are taken into account (de Bruijn, 2002). In addition the performance metrics is often ambiguous and therefore organisations can distrust the result of the benchmarking (Maheshwari, 2013).
How the benchmarking results are presented also poses to be challenging. Often the results are presented in a single score, because the scoring method reduces multiple measurements into a single number. This is challenging since the presentation of a single number results in an arbitrary score (Bannister, 2007; Maheshwari, 2013) and prevents the freedom of multiple perspectives on the organisations performance (de Bruijn, 2002). This is important since the results might be discussable and position dependant (Maheshwari, 2013).

The aforementioned challenges are mainly related with the contents of the design. Bannister (2007) supports de Bruijn that the main challenge is related to the design of the benchmarking method. Subsequently he indicates that the most challenges derive from the lack of theoretical foundation. Because benchmarking models often lack a methodological approach the used guidelines for benchmarking remain intuitive, presumptive and speculative since they lack or have a weak theoretical and empirical base (Bannister, 2007).

Related to the challenges with the design Susha (2015) found that organisations that want to benchmark cannot use a readily available benchmarking method. Since each method has to be uniquely designed (Susha et al., 2015, p. 13). This is necessary because each benchmark has its strengths and weaknesses and is applicable in specific situations (Susha et al., 2015, p. 13). Illustrative of this argument Anand and Kodali (2008) found a plethora of benchmarking methods (Anand & Kodali, 2008, p. 257) in which each model differs in terms of the number of phase, steps and application. The challenge that they see is that it creates uncertainty among the users as to which model to use.

2.4. Benchmarking criteria

The previous sections addressed benchmarking the definition, the purposes and its challenges. It will serve as a base for the following section. This section addresses the first part of research question one, namely what are the criteria for designing a benchmark? This is relevant because the main challenge with benchmarking is related to its design. The following section addresses the challenges from paragraph 2.3. and indicate criteria for the design of the benchmark in order to combat these. The

following authors are used for the criteria for designing a benchmark: de Bruijn (2002, 2007), Maheshwari (2013, 2015), and Janssen (2010).

De Bruijn stresses the importance of the design of the benchmarking model, his main focus is that the design has to take into account the complexity of the area in which it is operating. Therefore he defined three design criteria namely: interaction, variety, and dynamics (de Bruijn, 2002).
The first design criteria *interaction* is between the organisation that conduct the benchmark and the organisation that is benchmarked. It requires trust and fairness between the two parties. This can be achieved by providing understanding in the decisions related to the design of the benchmark. In other words indicate the way in which the products are defined. Specify the specific performance indicators for these products and give insight in how the performance is measured and assessed (de Bruijn, 2002, p. 55).
The way of presenting the benchmarking results appeared to be challenging since often in benchmarking only a single value as the result is presented. The second design principle *variety* states that the performance may be viewed from multiple perspectives and can therefore have more than one meaning.
Challenging for benchmarking is that it occasionally only focusses on easy to measure output. This can possibly result in a one sighted picture and does not display the whole performance. The third design criteria *dynamics* states that the measurement system should not only concentrate on the output but should be adaptable if the output changes over time. In addition it should make clear the process of creating the output.

Maheshwari (2013) points out seven principles as foundation for developing benchmarking models.
Firstly the benchmarking should take into account the organisations goals, focusing on measuring what should be accomplished, the desired effect of the output of the organisation.
Secondly differentiate between multiple measurement levels. The generic level allows for comparison with all other organisations. Industry level, enables comparison with companies operating in the same industry. And finally organisational level, allows for comparison inside one organization regarding the performance of related business parts.
Thirdly use different measurement methods and data sources. This principle combines different views to refrain from bias during the data collection.
Fourthly, define multiple dimensions within indicators in order to create views from different angles. This allows to focus on more aspects and less ambiguous performance metrics.
Fifthly, connect measures and benchmarking models, this addresses the lack of theoretical and empirical foundation. Provide for the indicators and models clear measurement guidelines and Build upon the current theoretical research.
Sixthly, Give suggestion for improvement. This addresses the challenge that learning is sometimes difficult since the focus is sometimes only on measuring and not on improving.
Seventhly and finally, Improve communication and interpretation by using a dashboard. The dashboard is used to counteract the lack of improvement support, ambiguous performance metrics, and different views on the performance. By visualising the results, it supports interpretation and avoid misinterpretation. In addition the dashboards can be used for future benchmarking results.

Janssen (2010) defined 5 criteria for a benchmarking instrument. The criteria for the benchmarking instrument include the following: simplicity, consume limited resources, measure in-depth, enable easy communication and give attention to a broad range of aspects (Janssen, 2010, p. 165).

- Simplicity, is targeting the use of objective criteria.
- Consume limited resources, more indicators result in the use of more resources making the comparison more complicated.
- Measure in-depth, use multiple criteria consisting of multiple aspects (indicators)
- Enable easy communication, results should be straightforward to visualize and communicate
- Give notice of a wide spectrum of aspects, avoid focussing on a single measure. Provide a variety of elements and let the organisation assess the performance themselves.

2.5. Benchmarking models

This paragraph focusses on the design of the benchmark and more specific on the design for the benchmarking model and the benchmarking process steps. The prefacing paragraphs stress the importance of the design of the benchmarking model and state that each model has its strengths and weaknesses and is applicable in specific situations (Susha et al., 2015, p. 13). Indicated as fundamental cause for the challenges related to benchmarking is that benchmarking methods often lack theoretical foundation (Bannister, 2007; Maheshwari, 2013). Therefor this paragraph uses the research and the benchmarking models of Anand and Kodali (2008), Bhutta and Huq (2008), and Camp (1989). Furthermore the research from the European Committee for Standardization (ECS) is used, they have established a benchmarking standard by scientific research and consensus. Between the models of the aforementioned authors a distinction can be made to objective of the benchmarking model. Camp developed a specific benchmarking methodology for the company Xerox. As opposed to Anand and Kodali (2008), Bhutta and Huq (2008) and ECS (2012) all three of them developed a universal benchmarking model. Camp is included as he is regarded the founding father of benchmarking and the other benchmarking models build upon his research.

Anand & Kodali (2008), suggest a benchmarking model that is applicable all types of benchmarking. In order to come to their model, they have benchmarked the benchmarking models. Based on a similairty analysis of 35 benchmarking models they came up with a 12-phase universal benchmarking model. Below the model is depicted as follows: phase number, phase name, followed by the description of the stage in the process.

12-phase universal benchmarking model (Anand & Kodali, 2008, pp. 269–287).

1. Team formation. Form the team that is going to conduct the benchmark. This can be within the organization or by external party.
2. Subject identification. Identify the benchmark subject.
3. Customer validation. Identify the clients for the benchmark information.
4. Management validation. Sketch the intentions and scope of the benchmarking project, and ask for authorization.
5. Self-analysis. Select the measurement method for the performance indicators.
6. Partner selection. Identify the potential benchmarking partners.
7. Pre-benchmarking activities. Determine data collection method and assess information needs.
8. Benchmarking. Perform the benchmarking study, arrange the collected data.
9. Gap analysis. Determine current gap and identify possible causes for the gaps.

10. Actions plans. Hand over the benchmarking results to the partners and establish functional goals.
11. Implementation. Implement proposed plans to overcome the gaps.
12. Continuous improvement. Following the implemented plans monitor the results, check, recalibrate, and update benchmarking.

Bhutta and Huq (1999) have analysed various benchmarking methods and concluded that the essence of all models is the same. Based on this analysis they suggest a five-step benchmarking model that follows the method of plan, do, act (Pulat, 1994). The five-step benchmarking model is outlined below.

1. Plan the study. Focusses on what to benchmark, the study should be aligned with the organisations strategic direction.
2. Form the benchmark team. This includes the team that is actually going to perform the benchmark. Furthermore it includes defining the method, and the performance indicators.
3. Identify partners. Selecting benchmarking partners, companies that are similar or that are considered world class in the specific process.
4. Collect data and analyse. Gather the need data and analyse it. Based on the analysed data identify strategies for improvement.
5. Adapt and improve. Based on the identified strategies adapt them in order to implement the specific improvements.

Camp developed the first benchmarking methodology for the company Xerox. According to Anand and Kodali (2008) this model is the fundamental classification scheme of benchmarking. The model is created in an industrial setting and can therefore not be implemented universally. This especially holds up for step 7 and 8. However it is the fundamental benchmarking model of which other models are formed after. The ten-step Xerox benchmarking model (Camp, 1989):

1. Identify what is to be benchmarked. The benchmarking subject for Xerox was the photocopier.
2. Identify comparative companies. Xerox chose to compare its performance to the performance of its rival that produced similar quality products but could sell them for less.
3. Determine data gathering method and collect data. At the time Xerox chose to visit the factory of its rival.
4. Establish current performance gap. The difference between the performance of Xerox and its rival.
5. Project future performance levels. Based on the gap analysis new objectives are created.
6. Communicate benchmark findings and gain acceptance. Benchmark results are conferred with the employers.
7. Establish functional goals. The objectives are translated into functional goals.
8. Develop action plans. Plan to put the functional goals in to action.
9. Implement specific actions and monitor results. Carry out the action plans.
10. Recalibrate benchmarks. With the improvements made start the benchmark process again.

The European Committee for Standardization (2012) has developed a standard for benchmarking. This standard is a document that sets out the requirements of benchmarking and provides a universal benchmarking model. The standard is established by scientific research and consensus. The guidelines for performance benchmarking are conform to NEN-EN 157221-7:2012 and is reproduced below:

1. Set the objectives. Determine the benchmarking purpose and scope.

2. Define the methodology. Select the indicators and the measures. For example measures of extremes, dispersion, and central tendency.
3. Select the team. Chose peers, a matching organisation to compare performance.
4. Collect data. Asses data availability, collect data and validate the data.
5. Analyse data. Determine the measurements per indicators based on step 2.
6. Determine gaps. Compare the measurements to find gaps. Subsequently check whether the gaps can be explained.
7. Report findings. Communicate and explain the findings of the study.
8. Develop action plan. When gaps are identified develop a procedure of tasks to eliminate the gaps.
9. Implement plan. Execute the action plan and start to monitor the change.
10. Process review. Based on the conducted benchmarking review the process.

2.6. Benchmarking conclusion

This chapter provided detailed insight regarding benchmarking, and answers the first research question. *What are the process steps and criteria for designing a benchmark?* With literature review and this study's scope first benchmarking is defined as: measuring some elements and comparing the results with comparable organisations. The reason why organisations measure and compare is because organisations want to know and improve their performance. The method that has been proven useful for measuring and comparing of performance is benchmarking. Benchmarking uses monitoring and evaluation information to compare results, with the aim of creating transparency and enable learning. This commonly occurs in the policy cycle, more precise in the performance evaluation step, through comparing output on different levels: organisational level, industry level, or generic level.

The process of benchmarking is not straight forward, as showed in paragraph 2.2. In that paragraph the benchmarking challenges were identified and there it became apparent that the design of benchmarking is crucial. A wrongly designed benchmark could render it ineffective. Indicated as leading cause was the lack of theoretical foundation for the benchmarking.

Based on the foregoing, it is obvious that specific attention should be paid to the design of the benchmark. Therefor firstly the specific criteria for designing a benchmark are combined in an overview in Table 2. Secondly the benchmarking process steps for designing a benchmark are established. Table 3 presents an overview of the benchmarking process steps.

Table 2, depicted below, gives an overview of the benchmarking criteria. Common elements are the usage of indicators, transparency in performance measurement method and self-assessment of performance

Author	Criteria for model	Code	Description criteria
Bruijn	1. Interaction: Trust, fairness	b1	Indicate how the products are defined.
Three criteria	1. Interaction: Trust, fairness	b2	Indicate the performance indicators.
	1. Interaction: Trust, fairness	b3	Explain how the performance is measured and assed.
	2. Variety: Content	b4	Performance may be viewed from multiple perspectives and can therefore have more than one meaning.
	3. Dynamics: liveliness	b5	PM must make clear the process of creating products.
	3. Dynamics: liveliness	b6	Attention to changes to an organizations product.
Maheshwari	1. benchmarking should take into account the organisations goals	m1	Aim attention on measuring and what should be accomplished.
Seven principles	2. Differentiate between multiple measurement levels	m2	Chose the benchmarking level organisational, industry, or generic.
	3. Use different measurement methods and data sources	m3	combine different views to refrain from bias during the data collection.
	4. Define multiple dimensions within indicators	m4	Focus on dimensions within the indicators to get better insights.
	5. Connect measurements and benchmarking models	m5	Provide clear measurement guidelines.
	6. Give suggestion for improvement.	m6	Provide improvement support.
	7. Improve communication and interpretation by using a dashboard	m7	Provide insight at a glance, support interpretation by showing various measures.
Janssen	1. Simplicity	j1	Use objective criteria.
Requirements	2. Consume limited resources	j2	More indicators result in the use of more resources making the comparison more complicated
	3. Measure in-depth	j3	Use multiple criteria consisting of multiple aspects (dimensions)
	4. Enable easy communication	j4	Results should be straightforward to visualize and communicate
	5. Give notice of a wide spectrum of aspects	j5	Avoid focus on a single measure. Provide a variety of elements. And let the organisation asses the performance them self.

Table 2 Overview of benchmarking criteria identified by the authors de Bruijn (2002)*, Maheshwari* (2013)*, and Janssen* (2010)*. The benchmarking criteria are identified with a code. The coding is as follows the first letter of the author's surname is used accompanied with the number of the criterion.*

For the second part of research question one the benchmarking process steps for designing a benchmark are pointed out. Table 3 presents an overview of the benchmarking process steps. The literature regarding benchmarking models showed that all benchmarking models have certain steps in order to compare the performance. Common terms in benchmarking are measurement, comparison, identification of best practices, implementation and improvement (Anand & Kodali, 2008, p. 258).

Steps in benchmarking model	Authors			
	Anand and Kodali (2008)	**Bhutta and Huq (1999)**	**Camp (1989)**	**ECS (2012)**
	A1. Team formation	B1. Plan the study	C1. Identify what is to be benchmarked	E1. Objectives
	A2. Subject identification	B2. Form the benchmark team	C2. Identify comparative companies	E2. Methodology
	A3. Customer validation	B3. Identify partners	C3. Determine data collection method and collect data	E3. Team
	A4. Management validation	B4. Collect data and analyse	C4. Determine current performance gap	E4. Collect data
	A5. Self-analysis	B5. Adapt and improve	C5. Project future performance levels	E5. Analyse data
	A6. Partner selection		C6. Communicate benchmark findings and gain acceptance	E6. Determine gaps
	A7. Pre-benchmarking activities		C7. Establish functional goals	E7. Report findings
	A8. Benchmarking		C8. Develop action plans	E8. Develop action plan
	A9. Gap analysis		C9. Implement specific actions and monitor results	E9. Implement plan
	A10. Actions plans		C10. Recalibrate benchmarks	E10. Process review
	A11. Implementation			
	A12. Continuous improvement			

Table 3 Overview of benchmarking models of Anand and Kodali (2008), Bhutta and Huq (1999), Camp (1989) and European Committee for Standardization (2012). The benchmarking process steps are identified with a code. The coding is as follows the first letter of the author's surname is used accompanied with the number for the process step.

The second goal of the literature review is to design a specific benchmarking method. Chapter three will use the overviews with the identified benchmarking criteria and benchmarking process steps and use it to design the specific benchmarking method, a sport policy benchmark.

3. Design municipal sport policy benchmarking

This chapter describes the design of the benchmarking model, that will serve as the foundation for benchmarking municipal sport policy. This chapter is directed to answer research question two: *How can we design a sport policy benchmark?* To answer this question the identified benchmarking criteria and benchmarking process steps from chapter 2 will be used.
For the literature review parts this chapter also uses the same approach as mentioned in the previous chapter and elaborated in paragraph 2.1.

3.1. Benchmarking criteria

In the previous chapter an overview of benchmarking criteria is constructed and depicted in Table 2 page 25. The overview consists of the benchmarking criteria as defined by de Bruijn (2002), Maheshwari (2013), and Janssen (2010). With the overview the following common elements can be derived:

- *Usage of indicators*. De Bruijn has one criteria (b2) totally directed at indicating what the performance indicators are. Both Maheshwari (2013) and Janssen (2010) have the same criteria but do not use the word exact same word 'indicator' but use a different word with an equivalent meaning. Maheshwari (2013) uses the word areas (m4,) and Janssen (2010) uses aspects (j2).
- *Transparency in performance measurement method.* For this criteria de Bruijn (2002) makes specific reference (b3) Give insight in how the performance is measured. However Maheshwari (2013) and Janssen (2010) do not make a specific reference in their criteria but it is implicit in their reporting of the results.
- *Self-assessment of performance*. For this element, all three agree with each other. Since the result of benchmarking can have more than one meaning the result should be viewed from multiple perspectives and preferably by the organisation them self, de Bruijn (2002) (b4) and Janssen (2010) (j5). Maheshwari (2013) goes one step further and also gives a direction to the solution (m7). For the solution he creates the condition of using a dashboard. A dashboard is an instrument that enhances communication and interpretation. This is achieved by visualising the result in one overview, enabling the organisation to assess the performance them self.

	Authors		
Benchmarking criteria	**Bruijn**	**Janssen**	**Maheshwari**
Usage of indicators	b2. Interaction. Indicate the performance indicators	j2. Consume limited resources	m4. Define multiple dimensions within indicators
Transparency in performance measurement method	b3. Interaction. Explain how the performance is measured and assed.	No specific reference. Implicit in reporting	No specific reference. Implicit in reporting
Self-assessment of performance	b4. Variety: Content. Performance viewed from multiple perspectives.	j5. Give notice of a wide spectrum of aspects	m7. Use dashboards to enhance communication and interpretation.

Table 4 Common elements in the benchmarking criteria ordered per author. With the common element indicated whose criterion contributes to the common element.

3.2. Benchmarking process steps

The previous chapter concluded with an overview of benchmarking models. In the overview the models of the following authors are depicted: Anand and Kodali (2008), Bhutta and Huq (2008), Camp (1989) and the European Committee for Standardization (2012). The overview is described in Table 3 page 26.

The four models differ in aim because Camp developed a specific benchmarking model for the company Xerox and the other three models are universal benchmarking models. The review of the literature regarding benchmarking showed five common terms in benchmarking: measurement, comparison, identification of best practices, implementation and improvement (Anand & Kodali, 2008, p. 258). When reviewing the benchmarking models it stand out that all benchmarking models have certain steps in order to compare the performance. Common elements in the benchmarking models are the general phases of preparing benchmarking, comparing performance, and improving by implementing improvements.

Preparing phase

For the preparing phase of the benchmark, one author uses seven steps while the others use only two or three steps. More specific for the preparing/planning of the benchmark Anand and Kodali (2008) use the steps, A1-A7, Bhutta and Huq (2008) the steps B1-B3, Camp (1989). C1, C2 and the European Committee for Standardization (2012) E1-E3.

Within this phase the common elements are the identification of objective (A2, B1, C1, E1) the methodology (A7, B1, C1, E2) and the partner selection to compare against (A6, B3, C2, E3). See Table 5 for the overview of steps in the preparing phase.

	Authors			
Steps in benchmarking model	**Anand and Kodali (2008)**	**Bhutta and Huq (1999)**	**Camp (1989)**	**ECS (2012)**
	A1. Team formation (p)	B1. Plan the study (p)	C1. Identify what is to be benchmarked (p)	E1. Objectives (p)
	A2. Subject identification (p)	B2. Form the benchmark team (p)	C2. Identify comparative companies (p)	E2. Methodology (p)
	A3. Customer validation (p)	B3. Identify partners (p)		E3. Team (p)
	A4. Management validation (p)			
	A5. Self-analysis (p)			
	A6. Partner selection (p)			
	A7. Pre-benchmarking activities (p)			

Table 5 Overview steps preparing phase per author. In the steps indicated that the performed step contributes the common element preparing (p).

Comparing phase

In the comparing phase the benchmark is carried out. The common elements are: collecting of data, analysing data, determine gaps, and report findings. However not all scholars make specific reference is to all elements. For example Anand and Kodali (2008) do not specifically set out the different steps in the benchmarking comparing phase and solely refer to the process step benchmarking (A8). Further Camp (1989) makes no specific reference to analysing the data and Bhuta (1999) makes no specific reference to determining the gaps.

The common elements are

- Collecting data, (B4, C3, E4)
- Analysing data, (B4, E5)
- Determining gaps, (A9, C4, E6)
- Reporting findings, (C6, E7)

Steps in benchmarking model	Authors			
	Anand and Kodali (2008)	**Bhutta and Huq (1999)**	**Camp (1989)**	**ECS (2012)**
	A8. Benchmarking (c)	B4. Collect data and analyse (c)	C3. Determine data collection method and collect data (c)	E4. Collecting data (c)
	A9. Gap analysis (c)		C4. Determine current performance gap. (c)	E5. Analysing data (c)
			C5. Project future performance levels. (c)	E6. Determining gaps (c)
			C6. Communicate benchmark findings and gain acceptance (c)	E7. Reporting findings (c)

Table 6 Overview steps comparing phase per author. In the steps indicated that the performed step contributes the common element comparing (c).

Improving phase

In this phase an action plan is developed and implemented in order to eliminate the gaps identified in the previous phase. Anand and Kodali (2008) and Camp (1989) make reference to the continuous character of benchmarking and define a criteria to amend the benchmark based on the implemented actions. However Bhutta and Huq (2008) and the European Committee for Standardization (2012) do not make this reference and therefore this omitted. The common elements are:

- Develop action plans. Based on the identified gaps the organisation develops a procedure of tasks to eliminate those gaps. (A10, C7, C8, E8)
- Implement plan. Execute the developed action plan. (A11, B5, C9, E9)

Steps in benchmarking model	Authors			
	Anand and Kodali (2008)	**Bhutta and Huq (1999)**	**Camp (1989)**	**ECS (2012)**
	A10. Actions plans (i)	B5. Adapt and improve (i)	C7. Establish functional goals (i)	E8. Develop action plan (i)
	A11. Implementation (i)		C8. Develop action plans (i)	E9. Implement plan (i)
	A12. Continuous improvement (i)		C9. Implement specific actions and monitor results (i)	E10. Process review (i)
			C10. Recalibrate benchmarks (i)	

Table 7 Overview steps improving phase per author. In the steps indicated that the performed step contributes the common element improving (i).

3.3. Conclusion design municipal sport policy benchmarking

This chapter answers the research question *How can we design a sport policy benchmark?* As indicated the design of the benchmarking is crucial and needs to be tailored to the specific situation. This chapter describes the design of the benchmarking model that will serve as the foundation for the specific benchmarking. In this study's specific situation, the benchmark is designed for benchmarking municipal sport policy. This study designs and conducts the benchmark and therefor it is executed by an external party that has no decision-making authority and can be treated as a consultancy firm. Hence due the influence power, the improving phase falls not within the scope of this study and will therefore be omitted.

The design of the benchmarking method consists of phases, process steps and criteria that are identified in the literature review. In this chapter the common elements in the benchmarking process steps and criteria are pointed out. Based on this study's scope, the benchmarking phases and the criteria the following model is designed, see Table 8.

The table gives an overview of the benchmarking phases: preparing and comparing. The benchmarking preparing phase consist of the following three benchmarking steps: setting objectives, defining indicators, selecting benchmark groups. The benchmarking comparing phase consist of the following four benchmarking steps: collecting data, preparing data, analysing data, determining significant different findings and report findings. For each benchmarking step an abbreviated description of the step is given and it is indicated which criteria it addresses from de Bruijn (2002), Maheshwari (2013), or Janssen (2010).

Phase	Steps	Criteria	Description step
Preparing	1.Setting objectives	b1, b6, m1, m2	Indicate the scope of the research: the domain, the frequency and the measurement level.
	2. Defining indicators	b1, b2, b3, b6, m1, m4, j1, j2, j3	Use indicators with dimensions that have a scientific proven causal relationship and that are measurable (take data availability in account). Determine the measures.
	3. Selecting benchmark groups	b1, b2, b3, j1	Compare against comparable organisation, under the same circumstances, and with the same measures. Use objective criteria with scientific proven relationship to determine comparable groups.
Comparing	4. Collecting data	b3, m3, j2	Use multiple data collection methods and reliable data sets that are readily available.
	5. Preparing data	b3	Combing and normalising data sets, show what's done.
	6. Analysing data	b3, m5	Examine the data with regard to the indicators and the dimensions.
	7. Determining significant different findings	b3	Determine the gaps between individual indicators and the measures. Compare individual score to the collective predefined score.
	8. Reporting findings	b4, b5, m1, m7, j4, j5	In dashboard give overview of the measurements.

Table 8 Design of the benchmarking method. Consisting of phases, steps and criteria.
The phases and steps are determined by the common elements of benchmarking models of Anand and Kodali (2008), Bhutta and Huq (2008), Camp (1989) and the European Committee for Standardization (2012). The benchmarking criteria are derived by combining the benchmarking criteria from de Bruijn (2002), Maheshwari (2013), and Janssen (2010).

The designed benchmarking method, depicted in Figure 5, serves as the foundation for the specific benchmarking. In this study the domain that is benchmarked is municipal sport policy. The next chapter

will elaborate on the designed benchmarking method and will implement it for the benchmarking of municipal sport policy.

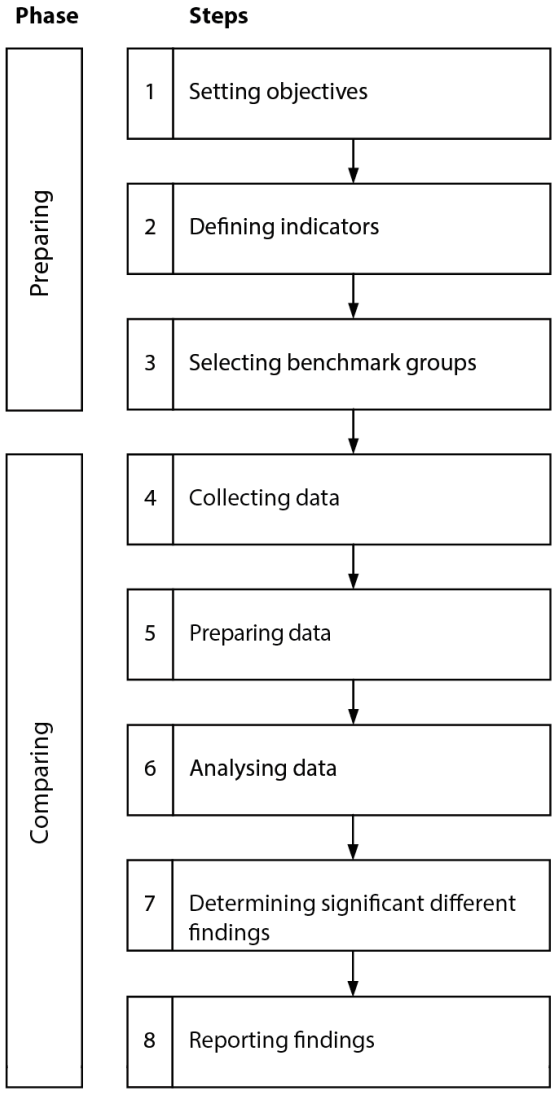


Figure 5 Overview of the designed benchmarking method.

4. Benchmarking application: preparing phase

In this chapter the benchmarking method that is developed in chapter 3 Design municipal sport policy benchmark is expanded and explained for the various steps. The benchmarking method can be divided into two main phases the preparing phase and comparing phase. This chapter focusses on the preparing phase of benchmarking and answers research question three: *What does the sport policy benchmark look like?* In order to answer the research question the preparing phase of the benchmarking is carried out. The preparing phase is divided into three steps, see Figure 6. Step one: setting objectives, step two: defining indicators, step three: defining benchmarking group. The steps will be answered according to their order of rank.

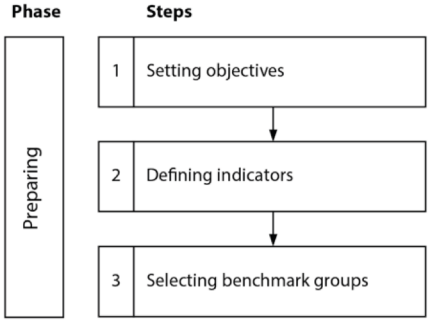


Figure 6 Benchmarking preparing phase.

4.1. Step 1 setting objectives

The criteria and description for step 1 as described in the design of the benchmarking method is:
Indicate the scope of the research: the domain, the frequency and the measurement level.
For the scope of the research we have to look back to the practical part of the problem statement. From there it became clear that municipal policy makers lack insight in the effectiveness of sport policy (van der Poel et al., 2016, p. 253). They do not know whether their budget for sport policy is spent effectively. There is a lack of insight in whether the output of their sport policy (namely the sport facilities) leads to the desired effects (namely, healthy citizens).

Insight in the effectiveness of a sport policy is crucial for municipalities, because it can be used in the policy evaluation step to assess the performance of this policy. It gives municipalities insight in whether they spend their sport budget in the way that leads to the desired effects. To increase the impact of the policy evaluation comparison with other municipalities can provide additional insights and stimulate learning (de Bruijn, 2007; Neely et al., 1995). It can help organizations understand their own position and identify growth opportunities (Maheshwari, 2013). With respect to the domain of the benchmarking, this study focusses on municipal sport policy, more specifically in the effectiveness of the sport policy.

Considering the frequency of the benchmarking this study designs and conducts a benchmark of municipal sport policy. The objective of the study is to design and perform the benchmarking method once. If the benchmarking proves to be successful municipal policy makers can use the design in order continue to measure and compare the performance.

Regarding the measurement level this study compares the sport facilities in each Dutch municipality on industry level. As defined in the design comparison can be made on three different levels on organisational level, industry level, or generic level (Anand & Kodali, 2008; Maheshwari, 2013). This study aims to compare against other municipalities since they are the only organisation that operates within the same industry and has relatable objectives.

Hence the objective is to increase insight in the municipal policy evaluation step. This is insight is regarding the effectiveness of the municipal sport policy. Therefore a benchmarking study of sport

policy to *measure effectiveness of sport policy* is conducted. The benchmark is performed once and compares with other municipalities in the Netherlands.

4.2. Step 2 defining indicators

The criteria for step 2 as defined in the design: *Use indicators with dimensions that have a scientific proven causal relationship and that are measurable (take data availability in account). Determine the measures.*
Thus, the criteria for the indicators can be divided in three parts. First part selecting indicators, with the condition that are aligned with the municipality's strategic direction and the objective of the benchmark. Second part selecting measurable dimensions, with the requirement that they have a scientific proven causal relationship. The final and third part determining the measures, the standard unit used to express the quantity of the dimension.

The objective of the benchmark is to increase insight for the policy-makers in the municipal policy evaluation step. Benchmarking the effectiveness of the sport policy of municipalities is expected to contribute to the insight of the policy evaluation. As mentioned in chapter 1, policy evaluation is the process that tries to assess the effectiveness of a policy. This is the extent to which the policy performance, the output, leads to the envisioned effects, the outcome (Hoogerwerf & Herweijer, 2014). Implicit in the objective are the indicators output and outcome. Output is directed to the policy performance and the outcome is the policy effect. Hence the indicators to determine effectiveness are: output and outcome.

Following the selected indicators, output and outcome, specific measurable dimensions need to be determined with the condition that they have a scientific proven causal relationship. Based on this criteria literature regarding sport facilities and municipal real estate needs to be reviewed. In an attempt to find dimensions related to the municipal sport policy's output and outcome various studies regarding these two topics are researched and dimensions are determined that are in line with the strategic direction of municipalities.

As elaborated in chapter 1 the role of the municipalities in sport policy is one of policy-maker. Their main interest is to ensure sport policy implementation effectively, in other words to provide an adequate sport service level with the objective to have healthy citizens (van der Poel et al., 2016, p. 253). To achieve this municipal sport policy concern two categories sport facilities and sport participation programs. This study focusses on sport facilities since this is measurable and in terms of money it by far the most important instrument used in municipal sport policy.

In recent years quiet some body of knowledge is accumulated regarding the subjects: sport policy, sport facilities and sport participation. As a starting point for the selection of dimensions for the output and outcome indicator the research of the European Sport for All Charter (Council of Europe, 2001) is used. In their research they make specific reference to the interdependence of *sport participation* and the *presence*, *proximity* and *variety* of sport facilities (Hoekman, Breedveld, & Kraaykamp, 2016, p. 2). These dimensions, though not always used in the same terminology, are extensive researched in various national and international studies and appear to be the main dimensions used to determine the link between sport facilities and physical activity. Which will be explained in the consecutive sections.

Based on the reviewed literature, the following dimensions can be observed for the output and outcome of municipal sport policy:

- Output dimensions: presence, proximity and variety of sport facilities
- Outcome dimensions: sport participation.

Below the reason for the selection of the specific dimensions will be elaborated. First the reasons for the selection of the output dimensions followed by the explanation for the outcome dimension.

	Indicators	Dimension	Referenced by
Step 2	*Output*	*Presence*	Atkinson (2005), Chad (2005), Haug (2008), Humpel (2002), Lim (2011), Limstrand (2008), Saelens (2003), Wicker (2009).
		Proximity	Limstrand (2008), Sallis (1990), Wicker (2009).
		Variety	Hallman (2011), Hoekman (2016), Limstrand (2008), Poel (2016), Wicker (2009).
	Outcome	*Sport participation*	Hildebrandt (2008), Hoekman (2016).

Table 9 Benchmarking step 2 Overview indicators, the corresponding dimensions and the author who references the relation between the dimension and physical activity.

Output dimension presence

Various scholars have concluded that there is an interdependence between physical activity and the presence of sport facilities (Humpel et al., 2002; Saelens et al., 2003; Wicker, et al., 2009). They derive to this conclusion from two ways. Some researchers see that low availability of sport facilities can restrict sport participation (Atkinson et al., 2005; Lim et al., 2011). While other come to this conclusion that a good supply of sport facilities promotes sport participation (Chad et al., 2005; Haug et al., 2008; Limstrand & Reher, 2008). In both ways they do not further indicate what good or low availability is. Wicker (2013) goes even further and states that all types of sport facilities have a positive impact on sport participation (Wicker et al., 2013, p. 57). Based on this and with the connection to municipal sport policy, the presence of sport facilities is an indication for the availability of sport facilities in the municipality. We define presence as the total number of sport facilities in the municipality.

Output dimension proximity

The nearness of sport facilities seems especially valuable since several studies have found that a small distance to sport facilities positively influences the sport participation. In addition other studies state that sport participation may be influenced by the proximity towards a variety and number of sport facilities (Limstrand & Rehrer, 2008; Wicker et al., 2009). For their analysis both Wicker (2009) and Limstrand (2008) use for proximity urban districts. More specific Wicker (2009) focusses on the urban districts of Stuttgard and Limstrand (2008) on Norwegian urban districts. Sallis (1990) and Wicker (2009) find directly that the distance is positively linked to sport participation (Sallis et al., 1990; Wicker et al., 2009). In addition Sallis (1990) states that the proximity to sport facilities is connected with exercise behaviour and that it independent of demographic variables (Sallis et al., 1990, p. 184). As a cause for a positive relation between proximity to sport facilities and physical activity they argue that the proximity reduces the psychological barrier to exercise.

Gilles-Corti (2002) argues that in some instances the proximity plays a smaller role than the attractiveness of the sport facility. The attractiveness is determined by its nearness to sport facilities or activities, for example the multi-sport complexes (Giles-Corti & Donovan, 2002, p. 1795).

Based on this and with the connection to municipal sport policy, the proximity of sport facilities is an indication for the nearness of sport facilities to an individual. We follow for our definition the approach of Wicker (2009) and Limstrand (2008) and use for calculating proximity urban districts. Therefore we define proximity as the amount of sport facilities on district level in a municipality (in Dutch: wijk).

Output dimension variety

The diversity of sport facilities in a municipality seems to be a relevant aspect. Van der Poel (2016, p. 105) makes a special reference to the ability to choose for a type sport and corresponding sport facility. The power to select a specific type of sport is in several studies indicated as a relevant aspect since it may influence their sport participation (Hallmann, Wicker, Breuer, & Schüttoff, 2011; Limstrand & Rehrer, 2008; Wicker et al., 2009). Again measurements were done on district level and the aspect was not further specified. In other words it was not indicated what a desirable choice amount of sport types is only that is a relevant aspect.

Hoekman (2016) has research the variety of sport facilities in the Netherlands. To define variety they used a radius of 1 km and counted the diversity of the types of sport facilities within the set range. With the measurements they concluded that the variety of sport facilities within 1 km is significant for explaining the sport participation levels of an individual. For causes of higher sport activity levels they indicate the smaller travel distance allows for more impromptu decisions to exercise (Hoekman et al., 2016, p. 11).

Based on the aforementioned and the relation to municipal sport policy, the variety of sport facilities is an indication for the number of sport facilities that are distinct in character. We follow for our definition the approach of Wicker (2009) and Limstrand (2008) and use for calculating variety urban districts. In addition to give additional insight the variety is also presented on municipal level since this is an indication for the sport service level. Therefore we define variety as the number of distinct sport facilities on district and municipal level.

Outcome dimension sport participation

The envisioned effect of the municipal sport policy, is to have healthy citizen. The dimensions related to this should be in line with this ambition. In determining the dimensions related to the output of municipal sport policy, the term sport participation was mentioned in each dimension. As a starting point the reference of the European Sport for All Charter (Council of Europe, 2001) was used, they specifically mention the interdependence of sport participation and the presence, proximity and variety of sport facilities (Hoekman et al., 2016, p. 2). Sport participation and accordingly physical activity provide health benefits (Ainsworth et al., 2011; Warburton, 2006). For the Dutch ministry of Health, Welfare and Sport, physical activity is used as an instrument for the creation of a healthy society (VWS, 2017). In order to measure physical activity levels of Dutch citizens, they have defined a standard, the fit-norm.

The fit-norm is a sport frequency norm and requires at least three times a week, intensive physical activity for at least 20 minutes (Hildebrandt, Ooijendijk, & Hopman-Rock, 2008). The fit-norm is a part of the larger study the Health monitor. This study is performed by the Municipal Health Services (GGD) and is directed to measure the health of citizen. For their results the health monitor have conducted a nation-wide research under the Dutch citizen. This research consisted of surveys and interviews, sample size n=402.195. Because this way of data collection requires self-assessment of physical activity they use the research of Ainsworth et al., (2000, 2011). Ainsworth (2000) developed the Compendium of Physical

Activities for the purpose: "to enhance the comparability of results across studies using self-report physical activity and is used to quantify the energy cost of a wide variety of physical activity" (Ainsworth et al., 2011, p. 1). The fit-norm will be elaborated more specifically in step 4 collecting data.

Based on the aforementioned and the relation to municipal sport policy, the dimension to the indicator outcome is sport participation. In this study sport participation is represented by the fit-norm. This is a standard determined by the ministry of Health, Welfare and Sport for the frequency of intensive physical activity each week. Therefore we define sport participation as: percentage of citizen that achieves the sport frequency norm, fit-norm.

Measures for the dimensions

Continuing on the dimensions for the indicators the measures need to be determined. Measures are standard units used to express the quantity of the dimension. To give insight in the effectiveness of the municipal sport policy the following measures for the dimensions will be used (NEN-EN 15221-7, 2012, p. 8):

- measures of dispersion, provides an indication of the typical bandwidths in a data set
- measure of extremes, provides an indication of the extreme scores in a data set

Conclusion defining indicators

To recapitulate, as part of the benchmarking preparing phase in step 2 the dimensions for the output and outcome indicators are defined. The dimensions are based on the criteria for dimensions depicted in Table 8 page 30. Specific literature regarding sport facilities was reviewed leading to the following dimensions for the indicators:

Dimensions for the indicator output. Representing the policy performance, the executed actions resulting from the municipal sport policy:

- *Presence*, number of sport facilities within the municipality (#).
- *Proximity*, number of sport facilities in district (km).
- *Variety*, number of different sport facilities with district (x,y,z).

Dimensions for the indicator outcome Representing the policy effect, the effect of the implemented measures, the output.

- *Sport participation*; percentage that achieves the fit-norm (%).

With literature review the above mentioned dimensions were found. It was found that these dimensions have a positive relation with sport participation. Hence based on the reviewed literature it is expected that higher numbers for presence, proximity, and variety positively influence the sport participation.
Below the relation between the indicators and the dimensions are depicted in Figure 7.

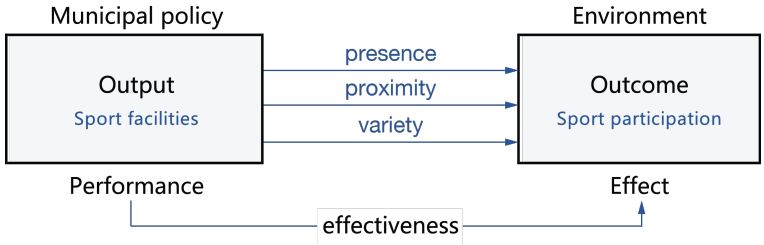


Figure 7 relation between indicators output and outcome and the dimensions in blue. The output dimensions are the presence, proximity and variety concerning sport facilities and the outcome dimension is sport participation. The figure is adapted from Hoogerwerf and Herwijer (2014) and towards this study's scope.

4.3. Step 3 defining comparable municipalities

The criterion and description for step 3 as described in the design of the benchmarking method is:
Comparison against comparable organisations, under the same circumstance and with the same measures. Use objective criteria with preferably scientific proven relationship.
This criterion implies two parts. First make groups of comparable organisations and use literature to determine criteria for groups. Second compare under the same circumstances and with the same measures.

The first part of step 3 is to compare against comparable organisations and preferably use scientific proven relationship to determine comparable. In the description of the criterion it is stated that comparable organisations are organisation for which the scope of services are comparable (NEN-EN 15221-7, 2012, p. 67). As determined in step 1 the organisations that are benchmarked are all the municipalities in the Netherlands. Therefore it is a benchmarking on industry level meaning comparison with companies operating in the same industry (Anand & Kodali, 2008; Maheshwari, 2013). Even though the organisations are the same kind of public organisation there are differences. A couple of examples of differences between municipalities are size, objectives, number inhabitants, budget, and urbanity. Due to these differences the formation of groups of comparable municipalities are expected to be beneficial and contribute to the relevance of the benchmarking results.

To select municipalities that benefit from comparing to each other benchmarking literature and sport literature is reviewed. Both topics are reviewed with the objective to determine how groups are formed in the specific literature. Benchmarking literature of the following authors is used: Aarts (2014), and Den Heijer and de Vries (2004). Sport literature of Gordon-Larsen (2006), Hallmann (2012), Hillsdon (2007), Hoekman (2016), van der Poel (2016), and Steinmayr (2011) is used.

Aarts (2014) takes in his benchmarking study of municipal road safety the following factors in to account: size of municipality and degree of urbanization (Aarts, 2014, p. 3). Den Heijer and de Vries (2004) used no factors to determine groups for their benchmarking study. They developed and conducted a benchmarking study between all universities in the Netherlands and their real estate. For the group selection, they made no distinction between the universities. Resulting in a benchmark on industry level with all universities in one group.

In sport literature the following two aspects are used: deprivation and urbanity. Hallman (2012) used the two aspects to determine the sport impact in Germany. Hoekman (2016) researched the importance of travel distance to the sport facility that is shortest in distance and found that different studies make distinction between areas based on two aspects. The first aspect: level of deprivation of an area (Gordon-Larsen et al. 2006, Hillsdon et al. 2007, Lamb et al. 2012 in Hoekman et al., 2016b). The second aspect: level of urbanity (Bale 1982, Hallmann et al. 2011, Steinmayr et al. 2011 in Hoekman et al., 2016b). Based on this they chose to utilizes both factors to explore the way sport facilities are spread out in the Netherlands (Hoekman et al., 2016). When both definitions are taken apart, either deprivation and urban areas come down to municipal size. Because urbanity and deprivation are based on the environments density measured with amount addresses per km^2.

Based on the reviewed literature the following aspects were used for forming group:

- All in one group (Heijer & de Vries, 2004),
- Municipal size (Aarts, 2014; Hallmann et al., 2012; Hoekman et al., 2016),

In the Netherlands CBS is the entity that collects and publishes statistical information regarding societal issues. According to the CBS the Netherlands has 383 municipalities. To enable benchmarking between comparable organisations this study aims to categorize the municipalities in groups. With the reviewed literature it is clear that in these studies no clear guidelines were formed for the formation of groups of municipalities. One aspects that differentiates between municipalities is found namely, municipal size. This study will use the aspect municipal size to form groups. It is expected that pooling on size will help with the dispersal of differences between the municipalities. In other words by arranging on size the objectives and budgets of the municipalities are expected to be less dispersed.

The aspect of municipal size is also acknowledged in practice. This is illustrated by CBS (2017) that classifies municipalities based on their population. This is further concentrated into five or three classes based on municipal size. The three classes are based on the cooperation that large municipalities in the Netherlands have formed. In the Netherlands, the large cities work together. There is cooperation between the four largest cities the G4 and the large cities, the G32. The initiatives are founded in order to exchange knowledge and to promote the interests of the cities to the first and second chambers in parliament. The G4 is a network of the largest cities in the Netherlands namely, Amsterdam, Rotterdam, The Hague and Utrecht. The G32-network is a network of 38 middle to large cities in the Netherlands. The 38 municipalities of the G32 are: Alkmaar, Almelo, Almere, Alphen aan den Rijn, Amersfoort, Apeldoorn, Arnhem, Breda, Delft, Deventer, Dordrecht, Ede, Eindhoven, Emmen, Enschede, Gouda, Groningen, Haarlem, Haarlemmermeer, Heerlen, Helmond, Hengelo, 's-Hertogenbosch, Hoorn, Leeuwarden, Leiden, Lelystad, Maastricht, Nijmegen, Oss, Roosendaal, Sittard-Geleen, Schiedam, Tilburg, Venlo, Zaanstad, Zoetermeer and Zwolle. The networks are determined based on municipal size and consist of the G4 and the G32.

To form comparable groups of municipalities the aspect municipal size is used. Currently the large municipalities in the Netherlands have formed a cooperation based on size in which they work together, the G4 and G32. These two initiatives are used for the classification of comparable groups of municipalities.

4.4. Benchmarking application conclusions

In order to answer research question 3: What does the sport policy benchmark look like? This chapter has completed the benchmarking preparing phase, by performing the process steps 1-3. Below the results of the steps will be shortly repeated.

Step 1: Was aimed towards the scope of the research. The objective of the benchmarking is to increase insight in the municipal policy evaluation step regarding the effectiveness of the municipal sport policy. Therefore a benchmarking study of sport policy to *measure effectiveness of sport policy* is conducted. The benchmark is performed once and compares with other municipalities in the Netherlands.

Step 2: Is directed to find indicators, dimensions and the measures. Based on literature review regarding sport facilities the following indicators and dimensions were determined:

Dimensions for the indicator output

- Presence, number of sport facilities within the municipality. (#)
- Proximity, number of sport facilities in district (km)
- Variety, number of different sport facilities within district (x,y,z)

Dimensions for the indicator outcome

- Sport participation; percentage that achieves the fit-norm.

Based on the reviewed literature and the found relations between the dimensions it is expected that higher scores for the presence, proximity and variety positively influence the sport participation levels.

Step 3: Addresses the formation of comparable groups of municipalities. With literature review an aspect was established for the formation of groups namely municipal size. Currently the large municipalities in the Netherlands work together and have formed a cooperation based on municipal size, the G4 and G32. These two initiatives are used for the classification of comparable groups of municipalities. This results in the following groups:

- G4 Amsterdam, Rotterdam, The Hague and Utrecht
- G32: Alkmaar, Almelo, Almere, Alphen aan den Rijn, Amersfoort, Apeldoorn, Arnhem, Breda, Delft, Deventer, Dordrecht, Ede, Eindhoven, Emmen, Enschede, Gouda, Groningen, Haarlem, Haarlemmermeer, Heerlen, Helmond, Hengelo, 's-Hertogenbosch, Hoorn, Leeuwarden, Leiden, Lelystad, Maastricht, Nijmegen, Oss, Roosendaal, Sittard-Geleen, Schiedam, Tilburg, Venlo, Zaanstad, Zoetermeer and Zwolle.
- Rest: 341 Municipalities.

In the following chapter the comparing phase of the benchmarking is conducted. Therefore the following steps are executed: collecting of data, preparing data, analysing data, determining significant different findings and reporting benchmarking results.

5. Benchmarking application: comparing phase

This chapter continues on the theoretical model that is developed in chapter 3 and elaborates on the predefined benchmarking steps. The benchmarking model can be divided into two main phases preparing and comparing. This chapter focusses on the comparing phase of benchmarking and answers research question four: *How can the developed sport policy benchmark be used?* In order to answer this question the benchmarking model has defined five steps and corresponding criteria. In order of appearance the following steps will be conducted. A visualization of the benchmarking comparing phase is depicted in Figure 8.

In step 4 the data that is required for the benchmarking is collected from multiple readily available sources. Following in step 5 the collected data sets are prepared for analysis, meaning normalising and combing of the data sets. Accordingly in step 6 the prepared data is analysed with respect to the defined indicators and dimensions from the preparing phase (step 2). With the analysed data in step 7 the significant different findings are established. Finally in step 8 the findings are reported in a dashboard.

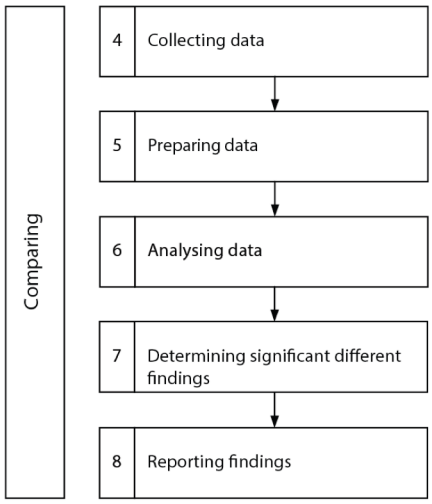


Figure 8 Benchmarking comparing phase

5.1. Step 4 collecting data

Step 4 has one objective, namely collecting data for the indicators. The in the design defined criterion for this step is: *Use multiple data collection methods and reliable data sets that are readily available.* For this objective, the data sets have to be in line with the established indicators and dimensions from step 2. Step 2 defined the indicators output and outcome and their corresponding dimensions presence, proximity, variety and sport participation. Firstly the gathering of data for the output indicator is depicted and secondly the collection of data for the outcome indicator is depicted.

Collecting data for output indicator

The output indicator gives insight in the output of the municipal sport policy. As mentioned before the output are decisions regarding budgets, subsidies, and permits for sport facilities. To measure this the following dimensions are established:

- Presence, number of sport facilities within the municipality. (#)
- Proximity, number of sport facilities in district (km)
- Variety, number of different sport facilities within district (x,y,z)

As mentioned in the introduction the ministry of Health Welfare and Sport has commissioned organisations to monitor certain aspects of sport in order to make the performance of sport policy transparent. One of the mentioned organisations is the Mulier institute. They monitor developments in sports and conduct scientific sport research. Specifically, they were commissioned to gather a national database of all sport facilities in the Netherlands, the Database Sport Provision (DSA). The DSA is a nationwide database with information regarding all sport facilities in the Netherlands. Examples are indoor halls, soccer courts, swimming pools, fitness centres, and ice skate rings. In total the DSA

comprehends over 22.000 sport facilities. Included in the database is the following: place, type sport facility, name, address, municipality, location with x-y-coordinates, and the source. This database is expected to provide comprehensive input for the output-dimensions.

The database is not available in an open data format and in order to obtain the database, permission is needed from the Mulier institute. For this research an export of the DSA is made available with the requisite that the source is mentioned.

The data is collected without difficulties, it is made available after inquiry. The data set provides the data to research the dimensions: presence, proximity and variety. An indication for the reliability and quality of the data is that the Mulier institute states that the DSA is a comprehensive overview of all sport facilities in the Netherlands and it is used as the main source for sport facility related research (van der Poel et al., 2016, p. 13). The database is constructed of provided data from the Mulier institute, municipalities, provinces, sport federations, and sport organisations.

Output indicator (presence, proximity, variety)	
Name	Database Sport Facilities
Source	Mulier institute
Owner	Mulier institute
Data representation	Name, address, x,y-coordinates, municipality, type, data source
Definition	The Database Sport Supply (DSA) is a nationwide database containing data about 22.035 sports facilities, for example sports fields, swimming pools, fitness centres and sport halls.
Explanation	From every sport facility the following information is available: name sport facility; address; x,y-coordinates; municipality; type (what sports can be practiced); indoor and or outdoor location; source; collection method

Table 10 Overview of characteristics of the dataset DSA.

Collecting data for outcome indicator

The outcome indicator is used in order to assess what the effect of the conducted municipal sport policy is. This indicator indicates the level of the effect of the conducted policy. In step 2 the dimension related to the indicator was established in relation with the objective of the municipality. Municipalities conduct sport policy that is mainly focussed on sport facilities. They do this in order to create the desired effects namely, healthy citizens. The dimension that was indicate for this is:

- sport participation, defined as the percentage that achieves the fit-norm. (%)

The reasoning regarding the fit-norm can be found in 4.2 step 2.

The fit-norm is a part of a larger study namely, the health-monitor. The data for this monitor is collected via surveys and interviews. The objective of the health-monitor is to provide an overview of the health, social situation, and lifestyle of the Dutch population aged 19 and over. The results are aimed to be used to research policy at both local and national level (VWS, 2017). The fit-norm is a sport frequency norm for adults and distinguishes based on age. The norm requires the following (Hildebrandt et al., 2008):

- For adults, 19-54 year, at least three times a week, intensive physical activity for at least 20 minutes (≥MET 6,5).
- For elderly, 55 year and older, at least three times a week, intensive physical activity for at least 20 minutes (≥MET 5).

Since the data collection method requires self-assessment of intense physical activity the research of Ainsworth et al., (2000, 2011) is used. They developed the Compendium of Physical Activities for the purpose: "to enhance the comparability of results across studies using self-report physical activity and is used to quantify the energy cost of a wide variety of physical activity with MET-values" (Ainsworth et al., 2011, p. 1).

The MET-value is used as a method to determine how heavy a specific kind of physical activity is. Therefore Ainsworth (2011) has determined what an individual consumes on average in rest (sleep). In rest an individual consumes an average of 3.5 millilitres of oxygen per kilogram of body weight per minute; is 1 MET. If the consumed energy rises to, for example to 7 millilitres of oxygen per kilogram of body weight per minute, there is an energy consumption of 2 METs (Ainsworth et al., 2011). Examples of MET scores are (Ainsworth et al., 2011):

- Game of tennis stands for a MET value of 7.3.
- Running at 15 kilometres per hour stands for a MET value of 12.8.
- Swimming stands for a MET value of 8.0.

A comprehensive list is provided in their article and an updated version on the website of the compendium of physical activity (Ainsworth et al., 2011).

Following the definitions and explanations of the fit-norm and the MET values we zoom in the data gathering process from the health-monitor. The data for the health-monitor is accumulated with surveys and interviews. Since the monitor uses sensitive personal information the results are anonymised. For analysis it is important to have the information related to health on geographical level (Brink, 2017). Due to the limited size of the respondents and privacy the ministry of Health, Welfare and Sport developed a model that calculates the figures at district level. According to ministry of Health, Welfare and Sport the dataset does not have enough respondents (n=387,195) to use weighing methods to calculate figures for all districts in the Netherlands (Brink, 2017).

The method is described by Brink (2017) on volksgezondheidenzorg.info as follows: The participants in the Health Monitor are anonymously linked in a secure environment to CBS registry files. These files contain information about a set of background features, such as age, gender, origin, household composition, income, and type of house. A statistical model has been used to relate health and lifestyle to these background features. Information from neighbouring areas is also included. Through this relationship, it is then possible to calculate their expected health and lifestyle for all adults. The results will then be average over the relevant neighbourhood.
Accordingly, it is significant for this study to be aware that the data are obtained via a model that attempts to approximate the reality as closely as possible, but the figures remain estimates of reality.

The health-monitor is publicly available as open data on municipal level. For this study a lower abstraction level is required namely district level as this level is expected to provide additional insights in the effectiveness of the municipal sport policy. Therefor the responsible researcher dr. C. van den Brink of the ministry of Health, Welfare and Sport was contacted directly to obtain the fit-norm data set on district level.

To sum, the outcome indicator of municipal sport policy has as a dimension sport participation, defined as the percentage that achieves the fit-norm. The required data set is obtained after inquiry. There are

some limitations towards the data set. The first limitation is due to the data collection method since the data is obtained with surveys and interviews and this required self-assessment of the respondents.
The second limitation is due to the data reliability since the data is anonymised and with a calculation model the figures are approximated at district level. Both these limitations can influence the quality and reliability of the data. This research requires readily available data and includes the limitations in the further analysis.

Outcome: Indicator (sport participation)	
Name	Health-monitor
Source	Open data health-monitor license: CC-BY-3.0
Owner	GGD, CBS, RIVM (Client)
Data representation	Data set.
Definition	% that achieves the fit-norm.
Explanation	Data is acquired via surveys (n=387,195) and interviews (n=15,000) under the Dutch citizens that are above 19 years old. The aim of the 'Health Monitor' is to provide an overview of the health, social situation and lifestyle of the Dutch population (aged 19 and over) for research and policy at both local and national level.

Table 11 Overview of characteristics of the dataset health-monitor.

Collecting data for location

The DSA and health-monitor have different structure to define the location. The DSA uses x,y-coordinates for the location of the sport facilities and the health-monitor uses district levels for the percentages that achieve the fit-norm. The objective is to match the DSA x-y coordinates with the district in which it is situated. In order to do that CBS has a database allocated for this kind of tasks the CBS District and neighbourhood shapes (in Dutch: Wijk- en Buurtkaart 2012). This dataset provides data in shape file format that corresponds with x-y coordinates.
The data for the location match is publicly available data. Making the access of the data straightforward. The data is expected to be reliable high quality data.

Location (district shapes)	
Name	Districts and neighbourhoods map 2012
Source	CBS Wijk- en buurtkaart 2012
Owner	CBS and Kadaster
Data representation	Data set in ESRI Shape-format
Definition	The districts and Neighbourhood Map 2012 includes a digital representation of the geometry of the boundaries of the districts, neighbourhoods and municipalities.
Explanation	The district and neighbourhood map is made up of three sources. The first source is used for the municipal boundaries and come from the Basic Registration Kadaster (BRK); The district boundaries come from the CBS database: Land use.

Table 12 Overview of characteristics of the dataset for matching the DSA x-y coordinates to district level.

Objectives for preparing data

Combining the datasets DSA and health-monitor requires three additional efforts.
The first effort is applicable to the DSA, because DSA and health-monitor use different structures for the location. Namely the DSA uses x,y-coordinates for the location of the sport facilities and the health-monitor uses district levels for the percentages that achieve the fit-norm. Since the dimensions of the output indicator require measurements on district level the location terms of the DSA need to be transformed from the current x,y-coordinates to the district in which the sport facility is located.

The second effort is only applicable for the DSA. Since the fit-norm can only be achieved by intense sport activity. Meaning that not all sport facilities assist in meeting the requirement. Only the sport facilities that enable physical activity levels of ≥MET 6,5 or ≥MET 5 are taken in to account. For example the sport bowling has a MET score of 3 (Ainsworth et al., 2011) and therefore it should be filtered out of the data set.

Hence the assignments for the preparing data for analysis are:

- Assignment A: Filter out the sport facilities that do not assist in achieving the fit-norm i.e. (≥MET 6 and ≥MET 5).
- Assignment B: Make the terms and structure consistent for the location terms in the DSA.
- Assignment C: Combine DSA and the health-monitor into one dataset.

5.2. Step 5 data preparing

The criteria for step 5 as defined in the design: *Combing and normalising data sets. Show what is done.* This step makes the data that is collected in step 4 ready for the analysis of the data. In this step the data sets need to be combined and normalised. To be more specific this step is tasked with three assignments A, B and C, visualised in Figure 9. The first assignment A: filter out the sport facilities that do not assist in achieving the fit-norm i.e. ≥MET 6 and ≥MET 5. This results in a filtered database. Following the second assignment B is make the terms and structure consistent for the location terms in the DSA, resulting in the DSA MET with district level codes. Finally the third assignment C: Combine DSA and the health-monitor into one dataset. Here the fit-norm scores from the health-monitor are added to the filtered dataset DSA MET with district level codes.

In the above described sequence the assignments will be conducted and elaborated below.

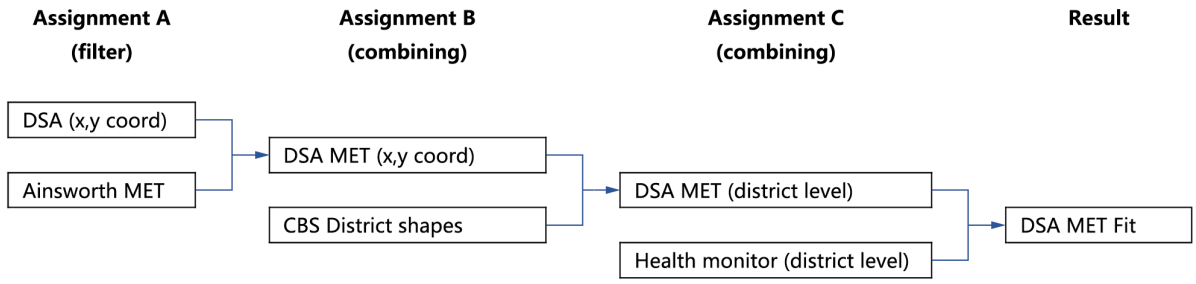


Figure 9 Visualisation of the of the assignments A, B, and C. First filter out sport facilities that do not assist in achieving the fit-norm. Second make terms and structure consistent for location terms DSA. Finally combine the data sets health-monitor and DSA.

The first assignment A is filter out sport facilities that do not assist in achieving the fit-norm i.e. ≥MET 6 and ≥MET 5. The data sets that are needed for this are the 'DSA' from the Mulier institute and the 'MET scores' from Ainsworth (2011).

In the appendix 9.1 the constraints for the filtering are depicted. Firstly all the type of sport that are in the DSA were identified. Secondly for each type sport the corresponding code number and MET value from Ainsworth (2011) is indicated. Not for all sport facilities types a specific MET score could be found. In these cases a score is calculated based on the activities that could be performed on a specific sport facility, the calculations are depicted in appendix 9.1

Based on the MET values the sport types that do not meet the required MET score of 5 or 6,5 are filtered out resulting in two new data sets 'DSA MET 5' and 'DSA MET 6,5'.

Following the filtered database 'DSA MET' the second assignment B is to make the terms and structure consistent for location terms DSA. This is needed since the 'DSA MET' uses x-y coordinates while the districts in which the sport facilities are located are needed for the analysis. The database 'CBS district shapes' can provide in this combining. The dataset provides shapes that corresponds with x-y coordinates.
A workflow is constructed combining the databases 'DSA MET' and 'CBS District shapes'. The workflow is constructed to find the corresponding district to each sport facility, depicted in Figure 10. The result of the work flow is the 'DSA MET' including the corresponding district in which it is situated.

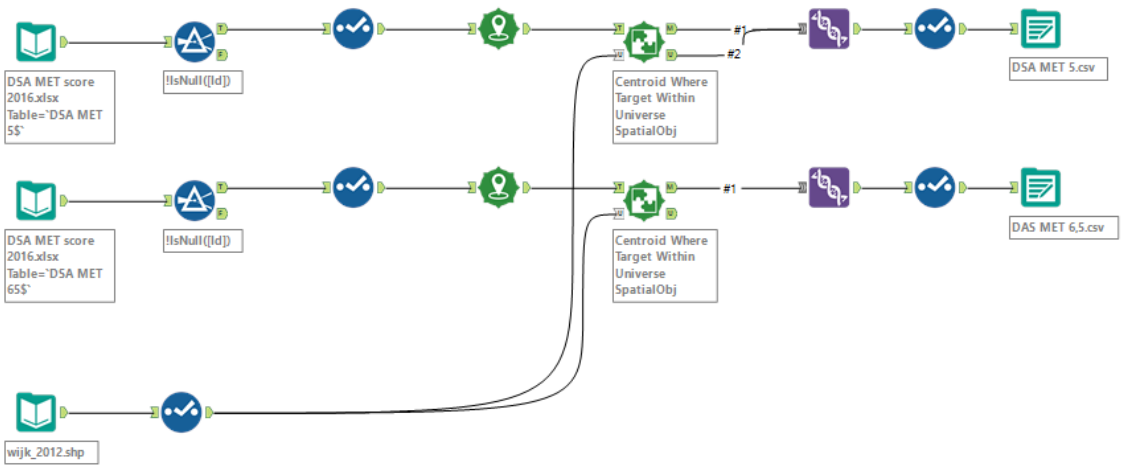


Figure 10 Alteryx workflow combining data set DSA (containing all sport facilities in the Netherlands) with data set CBS district codes (containing district shapes).

Finally after assignment B the last data preparing assignment is assignment C: Combine 'DSA MET' and the 'health-monitor' into one dataset. This is a simple process since the 'DSA MET' and the 'health-monitor' now both use district codes for the location. A match work flow is constructed between the two datasets. This results in an added column for each sport facility in the 'DSA MET' with the fit-norm score of the corresponding district, the dataset is called 'DSA MET fit'.

To summarize the description for step 5 as defined in the design is the following: *Combing and normalising data sets. Show what is done.* Step 5 has prepared the data, accordingly three assignments were performed. The assignments consisted of normalising, filtering and combing data. Step 5 has resulted in one normalised, filtered and combined data set: 'DSA MET fit' consisting of a filter for sport facilities that do not assist in meeting the fit-norm, an additional column for the districts in which the sport facilities are situated and an additional column with the districts fit-norm score.

5.3. Step 6 data analysing

The description for step 6 as described in the design of the benchmarking method is: *Examine the data with regard to the indicators and the dimensions.* In this step the data is analysed in order to compute the values for the dimensions. The in step 5 constructed dataset 'DSA MET fit' can be used for these computations. The specific dimensions for which the values need to be computed are determined in step 2 and are the following:

Dimensions for the indicator output:

- Presence, number of sport facilities within the municipality. (#)
- Proximity, number of sport facilities in district (km)
- Variety, number of different sport facilities within district (x,y,z)

Dimension for the indicator outcome

- Sport participation; percentage that achieves the fit-norm.

In the dataset 'DSA MET fit' all the dimensions are calculated. To give a better understanding of the process, a dataflow diagram is made Figure 11 and a visualisation from the actions that are taken after step 4 data collection. The visualisation is an example of the calculations and uses the city Delft and the district Voorhof of illustrating the process of data analysing. This is depicted in Figure 12. On the left the process step is visualised and on the right a short description of the process is depicted.

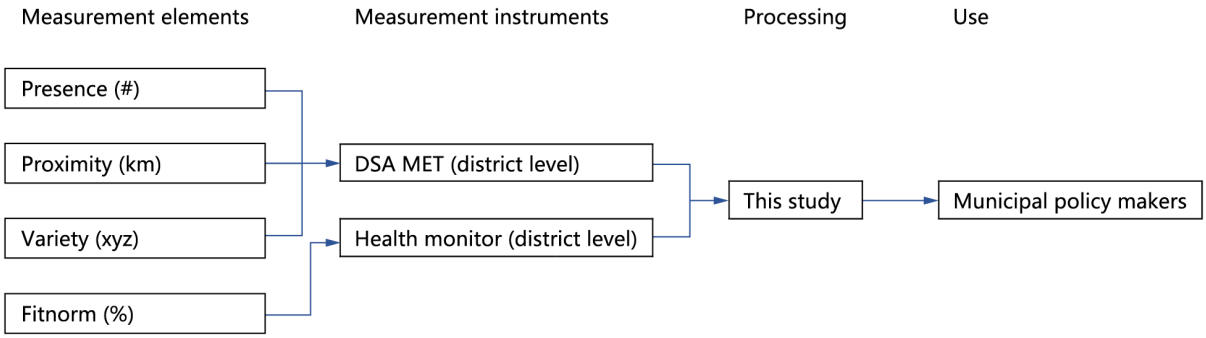


Figure 11 Dataflow diagram for this study.

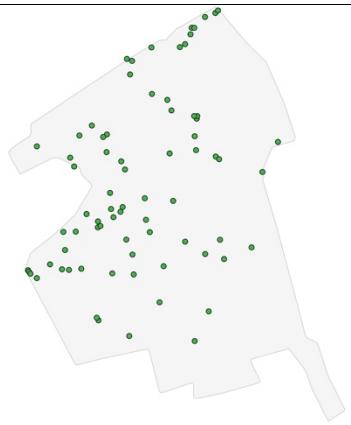
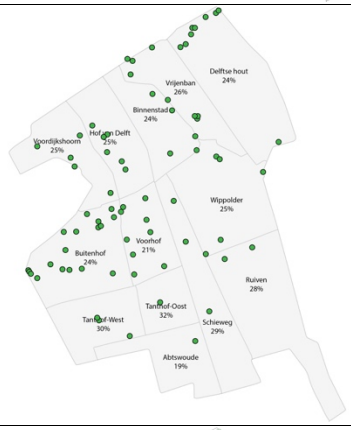
Process step	Description process
	**Step 5.1. data preparing: filter** Find type sport facilities that assist to achieving the fit-norm, with the help of the MET score ≥5.5 or ≥6.5 (Ainsworth et al., 2011). For the sport facilities that do not have a specific MET score, this score is calculated as an average of the activities that are performed on the sport facility.
	**Step 5.2. – data preparing: match** Find district code based on x-y coordinates. In the picture the fit-norm scores are already depicted for each district.
	**Step 5.3. – data preparing: combine** Match fit-norm score to the district code of the sport facilities.
	**Step 6 data analysing** Compute the values for the dimensions Results Municipality Delft: • Sport participation, fit-norm = 25,3,9% • Presence, sport facilities MET ≥ 5 = 83 • Presence, sport facilities MET ≥ 6.5 = 77 • Municipal variety, types sport facilities =17 Results District Voorhof within Delft: • Sport participation; fit-norm = 21% • Proximity, sport facilities MET≥ 5 = 9 • Proximity, sport facilities MET≥ 6.5 = 7 • District Variety, types sport facilities = 2

Figure 12 Visualisation of the benchmarking step 4-6 illustrated for the city Delft.

5.4. Step 7 determining significant different findings

The criteria and description for step 7 as described in the design of the benchmarking method are: *Determine the gaps between individual indicators and the measures. Compare the individual scores to the collective predefined score*. As stated in the problem statement, there is a lack of insight in whether the output of the sport policy (namely the quantity, type, and location of sport facilities) leads to the desired effects (namely healthy citizens), which is defined as a lack of insight in the effectiveness of the sport policy. The individual indicators are the output and outcome and consist of dimensions. Output is composed of the dimensions: presence, proximity, and variety and outcome is made of the dimension sport participation. This is a representation for the collective predefined score.

In this step the significant different findings are determined. To accomplish this first the gaps are determined. This is the process of ascertaining the differences between the scores of the dimensions. In step 6 the scores were calculated, following in this step the scores are compared to the average of the measurement. If the comparison shows large gaps, then it is investigated whether the gaps can be explained, i.e. placing the gaps in context and finding causes for the high or low scores.
Below first the identification of gaps is depicted followed by the correlation calculations between the dimensions.

Identification of gaps

For the groups G32 and G4 the following large differences in the number for sport facilities a municipality has could be identified. These identified gaps need to be place in context as stated by de Bruijn (2007) in criteria b3 and by Bouckaert (2011, p. 121) often only the focus is on the quantitative aspects and reflection is needed. By placing the gaps in context possible causes for the high or low scores could be found. As stated in step 2: based on the reviewed literature it is expected that higher numbers for presence, proximity, and variety positively influence the sport participation.

Measures of extremes, for the G32

Within the G32 there are considerable differences in the number of sport facilities in each municipality. The differences and the other results are depicted in Table 13. The top 10% scoring municipalities have around 170 sport facilities in their municipality (i.e. Apeldoorn, Arnhem, Eindhoven, Emmen, Groningen) and the bottom 10% scoring municipalities have around 60 sport facilities in their municipality (i.e. Almelo, Helmond, and Lelystad). In order to find possible effects of the large differences the corresponding sport participation scores are studied, this did not provide additional insight as the sport participation appears not to be affected by the presence of sport facilities in the municipality.
To explain the difference it is expected that the size of the municipality could be an critical factor. Since various scholars have pointed out the positive relations between the number of sport facilities and the sport participation and the variety of sport facility types and sport participation, it is expected that a larger number of sport facilities per 1000 citizens positively influences the sport participation. Although the reviewed literature pointed out the positive relations these connections are not supported by the data of this study. The presence and variety of the sport facilities per 1000 could also not explain the large differences.

	Almelo	Apeldoorn	Arnhem	Eindhoven	Emmen	Groningen	Helmond	Lelystad
Sport participation (%)	24,9	24,9	24,9	21,4	21,4	24,5	23,0	18,3
Presence MET5 (#)	57	160	160	178	178	170	60	62
Presence MET5 per 1000 citizen (#)	0,78	1,02	1,07	0,82	1,64	0,88	0,68	0,82
Variety MET5 per 1000 citizen (xyz)	0,19	0,14	0,15	0,11	0,22	0,10	0,16	0,21

Table 13 Overview of the results of municipalities in group G32. The top 10% and bottom 10% municipalities from the group G32 are depicted based on the amount of sport facilities in the municipality.

Measures of extremes, for the G4

Between the G4 there are large differences between the dimension presence of sport facilities, based on the reviewed sport literature it is expected that this will influence the outcome dimension: sport participation.

The municipality Amsterdam has 500 sport facilities and a sport participation of 19,9%. While Utrecht has 218 sport facilities and a sport participation of 19,2%. The large difference in number of sport facilities appears not to affect the sport participation. The difference can be explained by the large size difference, Amsterdam had in the time of the measurements 790.110 citizens and Utrecht 316.275 citizens. Consequently the municipalities have similar presence scores per 1000 citizens namely Amsterdam 0,63 and Utrecht 0,63.

Municipality	Amsterdam	The Hague	Rotterdam	Utrecht
Sport participation (%)	19,9	21,2	19,1	19,2
Presence MET5 (#)	500	323	442	218
Presence MET5 per 1000 citizen (#)	0,63	0,65	0,72	0,69
Variety MET5 per 1000 citizen (xyz)	0,04	0,05	0,04	0,08

Table 14 Overview of the results of the group G4.

Correlation

In the previous paragraph two gaps were identified. To gain additional insights in dataset 'DSA MET fit' two types of correlation calculations are performed. The first is on district level the correlation between sport participation and all individual aspects. The second is between the sport participation and the dimensions.

Correlation expresses the mutual relationship between variables, the way two random variables influence each other (Dekking, Kraaikamp, Lopuhaä, & Meester, 2005, p. 135). Correlation is defined as follows: Let X and Y be two random variables. The correlation coefficient ρ(X,Y) is defined to be 0 if Var(X) = 0 or Var(Y) = 0, (Dekking et al., 2005, p. 142).

- $\rho(X,Y) = \frac{Cov(X,Y0}{\sqrt{Var(X)Var(Y)}}$

In order for two dimensions to be correlated, a linear relationship should be found between the two dimensions. The following values and descriptions are used to determine the linear relationship (Dekking et al., 2005, p. 142):

- ρ(X,Y) = +1 Perfect linear relationship
- ρ(X,Y) = 0.70 A strong linear relationship
- ρ(X,Y) = 0.50 A moderate linear relationship

- ρ(X,Y) = 0.30 A weak linear relationship
- ρ(X,Y) = 0 No linear relationship

Note ρ(X,Y) could be either positive resulting in a positive relationship or negative resulting in a negative relationship.

Correlation between sport participation and individual aspects

To gain more insight in sport participation levels on district level it is investigated whether individual aspects correlate with sport participation levels. Based on the reviewed data and the correlation values no strong correlations were found between sport participation and all individual aspects of the dataset 'DSA MET fit'. For the scores see Table 19 page 71.

Individual aspects that are taken into account are aspects related to number of inhabitants, age, gender, ethnicity, income. In the appendix 9.2 the results of the correlation between sport participation and the individual aspects are depicted. A small negative correlation, -0,49, between lower income and sport participation was found. Citizen with low disposable income tend to not achieve the fit-norm. Possible causes for this weak relationship could not be found with the used dimensions. The correlation between the presence and low disposable income is, +0,12 and the correlation between the variety and low disposable income is, -0,01.

Correlation between sport participation and dimensions

Below in Table 15 the results of the correlation calculations between sport participation and the dimensions are depicted. The results are given for the groups G4 and G32.

With the correlation scores for the municipalities in the G32 no significant correlation could be established. For the municipalities in the G4 a relatively strong negative correlation is found between the dimension sport participation and number of sport facilities per 1000 citizens, -0,67. The association between the two was also made in the 'measures of extremes, for the G4'. In words there is a negative correlation between sport participation and the number of sport facilities per 1000 citizens in the G4. A reason for the negative correlation could come from the small size of the group, i.e. four namely: Amsterdam, The Hague, Utrecht, and Rotterdam.

Correlation		**Dimensions**
G4	**G32**	
-0,03	+0,05	Sport participation and presence MET 5
-0,05	+0,04	Sport participation and presence MET 65
+0,05	-0,05	Sport participation and variety
-0,27	-0,25	Sport participation and proximity district MET 5
-0,30	-0,26	Sport participation and proximity district MET 65
-0,41	-0,22	Sport participation and variety district MET 5
-0,44	-0,11	Sport participation and variety district MET 65
-0,67	+0,12	Sport participation and # per 1000
-0,21	+0,06	Sport participation and xyz per 1000

Table 15 Results of the correlation calculation between the dimensions for the groups G4 and G32.

Correlation between sport participation and dimensions

In the Netherlands municipalities provide a network of sport facilities (in presence, variety and proximity) in order to create healthy citizens (sport participation, defined as citizens who achieve the fit-norm). The reviewed sport literature reference the interdependence of sport participation and the

presence, *proximity* and *variety* of sport facilities (Hoekman et al., 2016, p. 2). The findings in this study give no indication that the sport facilities and the sport participation are correlated.

5.5. Step 8 report findings

Step 8 is the final step of the benchmarking of municipal sport policy. As described in the design of the benchmarking method the criterion and description for step 8 is: *report the findings as an overview in a dashboard.* The description of the step can be broken into two parts the first part is report the findings and the second part is present them in a dashboard. The first part reports the findings that are determined in the previous steps and the second part elaborates on the presentation of the findings in a dashboard.

A dashboard is used to present the findings of the benchmark. Maheshwari (2013) defines a dashboard as follows: a dashboard a visualization of benchmarking information given a certain objective and provide interpretation and improvement support (Maheshwari, 2013, p. S89). The benchmarking information are the findings that are determined in the previous steps. The objective is giving insight in the effectiveness of the municipal sport policy. Providing interpretation and improvement support is achieved by giving graphically an overview of the findings, i.e. an overview of the indicators and dimensions. This enables the municipal policy-makers to assess the effectiveness personally. This is important since the findings can be viewed from multiple perspectives and can therefore have more than one meaning.

The dashboard acts as a communication instrument of the benchmark municipal sport policy. It is directed at municipal policy-makers, since they create the sport policy and domain knowledge is needed to make sense of the benchmarking results. In order to present the findings to the municipal policy-makers choices have to be made for the constructs of the dashboard. In the next paragraphs the considerations for the constructs of the dashboard are elaborated.

In the previous steps the results for the dimensions are calculated, subsequently the gaps are identified and correlations are investigated. The determined gaps in presence and variety that were identified could be explained by taking out the size differences between the municipalities. Accordingly for the dimensions presence and variety also the weighted score of the dimension is depicted. In other words, the dimension is put in perspective by giving both the total number of the dimension as the number per 1000 citizens. As stated in step 7 no significant correlations were found between the dimensions. Accordingly for the dashboard no combination of the indicators are depicted and all the calculated dimensions are depicted.

The objective of this dashboard is to depict the results in order to let the municipal policy-makers assess the effectiveness of the municipal sport policy themselves. This is important since the assessment depends on the objectives, budgets, and ambitions of each municipality. To assist in the assessment of the effectiveness a special segment is allocated for this. The effectiveness of the municipal sport policy was defined as the extent to which the policy performance, the output, leads to the envisioned effects, the outcome (Hoogerwerf & Herweijer, 2014). However the data does not support this relation, that the output and outcome of the municipal sport policy are correlated. In order to assist the assessment an

objective calculated value for effectiveness is depicted and is defined as the ratio between actual outcome and theoretical outcome i.e. fit-norm score Delft / fit-norm score average group.

Apart from the effectiveness segment the dashboard is constructed of a mapping and an overview of the results of the dimensions, see Figure 13. The mapping is made from the chosen municipality in this case Delft and depicts the corresponding districts, sport facilities, and the fit-norm scores of the municipality. The overview of the results gives insights in the results of the municipality and the average of the group, in this case Delft and the average of the G32. The results consist of number of citizens, the effectiveness, the ranking, and the results for the dimensions. The effectiveness is defined as the ratio between the actual outcome and the average outcome. The ranking, represents the hierarchical position within the G32 based on the score of effectiveness. The dimensions, represent the calculated score for the presence, proximity, variety, and sport participation.

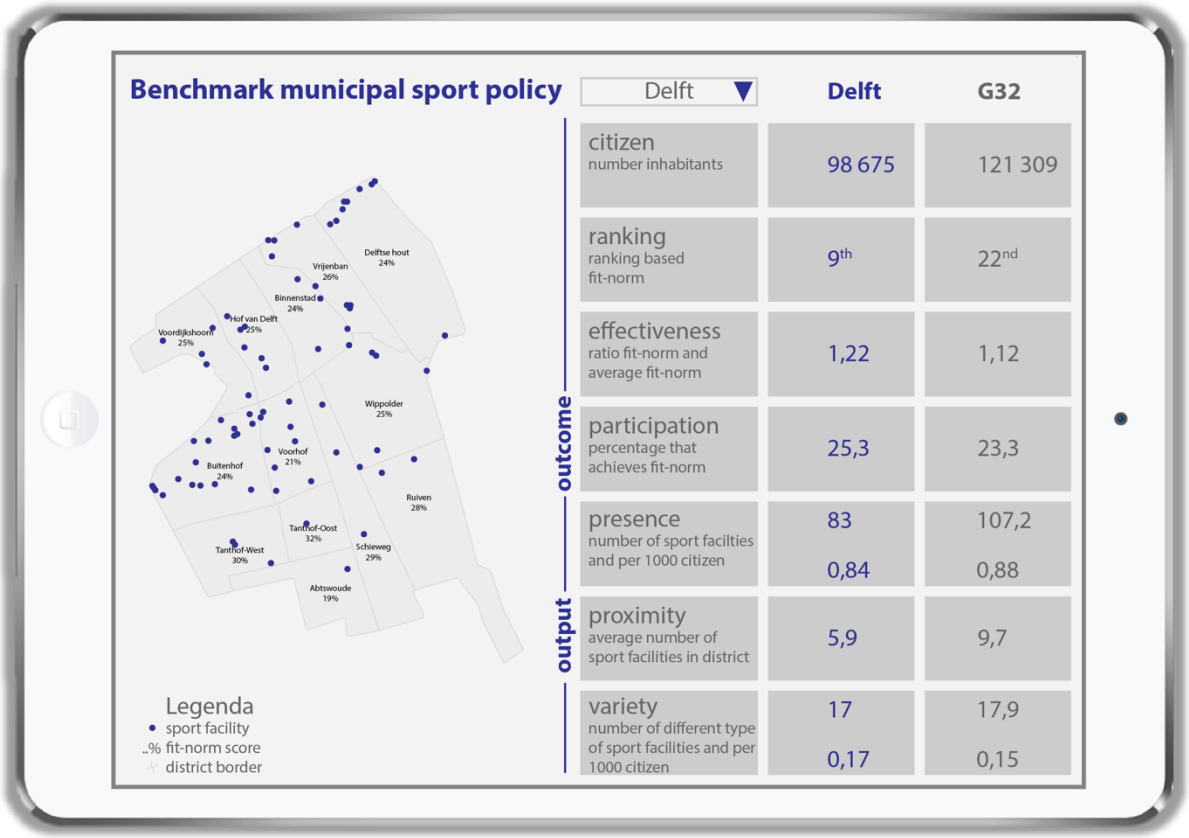


Figure 13 Dashboard for the benchmark municipal sport policy. The dashboard depicts the municipal sport policy results of from the municipality Delft based on the analysis of the dataset 'DSA MET fit'.

6. Evaluation

This chapter addresses the evaluation of the designed benchmarking method and the application of the benchmarking study. This chapter is directed to answer research question 5: *What are the effects of the developed municipal sport policy benchmark?* The prefacing research questions 1-4 are aimed at developing, designing, and executing the municipal sport policy benchmark. This chapter and the corresponding research question five is aimed at evaluating the designed and applied benchmarking method.

Here it is assessed whether the conducted study provides the beneficial effects and whether the benchmarking model is designed in such a way that it minimizes perverse effects. First the method is evaluated, here it is assessed which perverse effects apply to the method. Next it is evaluated if the conducted study improves the current situation, meaning that with the benchmarking study more insight in the effectiveness of the sport policy is obtained. To assess this the benchmarking findings and benchmarking process are evaluated.

6.1. Evaluation benchmarking perverse effects

In this paragraph the developed benchmarking method is evaluated. The design of the developed method is evaluated by assessing whether the method leads to the perverse effects as established by de Bruijn (2002). In other words whether the designed and applied benchmarking method consisting of phases, steps, and criteria minimizes the perverse effects as established by the Bruijn (2002).

Below the perverse effects of de Bruijn (2002) are enumerated and accompanied with an explanation whether the perverse effect apply or not.

Perverse effect 1 *Performance measurement is an encouragement for strategic behaviour.* The first perverse effect focusses on strategic behaviour in other words 'gaming the numbers', meaning increasing the output in accordance with the indicators and dimensions of the benchmark.

The benchmark municipal sport policy is developed in such a way that it measures the effectiveness of the municipal sport policy. Effectiveness is the extent to which the policy output leads to the envisioned effects, the outcome. The corresponding effectiveness indicators are output and outcome. This study found no relation between the two indicators and therefore redefined effectiveness to the ratio between the own outcome score and the outcome score of the average of the group. The results of the outcome indicator cannot be influenced by the municipality rendering strategic behaviour ineffective, since only the output can be influenced and the effectiveness is to a large extent depended on the outcome.

Perverse effect 2: *Performance measurement blocks innovations.* An organisation will focus on output that is measured and choses for the save option meaning output that is well-known and easy to make.

This perverse effect can be applicable for the design method of this benchmark, municipal sport policy concern two categories: sport facilities and sport participation programs. For the output indicator dimensions are chosen that solely focus on the first category namely sports facilities. Therefore municipalities could intensify their output even more on sport facilities since this is measured.

Perverse effect 3: *Performance measurement blocks ambitions.* A selection criteria for the input can result in strategically selecting the inputs in order to achieve the desired output with minimal effort. This

perverse effect of 'cherry picking' is not applicable for the design of this benchmark since it does not take input as an indicator into account.
Perverse effect 4: *Performance measurement veils actual performance.* This perverse effect can happen because during the measurement process, causal connections can be lost which give meaning to the results. Therefore the measurement results need to be placed in context by the municipal policy-makers in order to assess the performance.
In the design of the benchmark special attention is given towards the measurement process and depicting of the results. The results are graphically described in a dashboard encouraging the municipal policy makers to put the results in context and assess their performance.
One of the criteria for the measurement process was to take data availability into account. This criterion led to data sets that were cultivated and cleansed. This can result in poor assessment of the output and outcome since the quality and reliability of the data cannot be guaranteed.

Perverse effect 5: *Performance measurement drives out the professional approach: no system responsibility, no quality, more bureaucracy.* The output of an (public) organisation is consistently a trade-off between different values. The performance measurement focusses on measurable output and when only one value is measured the organization can concentrate on this value and relinquish the other values.
This perverse effect is applicable for the design of this benchmark the reasoning is the same as for perverse effect 2. The output indicator only focuses on one aspect of the municipal sport policy, i.e. sport facilities. When only the relation between sport facilities and sport participation is measured, possibly the sport participation programs may suffer and therefor the municipality will not comply to its system responsibility.

Perverse effect 6: *Performance measurement leads to copying, not learning.* Benchmarking may degenerate into duplicating best practices. This is risky because by measuring the performance the complexity is reduced resulting in a poor copy of reality. Consequently, duplicating best practices is difficult since each organization is different.
In the description of the results of the benchmark no best practices are indicated. Solely the results of the measurements are depicted in comparison with the group average. Therefore this perverse effect is not applicable for the design of this benchmark.

Perverse effect 7: *Performance measurement punishes good performance.* Due to the created transparency and good performance an organisation can receive higher targets.
The assessment whether performance is good or bad is not done by this study. The results are presented in a dashboard allowing the municipal policy makers to assess the performance themselves. Therefore this perverse effect is not applicable for the design of this benchmark.

Based on the evaluation of the 7 perverse effects of de Bruijn (2007), it was identified that three perverse effects occur and apply for the design of the benchmark of municipal sport policy. Namely the perverse effects 2, 4, and 5.
Two causes can be indicated for the possible perverting of the benchmarking. The first cause is due to the decision to focus the policy output indicator on one of the two categories that concern municipal sport policy. This can possibly lead to the second perverse effect, blocking innovation and the fifth perverse effect, driving out system responsibility. The second cause is due to the criterion 'use readily

available data'. Since the reliability and the data quality cannot be guaranteed possibly the actual performance is not the same as the measured performance, perverse effect four. Both causes are related to the premise of benchmarking and that is that benchmarking is restricted to the parts that are measureable. By adding the criterion of 'use readily available data' restricts the area that can be benchmarked even more.
Finding a new more criterion for the indicators that is more comprehensive in capturing the whole system is not an easy task since the indicators are restricted to what is measurable, the premise of benchmarking. The criterion use of readily available data is alleviated easily however imposes additional efforts in measuring and gathering data.

6.2. Evaluation application benchmarking

This paragraph gives insight if the conducted benchmarking study improves the current situation, meaning that by conducting the benchmarking study more insight in the effectiveness of the sport policy is obtained. In the previous paragraph the design of the benchmarking method is evaluated and this paragraph is directed at evaluating the application of the benchmarking, therefore the benchmarking findings and the benchmarking process are evaluated.

Benchmarking findings
Evaluating the benchmarking findings has led to the following points of interest:
On average a district in the Netherlands has 6,98 sport facilities that assists in achieving the fit-norm score. To put this in perspective this are on average 0,98 sport facilities per 1000 citizens. In this calculation the sport facilities that do not assist in exercising on high intensity physical activity levels are filtered out from the equation. This are the sports that have a MET score below 5 for example baseball, horseback riding, and golf have to low sport intensity scores. In each district a citizen has on average the choice between 4,04 different types of sport facilities. Per 1000 citizens the choice variety is 0,57. On average 23,83% of the citizen in the Netherlands achieves the fit-norm, a sport frequency norm that is set by the ministry of Health, Welfare and Sport. These results can be interesting for sport policy related research and for municipal policy-makers that are tasked with the municipal sport policy. Without context these findings are 'nice to know' but when put in context it can enable comparison and subsequently possibly provide insights. The context can provide additional insight regarding the meaning for certain scores, for example the municipality's sport ambition and its available budgets for sport policy or the number of inhabitants and sociodemographic characteristics.

This study has calculated for all municipalities in the Netherlands their municipal sport policy output and outcome. This study focusses on sport facilities as the municipal sport policy output with the corresponding dimensions: presence, variety, and proximity. For the outcome dimension this study uses sport participation defined as the percentage that achieves the fit-norm. Subsequently the correlation between all these dimensions have been calculated. Between the aforementioned dimensions no correlation was found. This was unexpected since various studies have pointed out a scientifically proven relationship between the dimensions. An overview of the scientific relations between sport participation and the dimensions is depicted in Table 9 on page 34.

No correlation was found between the number, variety and proximity of sport facilities and the achievement of the fit-norm. Additional research is done in whether certain sociodemographic factors

are circumstance that influence the achievement of the fit-norm. The research in the influence of sociodemographic factors yielded no significant correlation. The lack of correlation found between output and outcome in this study can be explained by practical explanations and by data related explanations.

There are two practical explanations that could explain the lack of correlation between the indicators output and outcome. It is possible that the Netherlands has a comprehensive network of sport facilities and that travel distance is not an issue for Dutch citizens. A second explanation could be from the indicator choice. For the output indicator, this research only takes specific designed sport facilities into account. However high intensity physical activity is also possible in non-specific designed sport facilities. The demarcation of this indicator fails to take non-specific designed sport facilities into consideration examples of these are the roads, parks, and the beaches.

Also data related explanations could explain the lack of correlation between the dimensions. The first data related explanation is due to data reliability and the second due to data preparing. The reliability of the data form the health-monitor is difficult to ascertain due to two limitations. The first limitation for the health-monitor is due to the data collection method, the data is obtained with surveys and interviews this way of data collecting requires self-assessment of the respondents. The second limitation is due to the data reliability of the health-monitor. The degree to which the measurements form the health-monitor can be depended on to be accurate is declined since the data is anonymised and with a calculation model the figures are approximated at district level. In these efforts, possibly causal relations are lost. Both the aforementioned limitations can influence the quality and reliability of the data from the health-monitor.
The second data related explanation can come from the data preparing efforts for this study. Due to the data structure differences between the DSA and the Health monitor three data preparing assignments needed to be performed in order to analyse the data. The three data preparing assignments were related to filtering, matching and combining of the data sets. Although these tasks were conducted with caution and precision it is possible that due wrong matches and combinations were made. This can be the result from the use of available data and therefore the quality and time stamps of the data sets is difficult to assess. Therefore based on the quality of the data and the data preparing efforts it is possible that these data related issues could have influenced the correlation between the dimensions.

Benchmarking process
Evaluating the benchmarking process led to two points of concern, namely forming benchmarking groups and data analytics required for measuring and comparing.
The formation of benchmarking groups that consist of comparable organisations has proven to be difficult. The reviewed literature provided no clear guidelines for making benchmarking groups. The only aspects that were found in both benchmarking literature and sport literature are related to municipal size. Although the benchmarking was conducted between the same kind of organisations (municipalities) still great differences between the organisations can be perceived. Examples of these differences are related to a municipality's: size, objectives, ambitions, inhabitants, and budget. For this specific study the process step for the formation of groups came to early in the process, namely process step 3. For this study data needed to be collected and analysed in order to determine the benchmarking groups. An adaptation in process step 3 could overcome this by amending this step to: define criteria for comparable organisations.

By assessing this study's benchmarking process, it appears that data analytics is a vital part of benchmarking especially in the benchmarking comparing phase. This is illustrated by the following benchmarking process steps: collecting data, data preparing, and data analysing. One of the criteria for the benchmarking method stated: *Give insight in how the performance is measured* (b3). Maheshwari (2015) does not give specific reference to this he only states: provide clear measurement guidelines (m5). However Janssen (2010) (j2) states that the more dimensions are used the more resources are needed and the more difficult the comparison is. Although this study only used four dimensions this statement from Janssen is found applicable. Because this benchmarking study already stated that data analytics skills are required in benchmarking when more dimensions are used more data is needed making data analytics even more complex. In order to conduct the benchmark multiple data sets are merged into one large data set, in this process of filtering, matching and combing it easy to lose oversight.

6.3. Conclusions evaluation

The assessment of the benchmarking design, the benchmarking findings, and benchmarking process is used to answer research question 5**:** *What are the effects of the developed municipal sport policy benchmark?* Below conclusions of the evaluation of benchmarking design, benchmarking findings, and benchmarking process are described.

The design of the developed method is evaluated. This is done by assessing whether the designed method minimizes the perverse effects of a performance measurement system. It was found that the design can pervert the benchmarking study. The first perverse effects found are the following: the design can block innovation and drive out system responsibility. Indicated as the cause for these effects is that the output dimensions focus on a part of the policy output. Therefore not the whole policy output is taken into account leading to the aforementioned effects.
The second perverse effect is that benchmarking can veil the actual performance, indicated as the cause was the criterion 'use readily available data'. This criterion enabled easily accessible data however it also led to the use of a data set consisting of data for which the reliability and quality was difficult to assess namely the health-monitor. For future research this criterion is easily alleviated, if alleviated it imposes additional efforts in measuring and gathering data possibly leading to new issues.

Evaluating the findings of the benchmarking study, led to the following beneficial effects. The conducted benchmarking study improves the current situation for municipal policy-makers. The first beneficial effect of this study is because this study designed a benchmarking method that can be used to measure and compare the municipal sport policy. The general method can also be used for other purposes for example benchmarking municipal real estate. The second beneficial effect of this study is that it made the presence, variety, and proximity of sport facilities in municipalities transparent. This information enables the municipal policy-makers to learn by comparing their numbers with other municipalities.
With the used data and the selected dimensions this study found no correlation between the output and outcome of the sport policy. Indicated as possible causes for the lack of correlation are the quality of the data and the data preparing efforts.

Evaluating the benchmarking process found the following two points of interest. The forming benchmarking groups proved to be challenging for this study. Because the literature provided few tools to form comparable groups. Based on the reviewed literature the only indicator that was found to form groups is municipal size. It was found that the step 'forming groups' came early in the benchmarking process. This was challenging because due to the large size of the benchmarking group (383 municipalities) data was required to form groups

The second point that was found in the evaluation of the benchmarking process is that *data analytics* is required for the measuring and comparing in a benchmarking study. The data analytics tasks that are performed can influence the results of the benchmarking and therefore should be executed with caution and by experts.

7. Conclusions and reflection

This study focusses on benchmarking municipal sport policy. Physical activity is used as an important tool to create healthy citizen. Therefore each municipality in the Netherlands conducts municipal sport policy. Decades of conducted municipal sport policy has resulted in a network of sport facilities, i.e. a number of sport facilities to play soccer, tennis, swimming or other sports. Currently there are trends asking for a new position of physical activity in the city: aging of society, population in- and decline, and change in the way we exercise (van der Poel et al., 2016, p. 16). Due to these trends sport policy and its effectiveness are being reconsidered and debated, however this evaluation of sport policy is complex. Municipalities lack insight in the effectiveness of sport policy (van der Poel et al., 2016, p. 253). The lack of insight is whether the output of the sport policy (namely the quantity, type, and location of sport facilities) leads to the desired effects (namely healthy citizens), which we define as a lack of insight in the effectiveness of the sport policy. This insight in the effectiveness of a sport policy is crucial for municipalities, since it can be used in the policy evaluation step to assess the performance of the policy. To increase the impact of the policy evaluation a benchmark of municipal sport policy designed and conducted. This chapter is aimed to specify the benchmarking findings and the recommendations regarding the benchmarking process and the importance of reliable data.

7.1. Benchmarking findings

Based on literature review and addressed to this study's scope first benchmarking is defined as: measuring some elements and comparing the results with comparable organisations. The reason why organisations use benchmarking is because they want to know their performance and improve it. Benchmarking uses monitoring and evaluation information to compare results with the aim of creating transparency and enable learning. This commonly occurs during the performance evaluation step as part of the policy cycle, by comparing output on different levels.

The process of benchmarking is not straightforward. In the identification of the benchmarking challenges, it became apparent that the design of benchmarking is crucial. A wrongly designed benchmark could render it ineffective. Indicated as leading cause was the lack of theoretical foundation for the benchmarking.

The main research question is:

> *How to design a sport policy benchmark that increases the effectiveness of a sport policy in municipalities in the Netherlands?*

To arrive to the answer of this research question the following research question are answered.

Research question 1: *What are the process steps and requirements for designing a benchmark?*
In the problem description it was stated that the leading problem with benchmarking is the lack of theoretical foundation, this is relevant since wrong design can lead to wrong conclusions and perverse effects. Accordingly it became apparent that specific attention needed to be paid to the design of the benchmark. Therefor first the benchmarking criteria for designing a benchmark are united. Table 2 on page 25 gives an overview of the benchmarking criteria. Common elements are the usage of indicators, transparency in performance measurement method and self-assessment of performance.

For the second part of research question one the benchmarking process steps for designing a benchmark are pointed out. Table 3 on page 26 presents an overview of the benchmarking process steps. The literature regarding benchmarking models showed that all benchmarking models have certain steps in order to compare the performance. Common terms in benchmarking are measurement, comparison, identification of best practices, implementation and improvement (Anand & Kodali, 2008, p. 258).

Research question 2: *How can we design a sport policy benchmark?*
The design of the benchmarking is crucial and needs to be tailored to the specific situation. The answer of this research question describes the design of the benchmarking model, that will serve as the foundation for the specific benchmarking. In this study the benchmark is designed for municipal sport policy. The design of the benchmarking method is derived from literature review and consists of phases, process steps and criteria. Based on this study's scope, a benchmarking model is designed see Table 8 on page 30. The table gives an overview of the benchmarking phases: preparing and comparing. The benchmarking steps: setting objectives, defining indicators, selecting benchmark groups, collecting data, preparing data, analyzing data, determining significant different findings and reporting findings. As well as the criteria that are addressed in each benchmarking step accompanied with an abbreviated description of the criterion. For each of the criterions it is indicated to who's criterion it refers to, the criteria from de Bruijn (2002), Maheshwari (2013), or Janssen (2010).

Research question 3: *What does the sport policy benchmark look like?*
This research question is answered in three steps. In the first step the scope and objective of the research was described. Conduct a benchmarking study of municipal sport policy to *measure effectiveness of sport policy*. In step 2 indicators and dimensions were established for this objective. Implicit in the objective are the effectiveness indicators output and outcome. Following the indicators, specific measurable dimensions that have a scientific proven causal relationship needed to be determined. With a review of sport literature and benchmarking literature the following dimensions were found for the indicators. For the indicator output the following three dimensions: Presence, number of sport facilities within the municipality. Proximity, number of sport facilities in district. And variety, number of different sport facilities within district. And for the indicator outcome the dimension: Sport participation; percentage that achieves the fit-norm. Finally in step 3 comparable groups were formed. With a review of benchmarking literature and sport literature an aspect was established for the formation of groups, namely municipal size. Based on municipal size municipalities already made two cooperation's, the G4 and the G32. These two initiatives are used for the classification of comparable groups of municipalities. This led to the following groups: G4, G32, and the smaller 341 municipalities. Together the steps 1-3 give the scope, indicators and dimensions, and the groups.

Research question 4: *How can the developed sport policy benchmark be used?*
This research question corresponds to the comparing phase of the benchmark and the process steps 4-8. In step 4 data for the dimensions was collected and it was found that additional efforts were required in order to analyse the data. First effort was to filter out sport facilities that do not assist in achieving the fit-norm. Second effort was to make the location terms between the data sets consistent. And finally the third effort combine the two data sets. With the combined data set the dimensions could be calculated. In the results two large gaps were found that also could be explained by differences in municipal size. Following correlation calculations were performed. With the selected dimensions and the used data no

significant correlations was found. Finally the results for the dimensions were depicted in a dashboard, enabling self-assessment of the municipal sport policy by the municipal policy-maker.

Research question 5: *What are the effects of the developed municipal sport policy benchmark?*
For research question 5 the benchmarking design, benchmarking findings, and benchmarking process are evaluated.
The design of the developed method is evaluated by assessing whether the method leads to perverse effects. It was found that the design can pervert the measurement system. The design can block innovation and drive out system responsibility, since not the whole municipal sport policy output is taken into account by the selected dimensions. The second perverse effect that applies to the benchmarking method is that benchmarking can veil the actual performance, this is due to the limitations toward the reliability and the quality of the data set from the Health-monitor.
Evaluating the findings of the benchmarking study, learned that no correlation was found between the output of sport policy and the desired effect namely healthy citizen. Evaluating the benchmarking process learned that data analytics are an important factor for this study.

In conclusion does the conducted benchmark improve the current situation?
As stated before there is a lack of insight in the effectiveness of sport policy which in this thesis is defined as: whether the output of the sport policy (namely the quantity, type, and location of sport facilities) leads to the desired effects (namely healthy citizens). It was not possible to calculate the effectiveness with the available data due to problems with the quality of the data. Therefor it is not possible to calculate whether the current situation significantly has improved but is has led to the following improvements:

- This study has made the presence, variety, and proximity of sport facilities in municipalities transparent. This information enables the municipal policy-makers to gain insight in their own network of sport facilities and learn by comparing their numbers with other municipalities.
- A benchmarking method is designed for measuring and comparing sport policy.
- Indicators are selected to measure and compare sport policy namely sport participation and the presence, proximity, and variety of sport facilities.

7.2. Research limitations

This part answers the question: What are the research limitations of this study? The following topics are attended to: the type of research and the replication of the study.

Data driven research.
This study has been designed with data availability in mind. It can be perceived as a data-driven research. The available data sets were from the Mulier institute and the health-monitor. It was expected that the data from these sources would enable accessible and trustworthy data therefore the selection of the dimensions was influenced by the available data. While this has limited the research, because it led to the use of the data set from the health-monitor. The health-monitor consist of privacy sensitive data and a relative small sample size. According to the RIVM in order for the data to be useful for research it needs to be published on a low abstraction level. Therefore the data is anonymised and with a calculation model the figures are approximated at district level and this approximated data set is published as open data. These efforts result in a data set in which the data is not trustworthy since it is

not a representation of the actual situation. Additionally it is possible that the aforementioned efforts have led to the loss causal relations.
Because this study has been designed as a data driven research, it is conceivable that this has limited the research findings. This study expected to find significant correlation between the number, type and location of sport facilities and sport participation levels. The significant correlation was not found and possibly the use of anonymised and calculated data has influenced the findings. Hence this study is limited by the availability of trustworthy data.

Replication question
As part of the research limitations the replication question is answered. This is the question whether new research that uses a similar methodology and design would get similar results (Yin, 2014). In order to gain insight in the replication of this research first the literature used for the design of the benchmarking method is reviewed followed by a review of the application of the benchmarking method.

For the design of the benchmarking method a systematic literature review is conducted, its approach is described in 2.1. Literature review approach. The literature review started with keyword searching, followed by forward searching based on selected articles. With the reviewed literature no specific literature regarding benchmarking and sport policy has been found. This led to the use of mainly non-specific benchmarking literature. More theoretical foundation was needed and the chair of this graduation committee Marijn Janssen suggested the research of Maheshwari. It is probable that without this suggestion the research of Maheshwari was not included in this research. Contemplating on the used scholars and their background, scholars from the TU Delft are represented more than average. The literature review approach and more specifically the forward search could have limited this research to a predominately TU Delft oriented research.

In the application of the benchmarking method it is assessed whether other research would select similar indicators and would get similar results from the data analytics. Based on sport literature review and influenced by data availability the following dimensions were selected: presence, variety, proximity and sport participation. These dimensions where found to be dimensions that are widely researched in sport literature. It is expected that similar research would select similar dimensions for sport facilities without the influence of the availability of data.
Alternative research is expected to yield the same benchmarking results if they use the same data and data preparing and analysing procedure. Possibly the reproducibility of this study is limited since data analytics skills are required due to the number of iterations that are involved in data preparing and analysing. When this study would be executed with the raw data form the health-monitor it is expected that correlation between the dimensions would be found.

7.3. Reflection & recommendations

This section contemplates on the following subjects: benchmarking method, benchmarking application, data, and benchmarking literature. Based on the considerations recommendations are formed.

Benchmarking method
Can municipal policy-makers use the municipal sport policy benchmark for policy-making? The benchmarking method and its findings can be used to gain insights in the number, types, and location

of sport facilities within a municipality. Subsequently this can be used to compare the results to other municipalities.
Benchmarking method recommendation 1: Use the gained insights for sport facilities to corroborate arguments.

The effectiveness score that was designed as the relation between sport facilities and sport participation was not found, indicated as cause for this was the unreliable data. Based on the reviewed sport literature it is expected that the two indicators influence each other. If municipalities have reliable internal data regarding sport participation they could use that data and the designed benchmarking method to determine their effectiveness.
Benchmarking method recommendation 2: Redo the municipal sport policy benchmark when reliable data is available for sport participation.

With the obtained insights form measuring and comparing sport facilities and the lack of reliable data regarding sport participation a following step for municipalities could be profiling. Whereas benchmarking is focussed on the performance of an organisation, profiling focusses on the characteristics of an organisation. The characteristics can assist in identifying categories of organisations i.e. elite sport municipality or a sports for all municipality. Subsequently profiling can help to narrow down where to optimize. For example a municipality can self-select certain dimensions and use it to form a character group which they want to achieve. The steps for this would be measuring, comparing, profiling.
Benchmarking method recommendation 2: further research in whether and how characterising could improve benchmarking in the public sector.

Benchmarking application
Can we use the designed benchmarking method outside the context of this study? This study focussed on designing and conducting a municipal sport policy benchmark. The designed benchmarking method uses: phases, steps, and criteria directed at assessing the performance of municipal sport policy. The designed benchmarking method can also be wider used for example to measure and compare other municipal real estate as libraries and community centres.
To illustrate that the benchmarking method can be wider used, a morphological chart is created. The morphological chart is used to make insightful what possible considerations are per step of the benchmarking method. The steps from the benchmarking method are formulated as functions and these are broken down into sub functions followed by the corresponding options. The morphological chart is depicted in the appendix 9.4 Figure 14.
Reflecting on the conducted benchmark and with the morphological chart the following steps are indicated as important:

- Step 2 the selection of indicators and dimensions is restricted to what is measurable. Hence it is sometimes not possible to cover the whole part of a system since this can consist of measurable and unmeasurable elements.
- Step 2 the selection of indicators and dimensions and step 4 collecting data are interrelated. The selection dictates the data needs or the choice is dictated by data availability which was the case in this study.
- Steps 5 and 6 the preparing and analysing of the data requires data analytics skills. This study found that these steps involved a number of iterations that requires data analytics skills.

Benchmarking application recommendation: The designed benchmarking method can be used for other studies, taking into consideration the importance of the selection of the indicators, and data collecting, preparing, analysing.

Open data

Can we use open data for policy-making? One of the results of this study is the importance of reliable data. In this study readily available data is used consisting of public and open data. It was found that the reliability of the open data was questionable. Since the data set from the health-monitor used sensitive personal information and the small sample the results were anonymised and with a calculation model the results were approximated at district level. These efforts possibly have influenced the reliability of the data.

This study recommends further research in the use of measurement tools for the health-monitor. Currently surveys and questionnaires are used as a measurement tool. Here the participants have to interpret their own sport activity levels leading possible bias and misinterpretations. To increase the reliability of the data an option could be to set up a platform and gather the data from fitness trackers, heart trackers, and GPS-locations. Anonymise the users and publish the data as raw as possible for example on pc-4 level. As a by-product this kind of data collection would give insight in were sport is done. Currently this research takes into account the specific designed sport facilities and omits the public space: streets, parks, and beaches.

Open data recommendation: Open data can be used for policy-making provided that it is raw data. Publication of calculated data is useless.

Benchmarking literature

How can benchmarking literature improve? The existing reviewed benchmarking literature focusses on the design of the benchmarking method. This study confirms that the design of benchmarking is important in addition to this it found that the benchmarking application proves to be an important factor. The current paradigm is that the possible adverse effects of benchmarking can be solved by its design. This study found that benchmarking application also can lead to adverse effects and further research is needed in benchmarking and data related tasks, the formation of groups, and presenting of benchmarking results.

This study found that data related tasks influenced the benchmarking application and results. Because data availability influenced the selection of benchmarking criteria. And the structure and the quality of the collected data influenced the data preparing and analysing efforts. These findings are not out of the ordinary since benchmarking falls under performance analytics which extensively uses data. As data sets are getting bigger and more data is made available as open data research in the influence of data related tasks and benchmarking is increasingly relevant.

Benchmarking literature recommendation 1: further research in the influence of data related tasks in benchmarking.

In the current benchmarking literature one indicator was found for the formation of benchmarking groups namely size. It is important to have more indicators to form groups of comparable organisations. All organisations are different and it is expected to contribute to the relevance of the benchmarking results and the implementation of the findings.

Benchmarking literature recommendation 2: further research in indicators for the formation of groups.

Current benchmarking literature states that self-assessment of the results is an important factor however generally it lacks to give guidance to the way of presenting results. To enable self-assessment a dashboard is used to visualize and present the results. The way of presenting the results is expected to influence the assessment. For example the assessment of results could be influenced by selecting a type of graph and use of colours certain results could be favoured.

Benchmarking literature recommendation 3: further research in way of presenting benchmarking results.

8. References

Aarts, L. T. (2014). *Benchmarking van gemeentelijke verkeersveiligheid in de praktijk*. Den Haag. Retrieved from https://www.swov.nl/sites/default/files/publicaties/rapport/r-2014-34.pdf

Ainsworth, B. E., Haskell, W. L., Herrmann, S. D., Meckes, N., Bassett, D. R., Tudor-Locke, C., … Leon, A. S. (2011). 2011 Compendium of Physical Activities: a second update of codes and MET values. *Medicine and Science in Sports and Exercise*, *43*(8), 1575–81. https://doi.org/10.1249/MSS.0b013e31821ece12

Ainsworth, B. E., Haskell, W. L., Whitt, M. C., Irwin, M. L., Swartz, A. M., Strath, S. J., … Leon, A. S. (2000). Compendium of Physical Activities: an update of activity codes and MET intensities. *Medicine & Science in Sports & Exercise*, *32*(Supplement), S498–S516. https://doi.org/10.1097/00005768-200009001-00009

Anand, G., & Kodali, R. (2008). Benchmarking the benchmarking models. *Benchmarking: An International Journal*, *15*(3), 257–291. https://doi.org/10.1108/14635770810876593

Bannister, F. (2007). The curse of the benchmark: an assessment of the validity and value of e-government comparisons. *International Review of Administrative Sciences*, *73*(2), 171–188. https://doi.org/10.1177/0020852307077959

Beemt, A. van den, & Veuger, J. (2016). Barometer for Municipal Real Estate: Developments form 2008 until 2015. In *Book of Proceedings 23nd European Real Estate Society Annual Conference*. Retrieved from https://research.hanze.nl/nl/publications/barometer-for-municipal-real-estate-developments-form-2008-until-

Bhutta, K. S., & Huq, F. (1999). Benchmarking – best practices: an integrated approach. *Benchmarking: An International Journal*, *6*(3), 254–268. https://doi.org/10.1108/14635779910289261

Bouckaert, G., Hondeghem, A., Voets, J. (Joris), Op de Beeck, S., & Cautaert, E. (2011). *Handboek overheidsmanagement*. Brugge: Vanden Broele.

Brink, C. van den. (2017). Sport en bewegen. Retrieved from https://www.volksgezondheidenzorg.info/onderwerp/sport-en-bewegen/cijfers-context/huidige-situatie#methoden

Camp, R. (1989). *Benchmarking The Search for Industry Best Practices that Lead to Superior Performance*. Milwaukee: ASQC Quality Press.

CBS. (2017). CBS municipal size. Retrieved from https://www.cbs.nl/nl-nl/onze-diensten/methoden/classificaties/overig/gemeentegrootte-en-stedelijkheid

de Bruijn, H. (2002). Performance measurement in the public sector: strategies to cope with the risks of performance measurement. *International Journal of Public Sector Management*, *15*(7), 578–594. https://doi.org/10.1108/09513550210448607

de Bruijn, H. (2007). *Managing Performance in the Public Sector* (2nd ed.). Routledge. https://doi.org/doi:10.4324/9780203946336

de Goede, M., Enserink, B., Worm, I., & van der Hoek, J. P. (2016). Drivers for performance improvement originating from the Dutch drinking water benchmark. *Water Policy*, *18*(5), 1247–1266. https://doi.org/10.2166/wp.2016.125

Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P., & Meester, L. E. (2005). *A Modern Introduction to Probability and Statistics*. London: Springer London. https://doi.org/10.1007/1-84628-168-7

Giles-Corti, B., & Donovan, R. J. (2002). The relative influence of individual, social and physical environment determinants of physical activity. *Social Science & Medicine*, *54*, 1793–1812. Retrieved from

https://www.researchgate.net/profile/Robert_Donovan/publication/11266044_The_relative_influence_of_individual_social_and_physical_environment_determinants_of_physical_activity/links/0c960515697626dc01000000.pdf

Hallmann, K., Wicker, P., Breuer, C., & Schönherr, L. (2012). Understanding the importance of sport infrastructure for participation in different sports – findings from multi-level modeling. *European Sport Management Quarterly*, *4742*(May 2015), 1–20. https://doi.org/10.1080/16184742.2012.687756

Hallmann, K., Wicker, P., Breuer, C., & Schüttoff, U. (2011). Interdependency of sport supply and sport demand in German metropolitan and medium-sized municipalities – findings from multi-level analyses. *European Journal for Sport and Society*, *8*(12), 65–84. Retrieved from https://www.waxmann.com/index.php?eID=download&id_artikel=ART101016&uid=frei

Heijer, A. C., & de Vries, J. C. (2004). *Benchmarking universiteitsvastgoecl managementinformatie bij vastgoedbeslissingen*. Delft: TU Delft, Real Estate & Housing.

Hildebrandt, V. H., Ooijendijk, W. T. M., & Hopman-Rock, M. (2008). *Trendrapport Bewegen en Gezondheid 2006/2007*. TNO Kwaliteit van Leven.

Hoekman, R., Breedveld, K., & Kraaykamp, G. (2016). Sport participation and the social and physical environment: explaining differences between urban and rural areas in the Netherlands. *Leisure Studies*, 1–14. https://doi.org/10.1080/02614367.2016.1182201

Hoogerwerf, A., & Herweijer, M. (2014). *Overheidsbeleid* (9th ed.). Wolters Kluwer.

Janssen, M. (2010). Measuring and Benchmarking the Back-end of E-Government: A Participative Self-assessment Approach. In *International Conference on Electronic Government* (pp. 156–167). Berlin: Springer. https://doi.org/10.1007/978-3-642-14799-9_14

Lasswell, H. (1951). *The Policy Orientation*. *The Policy Sciences*. Stanford: Stanford University Press. https://doi.org/10.1007/s13398-014-0173-7.2

Levy, Y., & Ellis, T. J. (2006). A Systems Approach to Conduct an Effective Literature Review in Support of Information Systems Research. *Informing Science Journal*, *9*. Retrieved from http://www.inform.nu/Articles/Vol9/V9p181-212Levy99.pdf

Limstrand, T., & Rehrer, N. J. (2008). Young people' s use of sports facilities: A Norwegian study on physical activity. *Scandinavian Journal of Social Medicine*, *36*(5), 452–459. https://doi.org/10.1177/1403494807088455

Maheshwari, D. (2013). Measurement and benchmarking foundations: Providing support to organizations in their development and growth using dashboards. *Government Information Quarterly*, *30*, S83–S93. https://doi.org/10.1016/j.giq.2012.11.002

Maheshwari, D. (2015, January). *Measuring and Benchmarking Public Service Organizations*. TU Delft. https://doi.org/10.4233/UUID:DB5220CA-A614-4A6F-B203-DD048A68C192

Neely, A., Gregory, M., & Platts, K. (1995). International Journal of Operations & Production Management Performance measurement system design: A literature review and research agenda For Authors Performance measurement system design A literature review and research agenda. *International Journal of Operations & Production Management International Journal of Operations &amp Production Management Iss International Journal of Service Industry Management Iss*, *15*(4), 80–116. Retrieved from https://doi.org/10.1108/01443579510083622

NEN-EN 15221-7. (2012). *Guidelines for Performance Benchmarking*. Brussels. https://doi.org/03.080.99

Sallis, J. F., Hovell, M. F., Hofstetter, C. R., Elder, J. P., Hackley, M., Caspersen, C. J., & Powell, K. E. (1990). Distance between homes and exercise facilities related to frequency of exercise among San Diego residents. *Public Health Reports (Washington, D.C.: 1974)*, *105*(2), 179–85. Retrieved from

http://www.ncbi.nlm.nih.gov/pubmed/2108465

Susha, I., Grönlund, Å., & Janssen, M. (2015). Organizational measures to stimulate user engagement with open data. *Transforming Government: People, Process and Policy*, *9*(2), 181–206. https://doi.org/10.1108/TG-05-2014-0016

Tiessen-Raaphorst, A., Verbeek, D., de Haan, J., & Breedveld, K. (2010). *Sport: een leven lang*. Den Haag: SCP,. Mulier Instituut. Retrieved from https://www.scp.nl/Publicaties/Alle_publicaties/Publicaties_2010/Sport_een_leven_lang

Treffinger, D. J. (2000). *Creative problem solver' s guidebook*. Waco: Prufrock Press.

van der Poel, H., Wezenberg-Hoenderkamp, K., & Hoekman, R. (2016). *Sportaccommodaties in Nederland*. Utrecht: Mulier Instituut. Retrieved from http://www.kennisbanksportenbewegen.nl/?file=7295&m=1475158489&action=file.download

Warburton, D. E. R. (2006). Health benefits of physical activity: the evidence. *Canadian Medical Association Journal*, *174*(6), 801–809. https://doi.org/10.1503/cmaj.051351

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: writing a literature review. *MIS Quarterly*, *26*(2). Retrieved from http://www.misq.org/misreview/announce.html

Wicker, P., Breuer, C., & Pawlowski, T. (2009). Promoting Sport for All to Age-specific Target Groups: the Impact of Sport Infrastructure. *European Sport Management Quarterly*, *9*(2), 103–118. https://doi.org/10.1080/16184740802571377

Wicker, P., Hallmann, K., & Breuer, C. (2013). Analyzing the impact of sport infrastructure on sport participation using geo-coded data: Evidence from multi-level models. *Sport Management Review*, *16*(1), 54–67. https://doi.org/10.1016/j.smr.2012.05.001

Yin, R. K. (2014). *Case study research : design and methods*. Sage publishing.

9. Appendices

Below the appendices of the thesis are depicted. 9.1. Gives an overview of the sport types that fall in the category of high intensity sports. 9.2. Gives the calculation results of the correlation scores of sport participation and various individual aspects. 9.3. gives an overview of the benchmarking results for the group G32. 9.4. Depicts the morphological chart of the designed benchmarking method.

9.1. Step 5 Overview type sports

In Table 16 an overview is given of all type sports that could be found in the DSA. Indicated in the first columns is the sport's corresponding code number and MET score as defined by Ainsworth (2011). Conditional that the MET score is equal or above 5 or 6,5 the type sport is taken into account, respectively column 4 and 5. Five sport types lack a specific MET scores, in those instances a value is calculated. Depicted in Table 17 and Table 18.

code number	METs (Ainsworth)	Type sport	Type ≤5	Type ≤6,5	Remark
3016	8,5	Aerobics	Aerobics	Aerobics	
15210	8	American Football	American Football	American Football	
15733*	6,7	Atletiek	Atletiek	Atletiek	* Athletics is a sport facility were you can achieve the required MET-score depending on the performed activity. Therefore it is included in the MET≥6,5 en MET ≥ 5. Calculation in.
15192	8,5	Autosport	Autosport	Autosport	
15725	8	Beachhandbal	Beachhandbal	Beachhandbal	
15725	8	Beachvolleybal	Beachvolleybal	Beachvolleybal	
15210	8,5	Betaald Voetbal	Betaald Voetbal	Betaald Voetbal	
15138	6	Bewegingsonderwijs	Bewegingsonderwijs	-	
1015	8	Bikepark	Bikepark	Bikepark	
15080	2,5	Biljart	-	-	
15120	7,8	Boksen	Boksen	Boksen	
15090	3	Bowlen	-	-	
15092	3	Bowls/Koersbal	-	-	
9010	1,5	Bridge	-	-	
15210**	7,9	Buitenspelen	Buitenspelen	Buitenspelen	** Cruyff en Krajicek foundation courts. Tennis and Soccer. Tennis [15685 MET 7,3], voetbal [15210 MET 8,5] = avg 7,9
15150	4,8	Cricket	-	-	
15170	4	Curling	-	-	
3010	5	Dansen	Dansen	-	
9013	1,5	Denksport	-	-	
18200	7	Duiken	Duiken	Duiken	
09115***	1,5	Evenementen	-	-	*** Events, one entry are the Brabant hallen used for markets, music events, and elite sport events (horse riding, kickboxing). Sports spectator [09115 1,5] **** Fitness, is a sport facility were you can achieve the required MET-score depending on the performed activity. Therefore it is included in the MET≥6,5 and MET≥5. Calculations in Table 18
1009	8,5	Fietscross	Fietscross	Fietscross	
2050*****	6,55	Fitness	Fitness	Fitness	
15265	4,5	Golf	-	-	
15300	3,8	Gymnastiek	-	-	
15330	8	Handbal	Handbal	Handbal	
15010	4,3	Handboogschieten	-	-	
1530	7,8	Hockey	Hockey	Hockey	
17165	3	Hondensport	-	-	
15620	5	Honkbal	Honkbal	-	
15360	8	Ijshockey	Ijshockey	Ijshockey	
15465	3,3	Jeu De Boules	-	-	
15430	10,3	Judo	Judo	Judo	
15465	3,3	Kaatsen	-	-	
18025	3,3	Kanosport	-	-	
15192	8,5	Karten	Karten	Karten	
15090	2,5	Kegelen	-	-	
4130	2,5	Kleiduivenschieten	-	-	
15535	7,5	Klimsport	Klimsport	Klimsport	
4130	2,5	Klootschieten	-	-	
15711	6	Korfbal	Korfbal	-	
15590	7	Kunstrolschaatsen	Kunstrolschaatsen	Kunstrolschaatsen	
1070	5	Kunstwielrijden	Kunstwielrijden	-	
15460	8	Lacrosse	Lacrosse	Lacrosse	
15265	3,5	Midgetgolf	-	-	
15470	4	Motorcross	-	-	
15470	4	Motorsport	-	-	

02001*****	3	Outdoor	-	-	***** Company activities, archery / yoga / playing / abseiling low intensity activity. Below 5 MET
15370	5,5	Paardensport	Paardensport	-	
2072	7	Roeien	Roeien	Roeien	
15560	8,3	Rugby	Rugby	Rugby	
19030	7	Schaatsen	Schaatsen	Schaatsen	
19030	7	Schaatsen-natuurijs	Schaatsen-natuurijs	Schaatsen-natuurijs	
9010	1,5	Schaken	-	-	
15200	6	Schermen	Schermen	-	
4130	2,5	Schieten	-	-	
15591	7,5	Skaten	Skaten	Skaten	
15590	7,5	Skeeleren	Skeeleren	Skeeleren	
19075	8	Skiën	Skiën	Skiën	
15600	3,5	Skydive	-	-	
15625	5	Softbal	Softbal	-	
15650	10	Squash	Squash	Squash	
18120	3	Surfen	-	-	
15660	4	Tafeltennis	-	-	
15685	7,3	Tennis	Tennis	Tennis	
15430	10,3	Vechtsport	Vechtsport	Vechtsport	
16020	2	Vliegen	-	-	
15210	8,5	Voetbal	Voetbal	Voetbal	
15711	8	Volleybal	Volleybal	Volleybal	
18355	6	Watersport	Watersport	-	
1015	8,5	Wielersport	Wielersport	Wielersport	
15711	7	Zaalsport	Zaalsport	Zaalsport	
15711	7	Zaalvoetbal	Zaalvoetbal	Zaalvoetbal	
18120	3	Zeilen	-	-	
18230	8	Zwemmen	Zwemmen	Zwemmen	

Table 16 Overview of all sport facilities in the DSA from Mulier institute combined with code number and MET value from Ainsworth (2011). If a sport facility does not meet the required MET score it is filtered out for example: golf, motor cross, chess, and table tennis.

Athletics		
code number	**METs**	**remark**
15732	4	track and field (e.g., shot, discus, hammer throw)
15733	6	track and field (e.g., high jump, long jump, triple jump, javelin, pole vault)
15734	10	track and field (e.g., steeplechase, hurdles)
	6,7	average Athletics score

Table 17 Step 5 calculations for an athletics sport facility. No specific MET value is available for this sport facility therefore it is calculated based on the common activities that can be executed on this sport facility. MET values, codes and remark from Ainsworth (2011)

Fitness		
code number	**METs**	**remark**
2050	6	resistance training (weight lifting, free weight, nautilus or universal), power lifting or body building, vigorous effort.
2052	5	resistance (weight) training, squats , slow or explosive effort
2054	3,5	resistance (weight) training, multiple exercises, 8-15 repetitions at varied resistance
2062	7,8	health club exercise (body combat class, body pump class)
2065	9	stair treadmill, ergometer
2040	8	circuit training
	6,55	average Fitness score

Table 18 Step 5 calculations for a fitness sport facility. No specific MET value is available for this sport facility therefore it is calculated based on the common activities that can be executed on this sport facility. MET values, codes and remark from Ainsworth (2011)

9.2. Step 7 Results correlation sport participation and individual aspects

Below an overview is given of the correlation results between various individual aspects and sport participation. With the used data no significant correlation was found between the aspects and sport participation. A medium correlation was found between the aspect 'low purchasing power' and 'sport participation'. The rest of the calculated correlations are negligible.

Correlation	Sport participation
Sport participation	1,00
OAD	-0,19
STED	0,12
Aantal	-0,06
Diversiteit	0,05
Fitnorm	1,00
Stedelijkheid	0,12
Aant inw	-0,10
AANT p1000	-0,04
DIV p1000	-0,05
AANT_INW	-0,10
AANT_MAN	-0,10
AANT_VROUW	-0,10
P_00_14_JR	0,00
P_15_24_JR	-0,07
P_25_44_JR	-0,28
P_45_64_JR	0,17
P_65_EO_JR	0,19
P_N_W_AL	-0,32
P_MAROKKO	-0,27
P_ANT_ARU	-0,30
P_SURINAM	-0,36
P_TURKIJE	-0,27
P_OVER_NW	-0,33
WONINGEN	-0,11
WOZ	0,21
P_KOOPWON	0,16
P_HUURWON	-0,15
P_HUURCORP	-0,06
P_HUUROVVH	-0,21
INK_ONTV	0,13
INK_INW	0,23
P_LAAGINKP	-0,11
P_HOOGINKP	0,14
P_LAAGINKH	-0,21
P_HOOGINKH	0,17
P_LKOOPKRH	**-0,49**
P_SOCMINH	**-0,48**
AO_UIT_TOT	-0,12
P_AO_UIT	-0,17
WW_UIT_TOT	-0,13
P_WW_UIT	-0,03
WWB_UITTOT	-0,21
P_WWB_UIT	-0,30
P_NIETACT	-0,27
AUTO_TOT	-0,05
AUTO_HH	0,04
AUTO_LAND	-0,15

Table 19 Overview of the correlation results between sport participation and the individual aspects

9.3. Step 7 Results for G32

Below the results of the dimensions are depicted for the municipalities in the group G32. The overview ranked and ordered based on the effectiveness score. Which is defined as the ratio between the municipal's fit-norm score and the fit-norm score from the average of the group.

Municipality	ranking	effectiveness	% fit-norm	# MET5	# MET65	xyz municipality	km MET5	km MET65	xyz district	xyz district	# avg	% avg	inhabitants 2012	# per 1000	xyz per 1000
Haarlem	1	1,20	27,9	140	129	18	14,0	12,9	1,0	4,6	107,2	23,3	150670	0,93	0,12
Hengelo	2	1,19	27,8	62	60	15	6,2	6,0	3,8	3,6	107,2	23,3	80939	0,77	0,19
Hoorn	3	1,15	26,7	65	63	17	5,0	4,8	3,5	3,4	107,2	23,3	70697	0,92	0,24
Gouda	4	1,14	26,6	69	65	10	8,6	8,1	3,9	3,5	107,2	23,3	71235	0,97	0,14
Deventer	5	1,14	26,5	83	81	19	7,5	7,4	4,7	4,5	107,2	23,3	98672	0,84	0,19
Alphen aan den Rijn	6	1,10	25,7	103	85	19	6,4	5,3	4,4	3,5	107,2	23,3	72853	1,41	0,26
Zwolle	7	1,10	25,6	121	118	18	7,6	7,4	3,6	3,4	107,2	23,3	121527	1,00	0,15
Venlo	8	1,10	25,6	111	100	19	5,0	4,5	3,7	3,4	107,2	23,3	100027	1,11	0,19
Delft	9	1,09	25,3	83	77	17	5,9	5,5	3,5	3,1	107,2	23,3	98675	0,84	0,17
Roosendaal	10	1,07	24,9	78	69	16	6,0	5,3	4,2	3,5	107,2	23,3	77426	1,01	0,21
Almelo	11	1,07	24,9	57	54	14	5,2	4,9	3,6	3,4	107,2	23,3	72757	0,78	0,19
Apeldoorn	12	1,07	24,9	160	152	22	10,7	10,1	5,0	4,5	107,2	23,3	156961	1,02	0,14
Arnhem	13	1,07	24,9	160	152	22	10,7	10,1	5,0	4,5	107,2	23,3	149271	1,07	0,15
Zaanstad	14	1,07	24,8	125	117	15	6,6	6,2	3,7	3,3	107,2	23,3	148281	0,84	0,10
Leiden	15	1,06	24,8	92	83	20	9,2	8,3	4,6	3,9	107,2	23,3	118748	0,77	0,17
Enschede	16	1,06	24,7	157	147	22	15,7	14,7	6,5	5,8	107,2	23,3	158048	0,99	0,14
Haarlemmermeer	17	1,06	24,6	111	103	16	9,3	8,6	4,3	3,7	107,2	23,3	143374	0,77	0,11
Breda	18	1,05	24,5	122	110	22	11,1	10,0	6,4	5,6	107,2	23,3	176401	0,69	0,12
Groningen	19	1,05	24,5	170	155	20	15,5	14,1	5,8	4,8	107,2	23,3	193127	0,88	0,10
Tilburg	20	1,01	23,5	126	121	21	12,6	12,1	6,2	5,7	107,2	23,3	207580	0,61	0,10
Oss	21	1,00	23,3	95	87	18	4,8	4,4	3,5	3,2	107,2	23,3	84639	1,12	0,21
average G38	22	1,00	23,3	107,2	100,0	17,9	9,7	9,1	4,5	4,1	107,2	23,3	121309	0,88	0,15
Helmond	23	0,99	23,0	60	57	14	5,0	4,8	3,3	3,0	107,2	23,3	88560	0,68	0,16
Zoetermeer	24	0,98	22,8	81	68	17	8,1	6,8	4,0	3,1	107,2	23,3	122331	0,66	0,14
's-Hertogenbosch	25	0,98	22,7	135	129	20	9,0	8,6	4,4	4,1	107,2	23,3	140786	0,96	0,14
Ede	26	0,97	22,7	95	89	12	5,6	5,2	3,0	2,7	107,2	23,3	108763	0,87	0,11
Amersfoort	27	0,97	22,6	121	117	18	4,2	4,0	2,9	2,7	107,2	23,3	148250	0,82	0,12
Maastricht	28	0,94	22,0	112	109	16	16,0	15,6	6,1	5,7	107,2	23,3	121050	0,93	0,13
Sittard-Geleen	29	0,93	21,6	110	104	14	13,8	13,0	5,4	4,8	107,2	23,3	94535	1,16	0,15
Nijmegen	30	0,93	21,6	115	110	20	12,8	12,2	7,0	6,4	107,2	23,3	165128	0,70	0,12
Eindhoven	31	0,92	21,4	178	167	24	22,3	20,9	8,4	7,5	107,2	23,3	217225	0,82	0,11
Emmen	32	0,92	21,4	178	167	24	22,3	20,9	8,4	7,5	107,2	23,3	108838	1,64	0,22
Leeuwarden	33	0,91	21,3	103	93	20	4,7	4,2	2,8	2,5	107,2	23,3	95321	1,08	0,21
Dordrecht	34	0,89	20,7	90	82	18	6,9	6,3	3,7	3,2	107,2	23,3	118862	0,76	0,15
Schiedam	35	0,87	20,4	63	60	15	7,9	7,5	3,6	3,3	107,2	23,3	76244	0,83	0,20
Alkmaar	36	0,85	19,9	86	84	19	8,9	8,7	4,4	4,2	107,2	23,3	94269	0,91	0,20
Lelystad	37	0,79	18,3	62,00	54	16	6,2	5,4	3,2	2,8	107,2	23,3	75312	0,82	0,21
Almere	38	0,68	15,8	109	105	20	27,3	26,3	8,0	7,0	107,2	23,3	193163	0,56	0,10
Heerlen	39	0,62	14,5	84	77	15	3,8	3,5	2,5	2,2	107,2	23,3	89212	0,94	0,17

Table 20 Overview of the results of the G32. Ordered by ranking for the effectiveness score.

9.4. Morphological chart

The morphological chart is depicted regarding the designed benchmarking method. The benchmarking steps are translated to functions and sub function. For each sub function possible options are depicted. The figure is followed by explanation per function.

Morphological chart benchmarking method					
Function	**Sub function**	**Options**			
1. Defining Objectives	*Domain*	Products	Services		
	Frequency	Once	Periodic	Continuous	
	Measurement level	Organisational	Industry	generic	
2.Defining indicators	*Indicators*	Input	Throughput	Output	outcome
	Dimensions	Number	Variety	Financial	Spatial
	Measures	Extremes	Dispersion	Central tendency	Example
3. Formation groups	*Criteria*	One group	Size		
4. Collecting data	*Multiple data collection methods*	Available quantitative data	Available qualitative data.	Collect quantitative data	Collect qualitative data
5. Preparing data	*Combing and normalising data sets*	Filter	Combining	Transferring	
6. Analysing data	*Examine data with regard to indicators*	Counting	Calculating		
7.Determining findings	*Determine gaps between indicators*	Top 10%	Bottom 10%	Standard deviation	
	Possible explanation gaps	Context	Correlation		
8.Reporting findings	*Results*	Individual results	Relative results		
	Visualize in a dashboard	Graph	Number	Table	Chart

Figure 14 Morphological chart, a display of the choice options in the benchmarking method directed to make the deliberations insightful per step. Morphological chart from Treffinger (2000) *adapted towards this study's designed benchmarking method.*

Explanation of the benchmarking function.

The benchmarking method can be used to compare performance of organisation. In order to use the benchmarking method the following steps need to be followed.

1. **Defining objectives**

Define the objectives of the benchmarking study: the domain, the frequency and the measurement level.

- *Domain:* Select the domain of the organisation that is benchmarked the performance of products or the performance of services.

- *Frequency*: Select the period of time for which the benchmarking occurs: periodic (daily, weekly, monthly, yearly), continuous (monitoring-data)
- *Measurement level:* Depending on the scope of the benchmark chose the measurement level:
    - Organisational level, comparison inside one organization regarding the performance of related business parts.
    - Industry level, comparison with companies operating in the same industry.
    - Generic level, comparison with all other organizations.

2. Defining indicators

Define indicators, dimensions for the indicators, and the measures.

- Select indicators: Based on the scope of the scope and domain of the benchmark select the indicators. If the benchmark focusses on: resources (inputs), activities (throughputs), performance (outputs) and effects (outcomes). With the four indicators combinations can measure the productivity, efficiency and effectiveness. Productivity is the ratio output and specific input. Efficiency is the ratio output and all inputs. Effectiveness is the ratio outcome and output.
- Select dimensions for the indicators. Use indicators with dimensions that have a scientific proven causal relationship and that are measurable.
- Select the measures for the dimensions, measures of: extremes, dispersion, or central tendency. These are used to determine the gaps in the results.

3. Define criteria for groups

Comparison against comparable organisations, under the same circumstance and with the same measures. Use objective criteria with preferably scientific proven relationship.

5. Collecting data

Use multiple data collection methods options are the use of: Available quantitative data, Available qualitative data or gathering your own data: collect quantitative data, and collect qualitative data.

7. Preparing data

Combing and normalising the collected data sets for the analysing of the data. This step involves the filtering, combining, and transferring of data. Depending on the scope and the size of the data sets data analytics skills are required.

9. Analysing data

With the obtained data set from the data preparing step examine the data with regard to defined indicators and dimensions in step 2. This can involve counting and calculating the data.

11. Determine findings

Determine gaps between the dimensions and the measures and put the gaps in to context and try to explain the gaps.

- Determine gaps. Compare the scores with regard to the selected measures.
- If possible explain the gaps based on the context and with correlation calculations.

12. Report findings

Report the findings as an overview in a dashboard. Publish the findings as individual results in a dashboard to enable self-assessment. Select the way of presenting the results: graph, single numbers, table, or a chart.