Anssi Juvela

Complexities of Benchmarking

Vaasa 2023

UNIVERSITY OF VAASA
Academic School: School of management
Writer: Anssi Juvela
Name of the paper: Complexities of Benchmarking
Degree: Master of Administrative Sciences
Academic Subject: Public Management
Supervisor: Professor, Dr. Christoph Demmke
Year of completion: 2023
Amount of pages: 85

SUMMARY:

In this master's thesis my central goal is to examine benchmarking as an academic phenomenon and moreover how principles of complexity thinking can be adapted into benchmarking and comparison of public governments. An overview of the history and ideology behind contemporary benchmarking boom is also laid. Focus in the case-study section is on two worldwide surveys: the Happiness Index and the InCiSE. My aim is to analyze the techniques used in both of the studies consisting of their advantages and disadvantages.

One aspect is how to perceive the vast amounts of comparative data that where the answer lies in complexity thinking. On a more practical level the thesis discusses certain political aspects related to happiness and complexity. One key element in the thesis is also to scrutinize why Finland did so well in both of the benchmarking surveys.

KEY WORDS: Benchmarking, Ranking, InCiSE, Happiness Index, Finland, Comparative Analysis, Complexity thinking, Systems of politics and state-apparatus

VAASAN YLIOPISTO
Akateeminen yksikkö: Johtamisen akateeminen yksikkö
Kirjoittaja: Anssi Juvela
Tutkielman nimi: Vertailukehittämisen kompleksisuudesta
Tutkinto: Hallintotieteiden maisteri
Pääaine: Julkisjohtaminen
Ohjaaja: Professori, tri. Christoph Demmke
Valmistumisvuosi: 2023
Sivumäärä: 85

TIIVISTELMÄ:

Tässä Pro Gradu-tutkielmassa keskeisenä tavoittenani on tutkia vertailukehittämistä tieteellisenä ilmiönä. Sen lisäksi tarkastelen sitä, kuinka kompleksisuusajattelun periaatteita voitaisiin soveltaa vertailukehittämiseen ja julkishallintojen vertaamiseen. Työ sisältää katsauksen vallitsevan vertailukehittämis-innon taustalla olevaan historialliseen ja ideologiseen kehitykseen. Painopiste tapaustutkimus-osiossa on kahdessa kansainvälisessä kyselytutkimuksessa: Onnellisuusideksissä ja InCiSEssä. Tavoitteenani on analysoida kummassakin tutkimuksessa käytettyjä tekiikoita vahvuuksineen ja kehityskohteineen.

Eräs tutkielman keskeisistä näkökulmista on se, kuinka kompleksisuus-ajattelun avulla vertailujen kautta saatuja valtavia tietomääriä voitaisiin entistäkin paremmin käsitellä. Käytännöllisempien tulokulmien osalta tutkielmassa käsitellään tiettyjä poliittisluonteisia näkökohtia liittyen onnellisuuteen ja kompleksisuuteen. Punaisena lankana on myös tarkastella Suomen hyvää menestystä molemmissa vertailukehittämis-tutkimuksissa.

AVAINSANAT: vertailukehittäminen, vertailututkimukset, InCiSE, Onnellisuusindeksi, Suomi, vertaileva analyysi, kompleksisuusajattelu, poliittiset- ja valtiokoneistot

Table of contents 4

No appendixes

1. INTRODUCTION

Administrative sciences are a relatively new and emerging branch of social sciences. In Finland they are taught only in a few universities. In addition to the University of Vaasa administrative sciences can be found at Universities of Tampere, Lapland, and Eastern Finland. There is also a collective of administration fellows called the Society of Administrative Research. Most commonly administrative sciences deal with public administration, public law, public policy, and various public organizational theories. Also, management and leadership are amongst top scientific interests amongst our branches. Administrative sciences are commonly regarded as “a more practical” version of the political sciences, which originate already in the German tradition of “Allgemeine Staatslehre” or more contemporary “Politikwissenschaft”. Of course, language structures our thinking and even way of perceiving the world and there can be communication difficulties in putting various languages or cultural traditions emerging from them into common universal measurements.

Political science focuses on the political parties, political history, and representative bodies in democracies as well as various forms of rule over history and still in modern use. Political science is also more engaged with ideologies and various philosophical schools of thought and their effect on society. Administrative sciences then again operate more on the field of “state apparatus” also the functionalities of public institutions. State law structures the foundations of civil service as all public service must in a constitutional state be based on law. Analysis of public organizations and various ways of leading public organizations are more emphasized in administrative science. With contemporary society being so sophisticated and having detailed knowledge on complex issues and phenomena our brain is prone to make comparisons in order to outline the (social) world we live in. And ultimately that is what benchmarking is also about; comparing things to make a sense out of them.

It might be overtly theoretical in the tradition of administrative sciences to further discuss complexity theory. However, in the modern society where relations of interdependence have become more and more networked and complicated it might be of both scientific and practical use to develop more sophisticated tools for perceiving the world itself, social realms, and their progression.

More profoundly studying the methodology of benchmarking has proved out that even in a relatively simple-looking benchmark-survey there is a vast amount of data and variables, in depth understanding of which very well requires the adaptations in approach related to complexity-thinking. Another notion related to benchmarking is that I was supposed to use the International Civil Service Effectiveness-index or as later referred to the InCiSE-benchmark-survey as a case example (Blavatnik School of Government, 2017). However, execution of it has for the time being put to a pause during the writing-process. Therefore, I ended up adding the benchmark-survey of Happiness Index and some comparison between these two on the recommendation of prof. Demmke.

On the scientific side. The main research question of this thesis crystallized is: can benchmarking benefit from the appliance of complexity thinking?
There are some assistant/sub research questions. Such as: how benchmarking results are formed? More elaborately, what are the background reasons for having certain outcomes in a benchmark-survey? In the case study-side sub-research question is: Why was Finland so successful in the Happiness-Index Survey? From the complexity theoretical side, the research question is: can complexity be used as an assistive tool to process vast amounts of data emerging in the benchmarking surveys? And can it be even applied when analyzing the results of benchmarking? The method used is analytical literature overview.

2. BENCHMARKING

2.1. Benchmarking as a phenomenon – background, history and ideology

Benchmarking is generally considered a relatively new phenomenon that has emerged hand in hand with the rise of public policy's paradigmatic change to governance. And indeed how suitable for subtle state-originated guidance in power-networks socially-structured inquiries may be. Thereby the larger march to public acknowledgement of benchmarking has been placed in the late 1980's and beginning of the 1990's. (Kelley & Simmons, B.A. 2015, p.2; Ringel & Werron 2020. p. 137). However early ancestors of benchmarking go longer back to history even up to the early Romantic era and actually were suited to entirely different spheres of abstraction as in today -- comparison of art and artists to be precise. Benchmarking was used on various art styles such as painters, composers, actors and writers but mainly in the visual arts. The first of all benchmarks was considered Roger de Piles *Balance des Peintres (1708).* It evaluates European painters from various epochs based on four criteria: composition, expression, coloring and design and puts them to an order. It even has a scale, although from 0-20 unlike the 10-step or 5-step ones we are accustomed to know. However this ranking does not have all characteristics of a modern benchmark. It is ranked in alphabetical order rather than in hierarchical order. It neither possesses quantitative measurements or overall score. (Ringel & Werron, 2020., p. 147.)

Some sort of an arts' comparing-boom was launched and during the course of the 18th century many other art-comparisons followed although constructed in a relatively similar way: without strict hierarchical order and basing the ranks on overall score. On the other hand many critiques also stated that such abstraction as arts couldn't be strictly ranked or evaluated on purely technical bases for they contain spiritual values and subjective experiments of the perceiver. Not to mention that these pieces of art contained some unique pieces of craftsmanship. Sorby introduced the first "proper"

arts-ranking in 1810, although with somewhat little resonance to it. Same destiny also awaited Schubart who enhanced Sobry's ranking. Various critics evaluated these in terms such as "foolish", "absurd" or "odd". In general characteristic to arts-critique was that they were highly elitist and at hand of a small group of professional intellectuals. As a result of the industrial revolution and mass production the situation changed, though. (Ringel & Werron, 2020, p. 147-150.)

A more popular approach to larger audiences of early benchmarking adaptation was amongst sports during the late 1800's. Namely the *American Baseball-Cup* in 1870's and relatively soon onwards *British Football League* where league tables were established in 1888. As a consequence serial forms of competition and the league system itself spread around the world changing the character of sport competition from single matches or knockout-tournaments into a league-system aiming to determine the league champion during the games season, typically one year. This also increased the number of participants to a typical amount of between 10 to 30. And now it was Sports Clubs that were competing against each other. This subsequently enabled modern ranking to take place between the clubs (or teams). And a tool for comparison was introduced: up-to-date league tables. Moreover a ranking for players too was invented. To summarize: already at the end of the nineteenth century benchmarking (or the drift to find sports champions) changed the core essence of sports for times to come and simultaneously raised the public interest in constancy of performance levels during long periods of time. (Ringel & Werron, 2020. p. 153-156.)

Benchmarking was now open to masses. But in order to profoundly understand the sociological background of benchmarking we need to have a look at the ideological developments behind the emergence of benchmarking. Many academicians indeed place roots of benchmarking into the Anglo-American cultural hemisphere and the larger emergence of benchmarking into the mid-to-late nineteenth century. Benchmarking is generally considered to be ranking between actors, products, factory workers, pupils, artists, athletes and later on more abstractly business, hotels,

restaurants and on public body side hospitals, universities and even nation states. So what would be more suitable to the competitive ethos of the fordian era than to compare productivity and quality of products and services or the human actors behind them. (Ringel & Werron, 2020, p. 137-138.)

One approach of benchmarking-like rating could be: a conduct in which variables on a multitude of various possible measurable or quantifiable characteristics are compared and put into an categorical order. These variables can be i.ex. business environment, budget transparency, aid transparency and economic freedom. Or simply just products and services. These rankings are made to ease the consumers' pain of difficult choices in a world filled with options and abundance of choice. (Mau, 2019, p. 10; Cooley & Snyder, 2016, p. 2.) On the business side some of the best-known and most prestigious benchmark-ratings are credit ratings for sovereign countries. The "three tenors'' here are rating agencies called Moody's, Standard and Poor's and Fitch. Other significant economical benchmarking are The World Bank's Ease of Doing Business Index (DBI), Transparency International's annual Corruption Perceptions Index (CPI) and Global Competitiveness Index. One might say that these RROs also Ratings and Rankings Organizations have a great deal of not only financial but also public power which will be examined more closely further on. (Cooley & Snyder, 2016, p. 2, 52.)

As the academic point of interest for administrative scientist is the public sector side of benchmarking within sophisticated surveys that can be used to measure state of civil society, democracy, levels of human development, quality of life, media freedoms, environmental performance and aid transparency amongst others (Cooley & Snyder 2020, p.2.) We shall nevertheless take the states' credit ratings into a more thorough scrutiny. This system is used to assess the financial performance and amount of sovereign debt (per GDP) and give a label of state's financial capability and incurring of state-dept. States are assessed and compared based on massive amounts of economical data. After the comparison is done states are given a rank expressed via letters from A to D where A stands for excellent rank and D very poor, in practical terms

``rubbish credit”. As a matter of fact the best rank is three A’s also AAA possessed by some of best-performing financial might in the world such as USA and Germany. Also Finland had for a long three A’s but the rating was dropped about a decade ago.

However the RROs are given tremendous power over the world's countries with this right to rank. One use of this power indeed is to act as a judge. In this case mostly judging the economical state of the world's countries. Secondly these judgements are based on careful assessment of monitoring. Thus RRO’s also act as monitors. Thirdly they have regulative power, for investment markets are sensitive to credibility changes. Therefore if the rater changes its evaluation on the creditworthiness of a country to one way or another it can have a huge impact over the interest rate of state debt and image of the state as a lucrative investment object. This occurs even though state bonds are generally considered very stable and reliable investment objects (varying from state to state obviously). Furthermore a brief psychological insight: forecasts can and even have the bad habit of becoming self-fulfilling prophecies. Hence the fourth of RROs functions: to exercise branding also creating images of its rating subjects. Fifth and the last being perhaps the most pragmatic: to claim ownership over issues (= national monetary funds) if needed. To confiscate government funds to secure creditors interests that is! (Cooley & Snyder 2016, p. 39-40.)

But how can the RRO’s have invested such a huge public power? Shouldn’t this rating be done by the state? Perhaps yes, for mastering and guiding international monetary functions and especially sovereign bonds should be done either by very stable banks or state treasuries. On the behalf of the benchmarking side also the gathering, measuring and analyzing data, not necessarily. That can be done by virtually any research oriented actor. It might even be so that an Non-Governmental Organization (NGO) has more objectivity to rate and rank public bodies. But then again economic and moreover moral liabilities can be a bit more complex than that. Ultimately the transfer of such an enormous amount of power is based on a governance doctrine called New Public Management (NPM). (Cooley & Snyder, 2016, p. 40-43.) The following characterisation

of NPM bases on a publication of Hyyryläinen ``Sopimuksellisuus, talous ja johtaminen" originally written in Finnish. Translation into English is made by the writer of this thesis and might not be perfectly accurate compared to a professional translator.

"NPM is a normative theory of how the government can use a new tool to get things done -- namely contractualisation" stated by Laine. Nearly all western countries have reformed their governance and service production machinery since the beginning of 1980's. Same tendency has then moved onwards to second and third-world countries or emerging and frontier markets as Cooley and Snyder prefer to call them. This nevertheless has happened with the support of the World Bank. This school of thought partly as an ascendant of the 1970's energy crisis emphasizes scarcity of public resources and therefore demands effective usage of them on all levels of public government ranging from municipalities to state administration (and why not federal government in federations like the USA). Socio-culturally NPM emerges from the Anglican-German cultural hemisphere where efficiency in governance has prevailed already from bismarckian times. Modern mentors of NPM lie within anglo-american conservative state-leaders such as Regan, Thatcher and Major. (Hyyryläinen, 2019, p. 44.)

Crudely and a bit controversially NPM can be regarded as an administrative doctrine that's main feature is to establish private sector or the financial life leading practices into the public sector. Aim here is yet again to make public government and service production more efficient. Perhaps the leading ideology behind NPM is neoliberalism. (Hyyryläinen, 2019, p. 44-45.) Neoliberalism is according to the Stanford Encyclopedia of Philosophy and accepted interpretation by a majority of scholars ``the philosophical view that a society's political and economic institutions should be robustly liberal and capitalist, but supplemented by a constitutionally limited democracy and a modest welfare state." Neoliberalism is considered a distinctive and coherent political philosophy based on principles, political concepts and policies formulated by thinkers such as Friedman, Hayek and Buchanan. (Stanford University, 2021.) Also Mau points

out the rise of neoliberalism as one of core background ideologies in the contemporary metric society. Mau also pinpoints market supremacy, underlined competition to foster growth and innovation and preference of private-sector solutions over public ones as neoliberal characteristics. In practice neoliberalism has also inflicted privatization of public enterprise, deregulation projects and various reforms in order to cut within administration, education and fiscal operations. (Mau, 2019, p. 22-23.)

Some of NPM's leading principles are the rule of democracy. Secondly administration that derives its right authorization from law (=legality principle) thirdly effective quality management. For example in Finland the most prominent occurrence of NPM appears in quality management. (Hyyryläinen, 2019, p. 45.) Indeed some attitude inherited from early managerialism is that good leadership can give a significant contribution in effort to solve economical and social problems. This may be more difficult to be verified in practice. Nevertheless characteristic to NPM leadership is that governmental entities responsible for policies or policy formulation are separated from executive entities aka. agetification whereas administrative units ordering services are separated from units that carry them out aka. purchaser-producer models. Also more attention is highlighted to alternative ways of service production, distinctive profit-responsibility, improvement of cost-effectiveness and general increase of managerial skillfulness. (Hyyryläinen 2019, 45-47.)

Within sectors of the society there are in general two sorts of law: public law and contractual law. Public law is considered one-way. That is rules, regulations and orders from the sovereign that can either be a monarch, dictator or in democracy a public-post office or authority to which or whom part of public power has been delegated according to Montesquieu's separation of powers principle or some other constitutional norm. In modern western countries certain bureaucracies and civil servants use public law. Alteration to usage of public power can be sought only via legal complaints or appeal either from the public body oneself or from a competent court. (Mäntylä, 2020.)

Private justice naturally has many branches but generally it is considered more based on contract law. That is two or more contracting party/parties mutually agree to something and make an agreement about it. These contracting parties can be legal persons, financial institutions like banks or companies as well as public bodies within matters that are not under public law by nature. Possible disputes are primarily attempted to be solved via negotiations between the interested parties. Help of solicitors may be used if needed. And justice ultimately from lower courts. (Mäntylä, 2020.)

Why so long lectures about constitutionality? Because the core essence of NPM is to bring more law of contract into actions of public authority. Of course even NPM does not replace the “hard” legal basis of public law that enables use of coercive powers by military or law enforcement in a weberian sovereign state. Or even means of compulsion within decisions that affect citizens rights and duties. However matters that can be agreed and negotiated without the must-tone of public law are implemented to governmental actions via NPM. One could say that NPM makes power usage softer and more mutual understanding- based and even more on communication between citizens and the users of sovereign power. (Mäntylä, 2020; Hyyryläinen 2019, p.46.)

One essential part of NPM is the client-agent theory. There obviously are client-agent relationships that are problem-free and the agent acts exactly according to clients assignment or there is no conflict between clients’ and agents’ interests. However the client-agent theory has focused on those relations or situations in which all agent actions are not controllable by the client and concerning the results outcome agents contribution is absolutely crucial. Therefore the client-agent relations are scrutinized as potentially realizing problems or as lack of confidence. The contracts are assumed to be broken both because making a perfect surveillance system doesn’t exist but the self-interest of both parties does. The fundamental problem can be condensed into a

question: “if both contracting parties are benefit-maximizers there is good reason to assume that the agent does not always act entirely according to clients’ interest”. (Hyyryläinen, 2019, p. 59-61.)

There are two main lines of agent-client problems. Firstly are the hidden action/ moral hazards where risk-take is highlighted. As if the one with better insurance takes bigger risks. Second line is hidden information/ adverse selection where the agent possesses pieces of information that the client lacks. In research it is assumed that the agent exploits this asymmetry in information to self-benefit. On the other hand, this applies also to situations where the client has better sources of information than the agent(s). Client-agent problems cause extra costs entitled as “agency loss” and “agency cost” Client cost are due to motivating and supervising the agent. Whereas agent costs are due to commitment to the company/client. Primarily client-agent theories were adopted into examination of the relationship between the owners (as clients) and executive management (as agents). But already in 1970’s it was observed that a wider appliance of this theory can be used. (Hyyryläinen, 2019, p. 61-62.)

Now here is also the lengthy and somewhat complicated answer to the power relations of credit raters. They actually take care of a public function that could also be taken care of by a public body, presumably ex officio by the world's hegemonical financial might -- which currently is the United States of America. USA also happens to be the father of NPM, a doctrine, core of which is to increase outsourcing, business-like principles in governance and contract law. So actually based on an agreement of outsourcing between the US Federal Governments and the trio of credit-raters they have assumed this part of US governments functions a little like freelancers. And actually it is quite handy that the US government doesn’t have to dirt its hand with sovereign credit rankings if something goes wrong or the placings alter dramatically. Furthermore this might be an ultimate example of a benchmarking survey that is made by private corporations. (Cooley & Snyder, 2016, p. 40-43.)

2.2. On the quantification of social realms -- insights to rational worldview and development of contemporary markets

There is one tectonic plate that has had to be put into a right place in order to enable the triumph of benchmarking, namely quantification of the society. Generally sociology as well as measuring the society has been considered qualitative amongst the so called “soft sciences” amongst which sociology includes. One of the key theorists of society's quantification, Mau has pondered the matter very profoundly in his work “The Metric Society” In his analysis quantities always means measurable and calculateable amounts. More deeply defined this means expressing social phenomena and stately matters in an abstract general and universally accessible language, the language of mathematics! (Mau, 2019, p. 12.)

This however requires that observations, insights and qualitative judgements are transferred or altered into numeric values. Quantification therefore in a way reduces “complex and confusing world to the standardized language of numbers in which there are clear proportional relations between large and small“. This might turn out to be somewhat of a difficult task for there are different ways of perceiving social phenomena or talking about them. In epistemology this is called the subjective observation-bias. At least the mathematical and physical way of perceiving the world understand that numbering makes phenomena at disposal more precise, variable and neutral having one-to-one correspondence and simplicity. (Mau, 2019, p 12.)

This on the other hand might not be such a novelty for social and administrative scientists who have over the decades struggled with the problematic nature of operationalization of the research questions and defining the measuring scales. In natural sciences the SI-system is a widely accepted paradigm (even in Anglo-American countries). It gives directly the framework within which to do research. So as a matter of fact this seems like a tendency to incorporate natural sciences’ methods into human sciences. Whether this is a good or a bad thing (if it even can be seen on such a

black-and white scale) is an entirely different discourse. Perhaps some qualitative aspects of studied matter are left to lesser attention. On the other hand it may very well be merely a sign that the model and arithmetics behind it aren't yet in a particular case honed to sufficient accuracy and therefore require adjustment into calculating models and processes.

Mau nevertheless responds to this critique by stating that quantification of social phenomena procedurally deliberately takes away local context and knowledge of social practices in order to generate more abstract and mobile information that can be put together or "melted" with information from other sources. Moreover such numbering practices tend to have certain procedural, objectivity and quality-demands generally accepted by the scientific community. Therefore results are expected to be determined by the research processes rather than people. And doesn't all that sound very scientific after all? (Mau, 2019, p.13.)

Many public bodies actually see quite a lot of effort in order to maintain their numeric data "clean". And that is for a good reason for all numbers in public discussion are more or less operationalizations of the social and they have to be considered correct (or at least as close to truthful as possible). There are a myriad of ways in which society attempts to keep their numeric data in good order, for example making extensive legislation related to statistical affairs, participating in international data-monitoring systems, creating statistical authorities and even elaborating standardized reporting systems. As it seems global social pressure in order to have numbers accurate is high. (Mau, 2019, p.13.)

Although numbers may very well be the answer for needs for rationalization, objectivity and relevance they are more than mathematics. Therefore care must be addressed to who makes the numerilization and how. It might sound pretty philosophical but as Heintz puts it "Statistic claim to demonstrate a reality which exists outside of them and is rendered visible by them" However numbers are not exact

replicas of the reality but rather “selective constructions” Thus making objectivity of numbers an attribution instead of a sole fact. Needless it may be to say that archetypes and ideals address people. And in fact numbers or inquiries containing them have first caused debate even battling over their relevance. And thereafter received approval. Even to such an extent that well-known ratings have caused social crises at worst even mild economic up/downturns and triggered public anger or contentment. Therefore politicians have to be pretty careful when dealing with numbers or statistical results in public. (Mau, 2019, p.13-14.)

Also Mau highlights the economical and stately development as background forces worming way to emergence of benchmarking. The so-called spirit of enlightenment era -- rationalization and utility -- has significantly affected the mindset behind benchmarking. Some classics of sociologists that May highlights are unsurprisingly Weber, whose ideas of separation between domestic and business spheres as well as rationalization of production have been essential for the social and financial world to become curved into the era of benchmarking. Somewhat less known landman of Weber, however, a prominent economist and sociologist, Sombart considered introduction of double-entry bookkeeping as a key step in establishing contemporary capitalist economic order. In his main work *Modern Capitalism* (1919) he also stated that this method of bookkeeping is not merely a technical achievement but also reflects the rational mindset of contemporary western society. (Mau, 2019, p. 14-15.)

Third classic theorist of this era worth mentioning is Taylor who developed the system of scientific management. Taylor’s functional theory of factory management consists of splitting each job or manufacturing process into pieces eliminating all unnecessary motions and hence increasing the production efficiency. Furthermore, providing the worker with sufficient training and tools increases productivity. Taylor has naturally been criticized over seeing working processes too mechanically. (Britannica, 2023.)

But as a product of his era, a vast majority of modern industrialized work was manual labor. This was indeed nearly a century before the invention of robots. But approximately a century after Smith had published his legendary cornerstone of capitalist theories: Wealth of Nations in 1776. Briefly put: pretty much everybody has heard the expression of the “invisible hand” and perhaps even the comparison of “the butcher, the brewer and the banker”. The key point of Mr. Smith was that trade, competition and division of labor is financially beneficial. Moreover Smith argues that the one who pursues his/her own good on competitive markets is “led by the invisible hand to promote public good”. (Arizona State University, 2023.) Whether this metaphor hangs up to modern comprehension is an entirely different question.

Both Weber and Sombart highlight that economic systems are characterized by socially highly developed intellectual calculative practices in which also the numerical medium (money) plays an important role. When the value of a commodity (physical, spiritual or service) can be measured in money, productive trade relations can be established. (Mau, 2019, p.15.) Perhaps this is why the European Monetary Union has a tremendous asset in standardizing its members’ mottled national currencies into a commensurable one. Not to forget that ultimately the US dollar is the currency into which all international trade sums are converted or at least compared to. Nevertheless Sombart’s and Weber's idea was also that standardized weights, length and various other variables were a prerequisite for international trade to flourish. Since trade is heavily based on trust between the business partners and even inside the same company, usage of heterogeneous measures easily created the image of foul play. And once trust is lost, it is difficult to re-establish. Same applies to sociological research, at least to some extent. (Mau, 2019., p.15-17.)

Mau also points out that states or public bodies have used statistical documentation, classification and accounting for a long period of time. The development of first fiscal states and later on democracies have virtually always needed the ability to count people and categorize them based on sociodemographic criteria. As a matter of fact

scriveners were an important profession already during the king-ruled ancient times. Originally the German *Staatswissenchaft* was far less numerological than it is today. Nevertheless numbers and rankings have always played an important role in stately affairs. Inflows, outflows, quantities, differences have had to be traceable to the rulers. Also topologies or classifications of the population based on for example such criteria as age, gender, socioeconomic status, ethnicity or estate have attempted to organize the occasionally complicated social realms. Statistics as well have been important for decision-making as they provide abstractions of reality. The dangers lie in oversimplification or viewpoint becoming too narrow though. (Mau, 2019., p. 17-19.)

But coming back to modern times after this history anecdote. The Second wave of quantification or "social engineering" rose its head during the 1960's and 1970's as politics became more rationalized and the importance of statistics rose. This scientific approach aimed for larger and more long-term politics as well as decision-making tools to pre-evaluate effects of political decisions. The social indicators-movement is worth mentioning here. Its aim was to create sets of indicators for measuring and evaluating social progress as broad as possible - not only the economical ones. Although there have been setbacks, the endeavor of measuring social progress has remained. (Mau, 2019, p. 17-20.)

The most postmodern tendencies within quantification have widely involved digitalization and digital data-collection techniques. As a matter of fact the amount of new rating procedures, indicators, polling instruments and performance measurements are "eating the breathing air of privacy from around us". Indeed, highly advanced ICT, algorithms and smart technologies (smart cities, smart homes, smartphones to mention a few.) have made collection, analysis and storage of data remarkably faster and easier. On the market side of this, it is easier to steer peoples economic decisions, win (new) customers and define people's commercial utility. But on the human side there probably lives more than merely "homo oeconomicus" within us. It seems quite bewildering if even areas that we have accustomed to belong to our private sphere of

life such as family relationships, emotional states, habits of behavior and hobbies all of a sudden become digitally measureable and part of business-marketing protocol. To condense this all. As the saying goes: data is the new oil! But what are we fuelling with it one might justifiably ask? Mau points out the concern that already Orwell presented; that the answer could be something as insidious as a network of data-based fiscal surveillance-states. (Mau, 2019, p.20-25.)

2.3. Definitions and characteristics of benchmarking

So, this is the neolibertarian financial and administrative worldview with NPM and quarreling client-agents and society being metrically measured that was set as a stage for emergence of a benchmarking school of thought. But what actually is benchmarking? It is a practice of comparison and ranking. A highly developed and sophisticated analytical comparison consists of four characteristics: firstly benchmarks are comparisons of performances. Secondly they are repeated either regularly or at least periodically. Moreover the results of these rankings are published and hence exposed to critique of the scientific community. Thirdly their core essence lies within quantification. Fourthly, results of the rankings are somehow visualized. Usually benchmark results reports contain graphs, oftentime very elaborate such. These put together contribute to social construction on fields of competition. (Ringel & Werron, 2020, p.137, 141.)

To further this definition a little. Comparison of performance consists of making entities the survey contains comparable in the same category. Therefore the matters compared in a benchmarking survey are defined relatively similarly if not even identically. Differentiation of results according to some additional criteria is also possible. It may though be noteworthy that many social processes are very complex and straightforward categorisation might lead to biases of various sorts. There has been critique concerning that there is nothing natural about social comparisons. More

mild tones emphasize that social practices can be captured in scientific form by applying or creating the right measuring instruments. (Ringel & Werron, 2020, p. 141-142.)

Furthermore it is in place to ponder whether benchmarks are zero-sum comparisons or not. Ringel's and Werron's view that they usually are not. This is because that quality ascribed to one entity is not directly related to other entities in the same category but are rather abstractions from reality that have been condensed in a comparable form. Therefore for example development of a country, outstandingness of a university or greatness of an artist doesn't automatically come at expense of others in the same category (unless in realms of competitive markets where this happens). However qualitative zero-sum comparisons are a bit risky for they can end up into potentially endless debates of the contains (even though quantitative ones as well but these discussions relate more on the operationalization conversions of the qualitative world into quantitative). But it may very well be that rankings aren't automatically zero-sum games but this is more case and category-dependent. (Ringel & Werron, 2020, p. 141-142.)

We live in an increasingly visual world so a few words about visualization in benchmarking surveys. There is a saying in Finland that "one picture tells more than a thousand words". And that is actually the role of visualizations in surveys: to condense the results of massive data amounts into a condense and lucrative form that is convenient and easy to read. Of course great care has to be addressed in the form of how results are visualized in order to keep the graphs informatively accurate. They also are performative elements in rankings. Oftentimes in the digital era these pictures and result rapports are in digital form and available via the Internet. A typical form of presentation is hierarchical descending order starting from the best performers on top. (Ringel & Werron, 2020, p. 143.) A traditional writing model called "cream on the top". At least theoretically the order could also be exactly the opposite, showing the "worst" performers first and then ascending to the top ones. A writing model called "sugar on

the bottom" on the other hand benchmarkings often aim to find the best performers and practices from whom others can learn so in the world of constant lack of time of busy readers perhaps it is more practical to present the top-ones first. And this ascending order might as well seem somewhat irrational.

From the point of view of social operations rankings also require the element of repeated publication, whether this is in real-time, daily, weekly, monthly or annually. The point of ranking is to bring the results of rankings available to larger professional and general audiences. Publicizing stuff to audiences can as a matter of fact be seen as a relatively complex social process per se. Especially within larger audiences the aim is to institutionalize the ranking and receive social prestige/status to it. So actually there are two performative elements within rankins: regular publication and visualization. There however is a certain nuances difference between a rating and a ranking. Ratings are perhaps a more broad concept which consists of virtually all quantificially made and visualized comparisons even those of academic purposes. They primarily strive to understand some phenomena or placement of individual actors within ratings. Rankings in contrast tend to have virtually always a more competitive and financially balanced character. The aim is to understand the market or general economic (or at least hard substantial) placement of some product or person's professional contribution within markets. Moreover the economically competitive nature is considerably more present within rankings. (Ringel & Werron, 2020, p.143-144.)

Benchmarking is actually a special sort of category within the larger pool of rankings. Benchmarking can, but it doesn't necessarily have to contain all four elements of rankings mentioned before. Benchmarking is very accurately conceptualized:
"Benchmarks define a standard based on the top result in a population of organizations, and they test the degree to which these organizations' performance corresponds to the said standard. By identifying a number one, benchmarks aim to motivate all other organizations to catch up." Thus benchmarking is a process in which outstanding practices and ways of doing things are identified, studied, and adapted

from one organization to another. Benchmarking can take place geographically anywhere worldwide. It can happen between two organizations in the same financial area or a country, but as well between organizations in many countries. (Ringel & Werron, 2020, p.144; Tuominen, 2016, p.6.) Indeed even countries – or at least their public governments can be benchmarked (Blavatnik School of Government, 2019b). The aim of benchmarking is to improve the performance of an organization (Ringel & Werron, 2020, p.144; Tuominen, 2016, p.6).

A benchmark means a top-rated way of an organizational practice, the best in its category. A benchmark can be considered as a reference for comparison as well as measurement standard. Typical for a benchmark is to be recognized as the outstanding standard within a specific business or administrative process. (Tuominen, 2016, p.6). Therefore, benchmarking can also be regarded as the practice of comparing processes and performances to reach best practices from other benchmarks (Demmke, 2020b). A potential example of a benchmark within the car manufacturing branch could possibly be Mercedes Benz.

This leads us to a more industrial point of view of benchmarking. It has similarities with lean thinking. That is benchmarking to be deemed as a continuous process where corporate processes per se, services and products are measured and compared against industry or branch leaders and even companies' worst competitors. (Tuominen, 2016, p.11.; Demmke 2020b). Benchmarking also progressively improves one's organization's performance via new procedures and ideas adopted from rival companies (Demmke, 2020b). For example, fresh orange juice squeezers were first introduced in widespread commercial use by the S-market chain. It did not take long for K-market chain and Minimani to benchmark this audience favorite service and thereafter freshly squeezed juice was available on all three major grocery store-chains in Vaasa

Why does benchmarking fit our times? There probably are many reasons for this. In economics, market liberalism and individualism have as aforementioned become the

prevailing doctrine in the western countries. Also, in the public sector, the leading school of thought in the OECD countries is the New Public Management (NPM) which emphasizes profit responsibility (Hyyryläinen, 2004, p. 44–54.) In this wide context of growing competition, benchmarking simply fits the popular zeitgeist. More deeply pondered it seems to be quite profound in human character to compare and compete. Seen positively, comparison can be regarded as the desire to improve, imitate good practices, and enhance performance (Demmke, 2020b.)

2.4. Examples of popular benchmark-surveys around 2020's

It seems to appear that the popularity of benchmarking has significantly increased over the past decade. Political, economic, and academic interest over the subject has been of much scrutiny. For example, when the first World Happiness index was published in 2012 the demand soon came so great that the decision to annually release a report discussing this topic emerged. (Helliwell, etc*, 2022*. p.3,7.) Also, the first International Civil Service Effectiveness Index, (InCiSE) was introduced at the end of the decade, 2017 second following at the turn of the decade in 2019. (Blavatnik School of Government, 2017 & 2019a.)

And no wonder politicians and even high-ranking civil servants can even become overly satisfied if their beloved country has done well in these international benchmarks. In addition to political pride some may even turn sharing their advice on how to make a successful country into a successful business. Not to mention that universities or institutions of higher education compete to attract an increased number of students – especially the brightest ones – partly with the power of excelling at international university benchmarks. On the other hand, low rankings in benchmarks can cause feelings of anxiety, distress and failure in countries or organizations ranking low on these surveys. In a more constructive sense perhaps build the courage to make positive change within the country/organization to increase its living standards and as a consequence its ranking in the surveys. So, there is a high demand for neutrally made

and objective international benchmarks. However international benchmarkings may prove out to be more complex than they surficially appear.

3. COMPLEXITY

3.1. Classical definitions of complexity

The term “complexity” comes from the Latin word plexus that has been derived from the latin verb “plectere” which means braided or intertwined. Complexity as a refinement means interconnectedness or interdependence. Oftentimes complexity is associated with the interactive character of interconnected organizational activities and wicked problems emerging from there. Operational dynamics of an organization is also often associated with complexity. (Vartiainen, etc, 2013, p. 52.)

In complexity-thinking there are seen to be two sorts of problems: tame problems and wicked problems. Tame problems are generally considered as straightforward problem solving like doing a mathematical calculation or finding a solution to a logic problem. Tame problems are characterized by six traits: 1) there is a problem that is profoundly and permanently defined, 2) there is a clear definition, 3) there ultimately is a limited amount of solutions although this amount can be very large, 4) the flawlessness of the denumenet can be objectively assessed, 5) once you have learned the rules of the game you can always play the game with same rules and 6) you can always restart the problem. Tame problems are of character every-day routine challenges. For example, if you run out of coffee-milk, go to the grocery store, and buy more milk. Tame problems are not always easy, but they are straightforward, and they can be solved with expertise. Usually tame problems can be solved via established processes and procedures. (Vartiainen, 2013, p.20.)

Wicked problems then again also have six most characteristic traits, but they are different to tame problems. Firstly, weighing the possible solutions helps to understand the problem itself. Secondly, the solutions are situational, not final. Thirdly, there are no right or wrong answers to wicked problems and the value-giving to "rightness" or "wrongness" of the solutions is at least somewhat subjective. The aim is to merely find a solution good enough to live with. Wicked problems are unique and new by their character. Usually, the specialness is the most common feature of a wicked problem although all wicked problems tend to have at least something in common. Fifthly, the suitable courses of action are always unique. Most of these problems emerge from social laws and interaction and therefore cannot be solved entirely via arithmetical methods. Sixthly and lastly, there are multiple suitable alternatives as solutions. Some of the solutions are more obvious and effective than others. Typically, not all solutions can be pondered and not all consequences can entirely be considered in advance. (Vartiainen, 2013, p. 22-27.) There have been many relatively colorful expressions used to characterize complexity at least in the research tradition of Vaasa University. Linear–non-linear/bi-polarity is one crucial separation between tame problems and wicked problems. Tame problems can be overcome with straightforward approaches. Wicked problems on the contrary require most often a more elaborate and multi-dimensional approach in problem solving.

Initial value sensitivity is again one of the major discoveries by weather scientist Edward Lorenz in the 1960's. Sensitivity to initial values or conditions means that a system's behavior may alter significantly even in smallest differences in conditions thus making them very difficult to predict. That also means that even predictable (or deterministic) systems can become surprisingly unpredictable (or indeterministic). Also, dependence on initial conditions or/and values means that the longer in future we go with our predictions the more inaccurate or even chaotic they become. And partly therefore we curse the meteorologists after evening news for being again "wrong". (Future Learn, 2022.)

Self-organization in complex systems is a characteristic originally introduced in cybernetics and system science. It means that a certain group of self-organized systems, also SOS or off-systems can change their internal function or structure as a response to external circumstances. Parts of SOS can reorganize or affect the organization of the same systems. This aims to stabilize the structure or functions of a SOS against external disturbances. In system science the self-organization of processes is achieved through increasing the system's internal time-space complexity to generate either hierarchical or multi-layered structures or functions. (Banzhaf, 2009.)

Faces of Janus is one metaphor used to describe ambiguity. Janus was of its historical origins a very popular Roman god who was worshiped on many occasions that brought change into life in example marriages, seasonal change in agriculture at the beginning and end of wars. Even the first month of the Julian calendar was named after Janus as January. Special about Janus was that he was pictured with two faces: one to see the future and another to see the past. (Bellesguard-Gaudi, 2021.) Cambridge dictionary defines ambiguity as a fact of something having more than one possible meaning and therefore possibly causing confusion (Cambridge Dictionary, 2022a). Applied into complexity thinking both the term ambiguity and “Faces of Janus”-symbolism illustrate that same phenomena or situations can have several approaches for various people and groups to see the same matters from even completely different points of view. (Raisio, 2013.) A more Bayesian way to phrase this: even with the same evidence, different priors will produce different models.

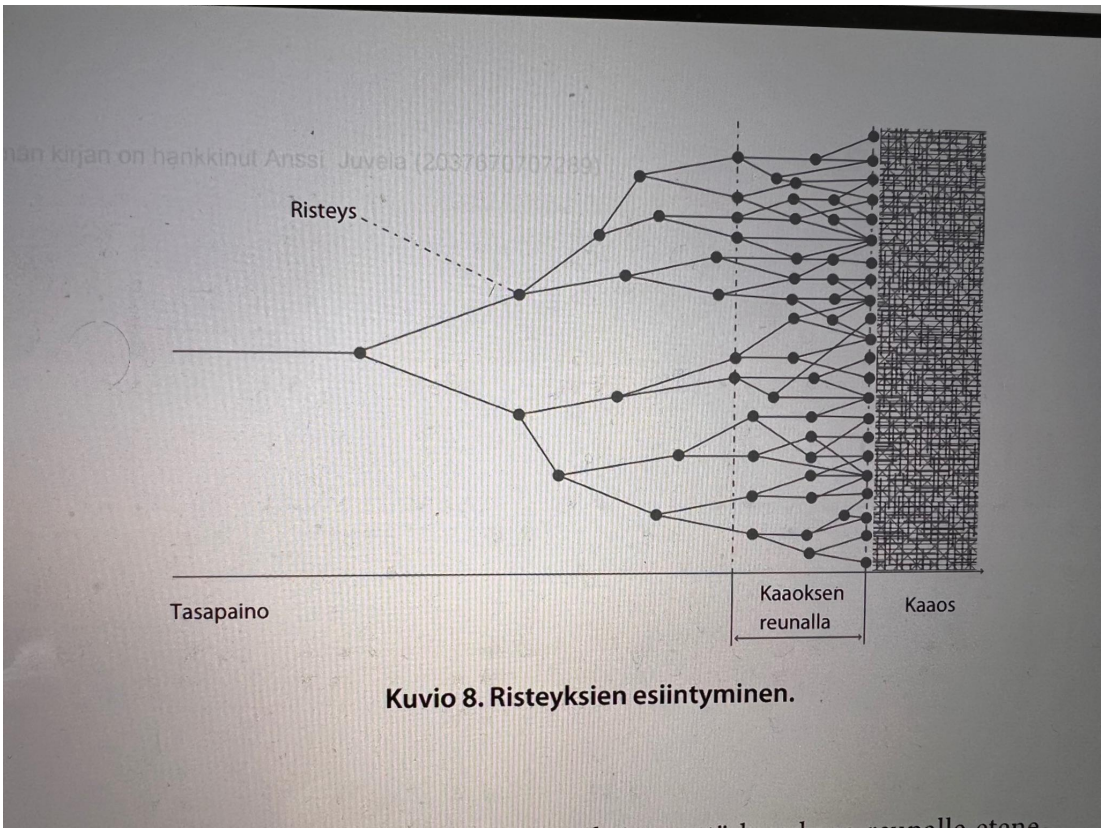
The living present as a complexity-concept emphasizes two main points: firstly, events and organizational life even with its renewal processes are binded into time. Secondly, it appears more prolific to think that no organization reaches an everlasting and stable situation (and presumably that is not even the goal). The present moment is seen as an anchor point on the bridge between organizational past and the future progression. (Ollila, 2013.) Obviously, it has to be acknowledged that the present is constantly moving on the timeline of existence and therefore even the present moment is living.

There are many forms of feedback processes. In everyday organizational activities they are seen as processes of personal and operational enhancement. Feedback-processes in complexity thinking are commonly seen as an outcome of information change between actors in a complex system and its surroundings. There are two types of feedback: positive and negative. Negative feedback processes tend to put pressure on conserving the established state of the system of situation (status quo). This can also cause change resistance. Positive feedback on the other hand highlights questioning of the current balance-situation. Usually this contributes towards better action in the whole of the organization and encourages change in co-operation. (Lindell, 2013, p.76.)

Joint evolution means that development in an organization is based on open and responsive interaction. That means an organization not only adapts to its environment but also shapes the environment by its own activities. Also, a holistic approach is adapted as one change in a sub-area affects the entirety and correspondingly the big picture affects the sub-categories. Since this interconnectedness often is somewhat ambiguous, it seems important to find these inter-relations and attempt to take them into consideration within change management. (Lindell 2013, p. 70.) Interaction is often associated with joint evolution. In an organizational interplay one member or group in the organization notices a matter that requires reaction and takes it up. Also, intra-organizationally there is interaction between the individuals in it and the organization itself. (Raisio, 2013, p. 63, 69.)

A closely related but not entirely the same concept to initial value-sensitivity is path dependence. In brief path dependence means that organizational endeavors and problem-solving modules are affected by their past in addition to the present moment. Previous decisions can open paths to new possibilities but on the other hand conventional habits and thinking models hinder, stop (or paradoxically it may also in some cases enhance) adaptation to new ways of doing things. (Lindell, 2013, p. 82–83.)

Space of possibilities or window of possibilities refers to a condition where many possibilities of new modus operandi or creation of novelty products/services are opened by change. Sometimes the timeline for spaces of possibilities is narrow, other times it may be wider. A classic example on space of possibilities is for a blacksmith to forge iron when it has been heated to the point that it is flexible and thus alterable. (Ollila, 2013, p. 137–138.) Bifurcation point, point of no return or just simply "cross-roads" is used to describe a situation in which organizations (or individuals) behavior turns into a new direction. A classic example of this is as a result of possibilities generated by globalization a vast majority of industrial production was transferred into low-cost production countries. (Ollila, 2013, 142.)



Kuvio 8. Risteyksien esiintyminen.

An illustrative picture concerning cross-roads, edge of chaos and full chaos, from book "As a Leader on the Verge of Chaos", p.82.

Tasapaino = balance

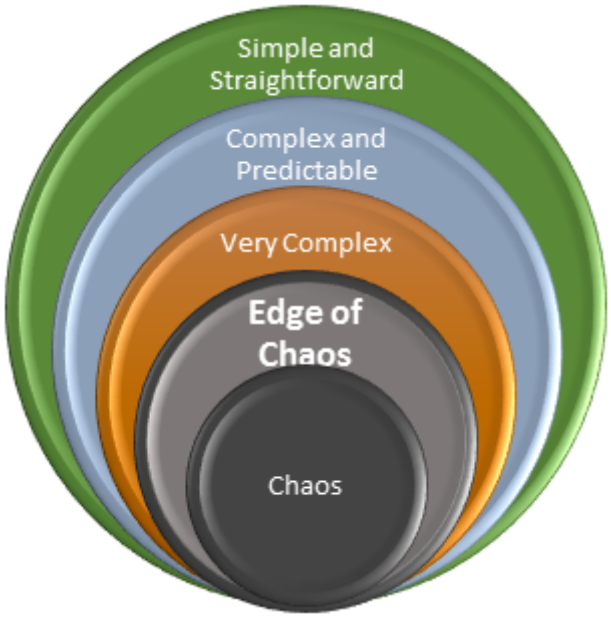
Risteys = crossroads/ bifurcation point

Kaaoksen reunalla = on the verge of chaos
Kaaos = Chaos

After a certain amount of bifurcation points, we enter the edge of chaos. “Edge of chaos refers to the transition between order and disorder, where there is complete randomness of chaos, where complexity is extremely high.” This situation can also be called “edge of order” (Balakrishnan, 2018). The point in which an organization is on the edge of chaos means that there are many interconnected crossroads in which choices have to be made. Superficially the edge of chaos can easily be considered as a negative phenomenon for it seems like a state that has gone or is about to go completely out of control. Nevertheless, more thoroughly seen it can also be considered as a creative and prolific state, in which a great deal of new possibilities can emerge – as long as the organization and people can be held in some extent of order. Obviously from the intensity of chaos, steps can also be taken backwards to return from chaos into its edge and even furthermore back to balance. Nevertheless, on the edge of chaos and especially in full chaos we can say that people and organizations are operating in a state far from balance. (Raisio, 2013.)

An example of increased complexity and chaos is war and the political tendencies preceding it. Normally states are well in order. For political reasons, often power-political engagement with obtaining natural resources, strategically important geographical targets, can cause political and economic bifurcation. Depending on diplomacy and negotiations, nations relations can return back to balance but if the negotiations fail the situation escalates on the verge of chaos. Usually, heavy intelligence operations and troop transportation follow at this stage. If the situation escalates even more and an attack-order is being given the situation enters the chaos of war, where virtually anything can happen. Unfortunately even a great deal of human lives lost, and a massive amount of property damage done. After the chaotic battling has raged for long enough, peace negotiations usually begin. When a cease-fire has been reached in them the hostilities end and troops are usually pulled back (in a good

scenario). After negotiations continue a peace treaty is made and compensation for damage or/and war-guilt has reached the nations relations slowly begin to normalize.



To summarize matters and situations from simple to increasingly complex and ultimately on the edge of chaos and full chaos. Picture from beeshankar.medium.com.

Emergence comes from the expression ‘to emerge’ which by Cambridge dictionary definition means “to appear by coming out of something or out from behind something” or to become known or develop as a result of something (Cambridge dictionary, 2022b). In complexity-thinking, the process means that especially in situations of change there are events and traits that spring up more or less surprisingly. Renewal processes per se may raise new kinds of situations, thoughts and contextual connections have the tendency that new matters will arise. (Vartiainen, 2013.) Perhaps this is due to some traits in human behavior: firstly, our long-term memory and even subconscious has surprisingly much data, thoughts and experiences stored into them. Secondly, usually there is a lot of wisdom also outside the executive management. Employees and people on the lower levels of organizational hierarchies tend to have a

lot of ideas and practical knowledge on how some matters are functionally carried out. When this silent knowledge is taken into use something new can emerge.

Adaptation is simply appliance of organizational practices either internally or externally. Complex Adaptive Systems (CAS) then again are basic units being analyzed in complexity sciences. It forms a neural network-like jointly collaborative and interactive entity in order to reach its goals. Typically, in these networks there is no central management, but it rather consists of independent agents. Action of these agents is based on its goals, and information implicitly received from its surroundings. Typically, these agents themselves are complex adaptive systems. CASs are formed in social surroundings relatively spontaneously; they learn quickly and are capable of rapid adaptation. (Pernaa, 2020, p.16.)

Paradox by a dictionary definition means a conflict or inconsistency within a person, action or situation that has seemingly contradictory qualities or characteristics. Paradox can also be a self-contradictory statement that at first appears to be true. In epistemology it can also be by appearing contradictory statements or statements that are opposed to common sense, but they still may be accurate. (Merriam-Webster, 2022a.) Since complexity and wicked problems are by their nature very multi-faced, paradoxicality can well be applied to them. Same problem may have several or partly even opposing traits. Moreover, a solution to a problem of this sort can have either paradoxical solutions or consequences.

3.2. Insight to History of complexity and complexity sciences

There are many theorists that have been associated with the development of various theoretical models associated with complexity, most famous of whom is Edward Lopez (1917–2008). He was the pioneer of chaos-theory with his '*butterfly-effect'* - metaphor although he was originally a meteorologist and a mathematician. Probably the original formulation was: "Does the flap of a butterfly's wings in Brazil set off a tornado in

Texas?" This has been in popular culture altered much from its original context and meaning. (Vernon, 2017.) The idea of Mr. Lorenz was to simulate that the movements in the atmosphere are tightly interconnected and a change in even one of the factors can cause a chain-reaction that will alter the eventual outcome. This is also called initial value-sensitivity. (Lindell, 2013.)

Roots of complexity-like thinking go even longer back to history. It was already a brilliant baroque physicist, Isaac Newton, who presented the idea of pre-deterministic clock-work-universe where the will of God has set all bits and pieces of existence into their perfect place. (Vernon, 2017.) The Lorentzian approach derives from the notion that laws and functions of nature are far more complicated than exactly-ticking Swiss-made universal clockwork. The butterfly-effect evolved into chaos theory in which mathematical formulas are utilized. Mainly chaos theorists attempt to explain phenomena that are pre-determined by their character. Chaos theorists believe that in superficially random-appearing phenomena profoundly a "higher degree of order" is concealed in the background. (Ollila, 2013.) Nonetheless, as chaos theory spread out to larger knowledge during the 1970's the groundwork for complexity sciences was laid. During the 1980's complexity-sciences expanded from physics to biology. During the first decades of the new millennium complexity-thinking has also occupied social sciences, sociology, and social psychology. Nowadays and especially in administrative sciences complexity-thinking has been harnessed as a tool for management and leadership in organizational studies. (Complexity Research Group, 2023.)

Although somewhat puzzling, complexity sciences are also often referred to as "complexity theory". This might cause some confusion for there is no one particular systematic theory but rather a collection of multiple branches of sciences that use systematic key concepts related to complexity-theory, for example physics, chemistry, mathematics, biology, neurology, ecology, and sociology. As methodologically in science there are two major categories of approach: quantitative and qualitative. If slightly generalized: quantitative method based on natural sciences, exact formulas,

statistical analyses, and computer-based modellings. Also, things that can be counted and calculated via units defined in a relatively similar manner universally. Qualitative method is more based on the philosophical and metaphorical aspects of complexity. Understanding the character of complex phenomena that is. (CRG, 2023.)

4) CONCLUSIONS AND DISCUSSION ON BENCHMARKING AND COMPLEXITY -- Can complexity help benchmarking?

Complexity-theory is a modern and versatile tool that has been adapted into many fields of sciences. It can help to understand complicated and even subtle reasons of causality in the increasingly more interconnected global world of 2020. Complexity as well as wicked problems emerge oftentimes from the social structure. In order to be able to quantify or make social phenomena and processes measurable highly advanced calculation models are needed. Complexity-thinking can indeed be of great assistance within the quantification processes for it can help to illuminate complicated causal relations that have an effect within defining the quantifical concepts per se as well as within actual arithmetical models used to calculate social quantities.

Furthermore in a globalized and digitalisized postmodern world financial, political and socio-cultural phenomena and interactions are increasingly more interconnected. Complexity can help not only within academic research but also within economical life. Trade and monetary systems after all are complicated agreement-based social apparatus where laws on social structure and negotiation apply and therefore also complexity-thinking is applicable. In the article within *Ranking the World* (Abdeal & Blyth, 2016) writers pinpoint that contemporary investment markets deal and trade with such complex investment instruments such as bonds, futures contracts, interest derivatives, invest/mutual funds and warrants that understanding complexity could also help economic actors to better understand the character of their trade objects and risks included. Idealistically and a bit exaggerated this could if not prevent at least

hinder future financial bubbles to be generated and therefore bring more stability to international financial markets.

But even in the world of increased complexity we must bear common sense in mind, for the mind can play its tricks and start to look for complexity also in places where it does not genuinely occur. Complexity should not per se be seeked or added in places where it does not belong. Linear progression and problem solving can still be applied in many if not even most contexts. However, benchmarking – and especially advanced benchmarking can benefit from complexity thinking. Especially the two worldwide surveys that are at hand in this thesis could perhaps be enhanced with insights from complexity-theory.

As for the benchmarking surveys both possess certain elements of complexity. The World Happiness Index perhaps to a lesser degree, for the number of variables and sub-categories in their operationalizations is more limited and specified. Scrutinizing nearly 150 countries is a relatively complicated study, nevertheless. And if the definitions on what to be researched would alter in this study or either the amount or quality of variables measured changed, for example by also adding elements of experiences on subjective happiness introducing, elements of complexity thinking would be welcome.

As for the InCiSE one area of improvement could be directly applying complexity theories behind the operationalizations and data procession proving out to be handy. Also slightly more carefully pondering the culturally sensitive insights of welfare-states and state governance philosophy could be operationalized more profoundly in possible further InCiSE-rounds. This is for their target of study and analysis with all its sub-variables being a far more delicate and complicated interdependent pool of interaction-dynamics than studies on world happiness. All in all, the key lies within understanding that the more international and multi-faced a study or survey is, the more complex it in general tends to become.

With much pondering about political analysis in this work; introducing complexity-thinking to politics could be a useful tool if not already applied. At least on the higher level of politics: national, continental, and international, things have started to become so tightly connected with strands reaching to even unpredicted directions understanding of at least elementary complexity theories seems useful to leading politicians. As a matter of fact, some of them have already done that. The former minister for sciences and culture, Antti Kurvinen mentioned in his jubilee speech at Vaasa University Society's 50th anniversary in December 2021 that politicians have to be prepared to face and find satisfactory answers to emerging wicked problems in society.

Indeed, politics may be the branch in which most wicked problems arise to be dealt with, some perhaps emerging from internal and external social structures while other from sheer increased information-intensity of the contemporary era. It is fruitful for politicians to bear in mind that benchmarks and various other rankings are here to stay. And some of them are very influential such as the country's credit ratings, Ease of Doing Business Index and various higher education comparisons. And they can affect both state economics and public opinion.

Moreover as there is motion a counter-motion tends to occur. Within politics and economics the counter-motion to neoliberalism and NPM could be right-wing state-corporatism with conservative populism. The more complex and hectic the world becomes the more difficult it is to perceive without academic training or/and a high IQ. And as long as Gaussian curves apply to population distribution there will be people hankering for a more simple old-fashioned world. It is of course an entirely different question was the world or it's social structures actually anyhow more simple in the past, say a century ago. This picture might be merely a romantic mindset generated by lots of historical information-bias. And to pop a bubble of romantic historic imaginary: probably the invisible hand Smith praised wasn't the hand of God. Probably it is merely a consequence of complex down-up-driven market forces. Or more precisely individual

economic decisions made by enterprises, public bodies and individual consumers that are interconnected and thereby forming the socioeconomic fabric of markets.

Although there are wicked people in politics as in other areas of life and wicked political gaming occasionally waged, we must not fall to the semantic illusion that politics itself would be a wicked game. On the contrary, politics is one of the primary tools in attempting to solve these new-era problems. And besides politics is based on relatively predictable structures of political parties, coalitions, and negotiation traditions. Speed of modern politics has increased with the development in communication technologies but this itself doesn't make politics more "wicked" , merely more hectic. Within the two-fold of politics and administration it is fruitful also to bear in mind for at least the top- ranking civil servants most closely connected with politics that complexity and wicked problems are going to affect their job descriptions as well.

As for the assistant research questions. It seems that results in benchmarking surveys are a result of many complex procedures. Firstly the operationalization and quantification of the study subject is extremely important for many of the matters studied are more qualitative social phenomena. And preference over qualitative matters itself is almost a wicked problem. There can be arguments on whether the subject studied has been correctly transferred into numbers. In order to receive as accurate numberizations and arithmetical functions as possible complexity-thinking can yet again help to deeply understand and quantify the complicated causal rations on matters at hand.

If operationalization has been made accurately or accurately enough the next step in results forming actually is the gathering of data. Methods to this can be a myriad, for example big data mining, surveys, collection of the data from corporate or public body archives or even purchasing it from data-mining companies. Thereafter the quality of data has to be evaluated. There are risks such as getting too much data or too little data or data from a wrong matter as well as corrupt data. Moreover cultural

perceptions and linguistic translations biases can affect the contents of data. If it is of sufficient high quality it can be computed. After the massive amount of data has been summarized it is time to draw graphs condensing the results. After that analyzing the results and publishing them depended on their nature in a scientific or economical journal or in mass media. So the results of a benchmark-survey depend on how the study subject has been operationalized, what kind of data is collected, how it is processed and interpreted.

5 CASE STUDIES

5.1. The World Happiness Report

The World Happiness Report (later often referred to as WHR) celebrates its tenth anniversary this year at the time of writing this section in 2022. Within this survey there are 146 countries all around the world that have provided a sufficient amount of data on the requested variables. One regional categorization can be described as following: “Western Europe, Central and Eastern Europe, Commonwealth of Independent States, Southeast Asia, South Asia, East Asia, Latin America, and Caribbean, North America and ANZ, Middle East and North Africa, and Sub-Saharan Africa”. These vast amounts of data have been primarily collected by the independent data-production organization Gallup World Poll (GWP). There seems to have been countries over the past in the materials that have not submitted sufficient amounts of data, or the participation has been so occasional that they could not have been compared with reasonable effort. As a matter of fact, the survey has had the principle of gathering missing data on individual factors from previous years, but no more than three years backwards. So, if there have been longer scarcities or larger amounts of missing data an individual country has been dropped out. (Helliwell, etc., 2022b, p.3–5&11–16.) However, when applicable without causing disproportionate distortions in the raw data in some cases missing information has been calculated with closest

variables available (Helliwell, etc, 2022b, p.4). It seems that the data-processing principles in this survey are clear, transparent, and relatively strict.

Happiness can be a quite difficult concept to define and study due to its ambiguous, even complex nature. Happiness in the WHR has been approached via dividing it into eight sub-factors. Four of them being more or less “hard” or quantitative variables while the remaining four are “soft” or mostly qualitative variables. First one being unsurprisingly money, in scientific terms ranked in per-capita gross-domestic-production (GDP) in US dollars ($). This is used to measure purchasing power parity (PPP) in the country. Also, GINI-index data obtained from the World Bank has been used to calculate the financial cornerstones of the survey. Second factor is healthy life expectancy (HLE). This crudely means the amount of years a newly born baby can on average expect to live healthy without any major diseases. So, this does not mean life expectancy which can be from some to several years higher depending on the country. Data for this variable has been collected from an independent source, the World Health Organization (WHO) with some minor chronological restrictions. Results of these surveys are displayed both on pages 16-18 and in the original results report in graph form. (Helliwell, etc, 2022b, p. 1-4, 32-55.)

Thirdly, corruption perception is measured as the national average of the survey responses in whether corruption is widespread in the governmental and business structures in the named country. With some methodological side notes including that the answer is an average of 0–1 response to those two questions on corruption. If there hasn’t been data on public body corruption available, the mean value has been calculated solely on business corruption. The data comes again from GWP. This value is closely related to experienced institutional trust. The Fourth quantitative factor is generosity. In simple form GWP has asked the recipient if they have donated money to charity in the past month. This is elaborated when sums of donation and their proportion on GDP per capita comes on display. (Helliwell, etc, 2022b, p. 1–4, 32–55.)

On the so called ‘soft’ side of happiness factors there is *social support*. This fifth variable means a person has someone to rely on, especially in times of trouble. This is again put in a binary yes–no form within the answer-options. Another indicator of happiness, the sixth factor is freedom to make life choices. GWP has studied this via asking whether the recipient is satisfied or dissatisfied with the freedom to choose what one does within life. Then on the emotional side are the counterparts positive and negative affection – the seventh and eight variables. In brief GWP studies map positive emotions/affect as laughter, enjoyment and doing interesting things. Whereas in GWP mappings negative emotions [affect] are worry, sadness and anger. The outcomes are yet again displayed as a graph at the end of the Statistical Appendix. (Helliwell, etc, 2022b, p. 1-4.) Of course, ultimately the categorization to hard and soft measurements is no objective, it is merely a generalization made by the writer to classify these many factors. In addition, a practical ‘humanization’ in framing the research question is the usage of the so-called Cantril ladder. In it the recipients visualized in their mind a ladder with steps from 0 equaling the worst possible life to 10 equaling the best possible life. Then recipients placed themselves on the ladder depending on how they imagined their current course of life to be. (Helliwell, & al, 2022a, p. 15.)

There are some other methodologically interesting notions within the World Happiness Report. A rising trend of artificial intelligence (AI) has been used to track down emotions in social media before and during the COVID-19 pandemic. The research team made a pre-hypothesis that there would be a positive correlation between self-reported and via social media expressed emotions. The primary social media platform studied was Twitter. Since men are more active on Twitter, balancing had to be made while analyzing the results. Methodologically it has to be noted that emotions in this study are only expressed in text-form, at maximum as emojis. On the other hand, it would become extremely difficult and expensive to insert advanced medical sensors to observe the biological emotion-responses to a group of recipients this large. There were two data pools collected, one from ‘historical’ existing data between June

2019 and October 2020. And then predictive data was collected during November 2020. When data collected was analyzed by using AI and machine learning, a correlation was indeed found, and it appears that especially negative emotions seem to be strongly expressed in social media. (Helliwell, etc., & al, 2022a, p. 77-84.) This can be an outcome of anxiety levels during the second worldwide corona-wave have been particularly stressed and filled with anxiety. Although in empirical experience top moments of life seem to be more often expressed via SoMe, it is noteworthy that also negative emotions can be expressed there as well which might make SoMe a bit more credible augmentation of peoples non-virtual, real life. Perhaps frustration and other negative emotions expressed in social media are also to some extent a way of seeking social support for the hardships.

Nonetheless as digitalization roars onwards, the UN has defined access to the Internet as human right and smartphones as common every-person's commodity in at least the highly developed industrialized countries and many developing countries, significance of AI and machine learning can and should not be omitted as handy tools for researchers. However, realities in such methodology need to be recognized. This big data can more easily be if not corrupt, at least prone to noise due to bots and fake accounts actions. Although samplings can become wider than in most cases of traditional information collecting it can also become somehow balanced, which has to be considered while analyzing or either extrapolating or interpolating the results. Some technical bugs in AI or machine learning might yet occur, although the softwares and technical systems for these have developed enormously over the past decade. All these potential pitfalls seem to have been duly taken into critical pondering within the study of social media emotion tracking survey in the World Happiness report (Helliwell, etc, 2022a, p. 95-98). However as in many bits the world has become and continues to be more technologically advanced and complicated it may very well be necessary to introduce computer-based or -assisted methods in order to handle the increasingly massive amounts of data in the postmodern world.

Oftentimes a holistic approach to the matter at hand is helpful. And so is the case within the Happiness Index. In order to deepen the comprehension of happiness, its biological basis has been covered. Nature vs. nurture, this eternal source of contention amongst scientists is at hand once again. The most prolific way to deal with it has been to approach the genetic effect of happiness via studies of twins. Massive data banks over twin-studies all around the world have been scrutinized in order to find a sufficient information-basis. Especially those twin studies, where the twins have raised or/and lived separately have been insightful, since there the genetical similarities come more forward. When these vast amounts of data have been analyzed the researchers have concluded that "30-40% of differences in happiness between people is accounted for by genetic differences between people". (Helliwell, etc, 2022a, p. 107–113.)

Consequently 60-70% of the variance in happiness-spectrum can be explained by differences inheriting from experiences and exposures to the living environment(s). Generally, it appears that the physiological basis of happiness is likely related to high levels of the neurotransmitter serotonin (Helliwell, etc, 2022a, p. 10). On a hormonal level, the amount of a hormone called cortisol or as in common language "the stress-hormone" affects the biology of happiness. The higher and longer periods of time levels of cortisol are increased the worse an individual human being tends to feel. Cortisol also affects the activity of the immune system and the more active one's immune system is the lower well-being in long term most commonly appear to be. (Helliwell, etc, 2022a, 107–113.)

Also, an important finding is that all genetic effects are not automatically manifested but rather can occur and activate in certain circumstances and life events caused by external events. So, although some people are fortunate enough to have been born with characteristics to be able to feel more easily happy, the interplay between inborn genetic pool and the life-events that shape people is yet again far more complex. (Helliwell, etc, 2022a, p. 107-113.) As in the current conception of modern gene-research suggest there are approximately 20 000 genes in a human being (Juan,

etc, 2018). So, the combination of two people's gene pools uniting during fertilization can cause an immense amount of combinations which can contribute to a surprisingly complex genetic set affecting just the biological basis of happiness.

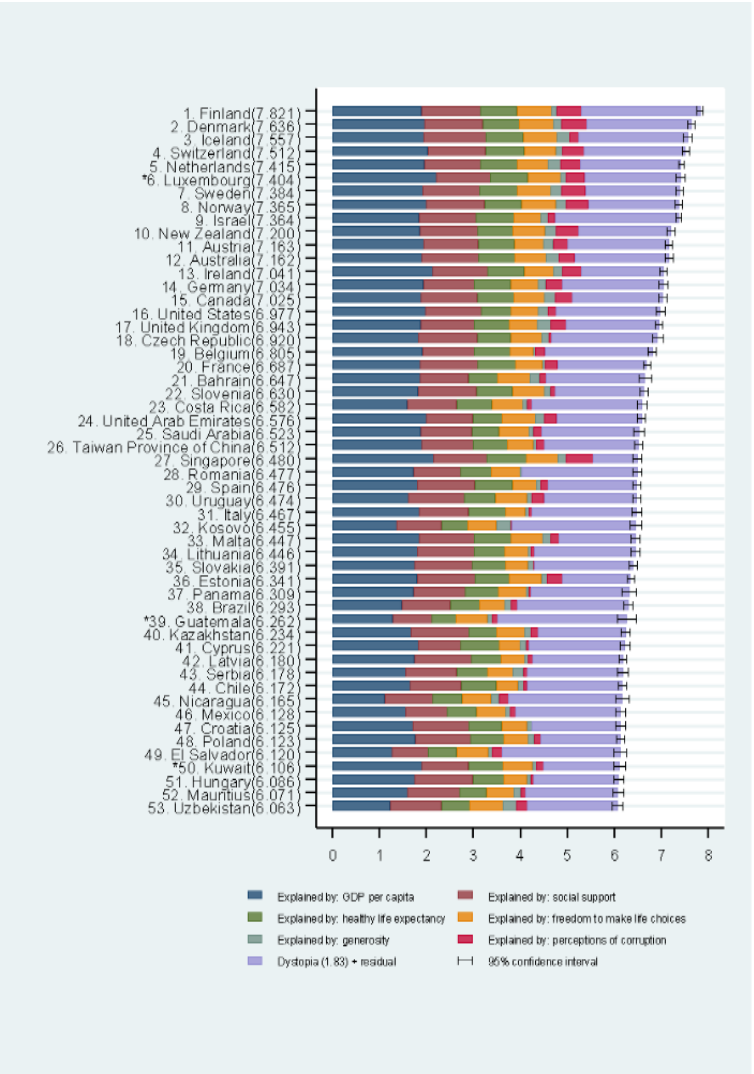
An interesting insight with this version of Happiness Index report is Chapter 6 in which results of the First Global Survey of Balance and Harmony are revealed. Although perhaps in popular literature in the western world these conceptions mainly associated with the Eastern-Asian and Indian philosophy are also of scholarly interest and hence research over the subject has also experienced some sort of “boom”. When attempting to understand happiness on a very profound level, also the somewhat deep experiences of life in harmony with the surroundings might deem to be considered. Like often in human sciences there is no one absolute and exact definition for both harmony and balance. They are somewhat similar concepts with slight cosmetic differences. However, they both consist of sub-phenomena that include a varied spectrum containing work-life-patterns, society, politics, motivation, relationships, emotions, attention, character, nature, exercise, diet, and sleep. (Helliwell, etc, 2022a, p.129.)

So far there appears to have been a research gap in insight into how balance or harmony might be linked to happiness. Nevertheless, with new studies this gap is gradually being filled. Probably this is not going to be the last time these matters are discussed in an international happiness report. More thoroughly seen, this area of research consists of such mindsets that include balance, peace, self-other-prioritization, calmness preference and calmness itself. Surprisingly though it may seem, after the GWP made extensive inquiries and data collection over the subject in analysis-phase, researchers noticed that experiences of harmony and balance are not the privilege of only traditionally more collective-oriented eastern cultures. Although schools of thought like Buddhism, Taoism, Confucianism and Shintoism tend to be more balance-seeking than the monotheistic western-associated book-religions, balance & harmony questions matter to all people around the world. Also, these

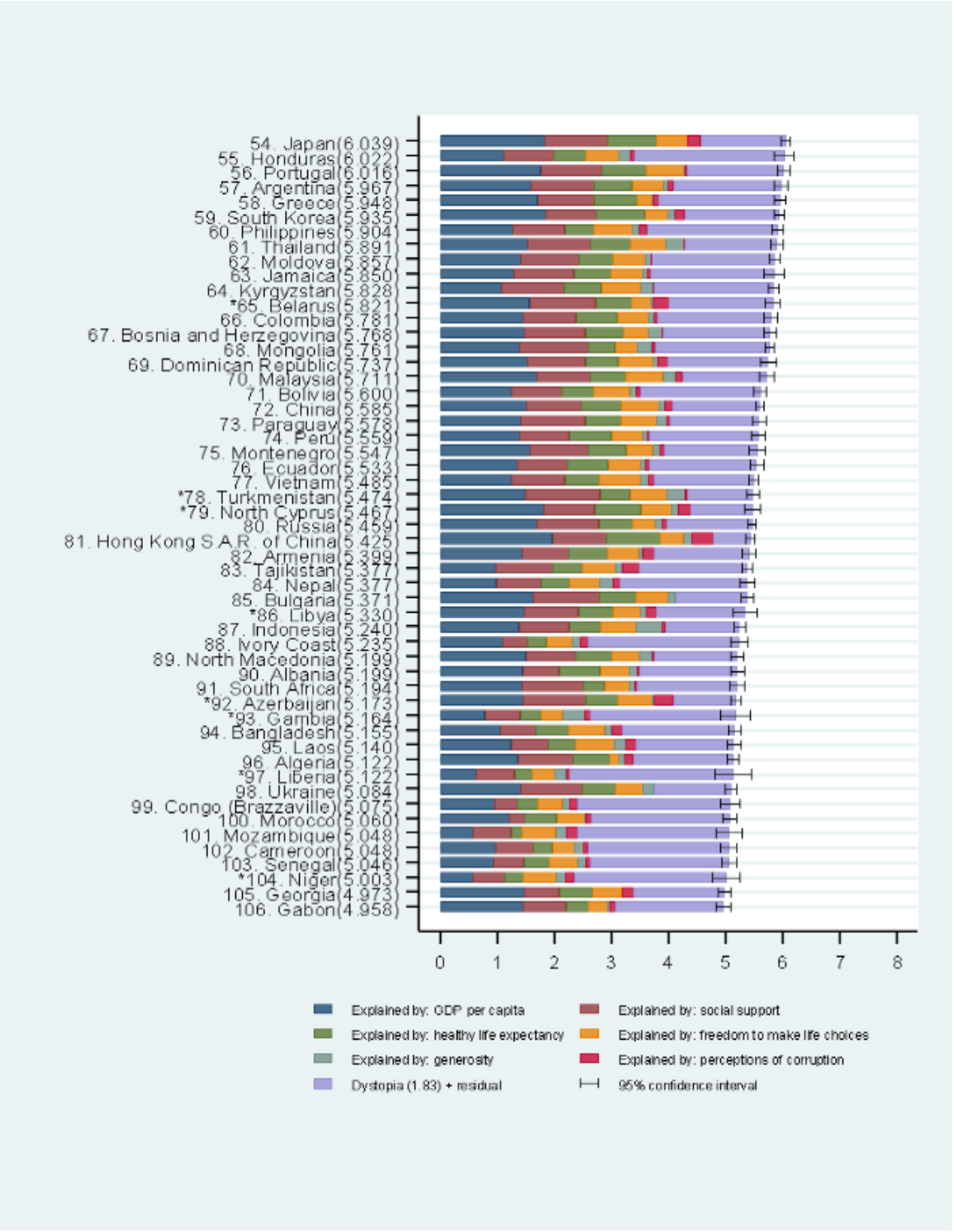
balanced-life issues lie in the center of well-being. Moreover, studies showed that western countries on average rank higher in balance-harmony ratings, with the top rankings somewhat surprisingly located in the Nordic Countries. (Helliwell, etc, 2022a, p.130–145.)

5.1.1. Results of the Happiness Index

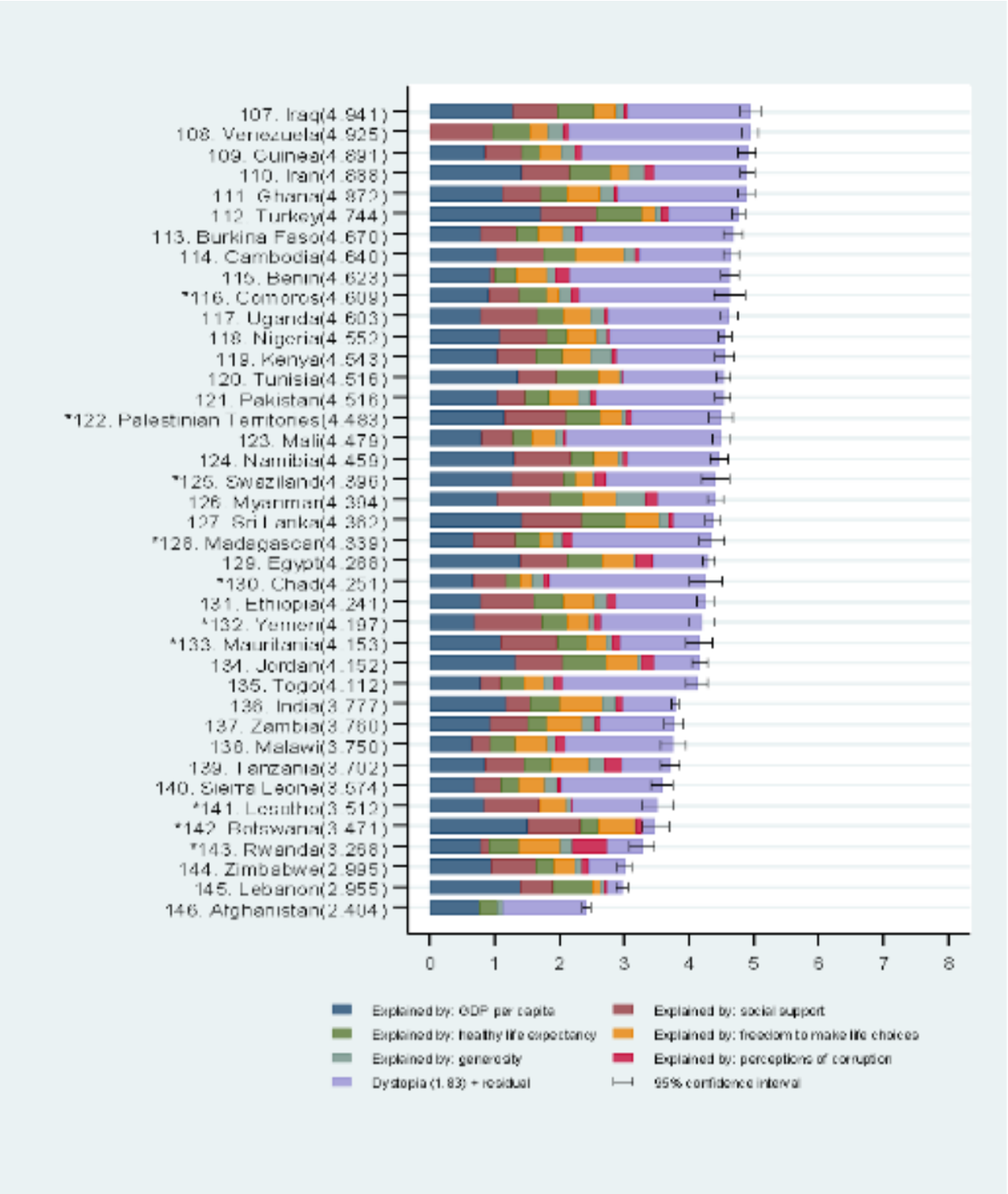
With all the great amounts of data collected over the eight variables put together the happiness index for year 2022 looks as following:



WHI 2022, first tercile. Picture from World Happiness Report, Appendix 2.



The second tertile of happiness index. Picture from World Happiness Report 2022, Appendix 2.



World Happiness Index, third tercile. Picture from World Happiness Report, Appendix 2.

As for a compact analysis for the results and trends in the index. The TOP-10 countries in the Happiness Index 2022 are somewhat expectedly highly developed, post-industrialized western service-economies, also including Israel and New Zealand. All five Nordic welfare states occur in the TOP-10, while three even in TOP-5. Furthermore, the TOP-20 countries include only rich, highly developed western countries, most of which are in either Europe or North America. Diversity takes place in TOP-30 where a wealth of developed Arab oil-producer countries and technologically advanced nations of the Far East come in. Countries of Southern Europe and the East Block are in very diverse settings. Virtually none of the heavily populated (ca. 200 million+) countries are on top rankings as China holding the place 72, Russia 80, India 136, virtually the only exception being the United States of America on place 16.

Of the representatives of Anglo-American public sector doctrine in TOP-10 the only one is New Zealand. In TOP-20 though a vast majority of these countries: UK, USA, Australia, and Canada. Japan is only in 54th place even though it is prosperous and technologically advanced. In the middle-places 50–100 there is a varied array of countries from all over the world, mainly South America, Far-East, Oceania, and Africa. (Obviously this cartographization and the Air Directions are heavily depending on the ones established during the rule of the British Empire.) On the lowest-ranking tertile there is a setting of countries mostly from Africa, Middle East, and Muslim-countries in the Far East. Last one being war-ruined Afghanistan. As a matter of fact, the six lowest-ranking countries: Afghanistan, Lebanon, Zimbabwe, Rwanda, Botswana, Lesotho, and Sierra Leone have in the past 20 years been either at war, civil war, internal turmoil or experienced a brutal genocide.

More of the trends in the survey are amongst other following. The gap between top and bottom countries is vast which perhaps implicates inequal living standards globally. It is neither certain that a country stays in its previous post. That is for better and for worse, for example France has climbed to its so-far peak of 20th when again the haven for North American happiness, Canada, has dropped from place 5 to 15. In Latin

America it appears that certain specific forms of social and family life increase the rating by app. 0,5 compared to a purely arithmetical model. On the contrary, countries in East Asia appear to compare slightly lower than they theoretically should. Accuracy of this comparison also describes a statistical certainty factor called the 95% confidence interval which has duly been taken into consideration while making this arrangement. In the long term, ten years of life evaluation has risen by more than a full point (on a ten-point scale) in 15 countries whereas it has sunk with the same amount or even more in eight countries. Positive emotion has stayed relatively stable but mostly due to COVID-19 pandemic, negative affections have risen. Perhaps also due to the global pandemics measures of prosocial behavior including helping strangers, donating, and volunteering seem to have increased. (Helliwell, etc, 2022a p.24 & 26-33.)

5.2. The InCiSE benchmarking-method

The International Civil Service Effectiveness, InCiSE method was originally ordered by the European Commission. It is executed by the Blavatnik School of Government at the University of Oxford. It is a sort of “umbrella-study” that summarizes existing data on civil service for more thorough studying. So far there have been two rounds of the survey, pilot in 2017 and enhanced version in 2019. There are three essential functions of governance that have been highlighted in the study. First of them is a well-organized central executive power possessing the abilities to manage fiscal resources, coordinate policies and assess them as well as manage risks. Second important aspect is mission support which contains financial management, Human Resources Management, ICT, and procurement. Third aspect is the outcome of this, so called de facto operative administrative activities, that is military, law enforcement, healthcare, social services, and education on all of its levels. Six most fundamental attributes highlighted in this study: integrity, openness, innovation capabilities, staff engagement, and inclusiveness. (Van Dooren, 2018. p.16-17.)

The background idea behind the InCiSE method was a need to have an international benchmarking method to define and compare effectiveness, transparency, and well-functionality of different countries' civil services. But the problem that arose was how to define effectiveness satisfactorily. There certainly is a lot of existing data related to indicators of civil service effectiveness. However, the aim in the InCiSE method is to gather and analyze this data more thoroughly. (Blavatnik School of Government, 2019b. p.9.)

As we have seen, benchmarking in general can be used to both business organizations and those of the public sector. However, in the latter case a need emerges to define the concept of civil service. There are three classical approaches explaining civil service. First is functional. In broad view civil service consists of the agencies responsible for service delivery. A narrower view of this approach emphasizes on the hierarchical structure of the civil service. That is the central national agencies or ministries regulating regional offices and authority institutions hierarchically below the central agencies. (Blavatnik School of Government, 2019b. p.8.)

The second approach relates to national accounts. Semantically account or accountability can mean both the civil service responsibilities and concrete bank accounts. But this interpretation indeed emphasizes more on the financial side of accounts. This approach is based on the idea that a state owns its subordinate governmental agencies. Hence bringing the agencies responsibility to report on their payment transactions and other financial activities. There likely is also a (national) system of accounting that can be used for this purpose. Third approach to civil service is via employment regimes. Recruitment and activities of civil servants tend to be regulated by specific and elaborate civil service laws. Therefore, civil service organizations can be deemed as those in which a majority of the employment and employees operate under various laws related to civil servants. Especially under the times of NPM this definition can be problematic. Increasingly more people working in

the public sector are no longer in service but rather in normal employment relationships. (Blavatnik School of Government, 2019b. p.8.)

How does InSiCE approach the definition of civil service? It presents a fourth and sort of alternative approach “by outlining and measuring performance on the core functions of civil services”. Some parts of the civil service can be identified rather universally. Therefore, InCiSE focuses on policy functions and public management conducted by the central government. InCiSE also takes into consideration actions that produce services or have direct effect on citizens. Thus, InCiSE focuses more on the civil service rather than merely the public sector. InCiSE has been developed for supranational purposes. Therefore, it focuses on the highest level of government in a state or a country. Typically, that is either the level of federal or central government. Therefore, it is fairly difficult if not even inappropriate to use this methodology to a single state agency. (Blavatnik School of Government, 2019b. p 8.) Hence, I will later on examine the case of Finland's placement in the InSiCE study.

Overall, the InCise potentials usage can be seen as a performance improvement tool which allows senior decision makers (both political and government officials) to see which countries perform the best in certain areas for the not-so-well succeeding governments to improve by benchmarking the top-ranked ones. Secondly the survey can also be used as an accountability tool which enables the citizens, politicians, and government officials to see and compare how their civil service is operating. However, as the main pages of the UK government related to the InCiSE state: “InCiSE is not claiming at this stage to be a robust, comparative measure of civil service performance. It is therefore important to view this inaugural Index as a pilot only, to be refined and improved over time. “(UK Institute of Government, 2022.)

InCiSE appears to be a very sophisticated instrument. Its focus is indeed divided into four main categories: 1) core executive functions (CEF) 2) mission support facilities (MSF) and service delivery functions (SDF) and 4) attributes (A). More detailed, the

Core Executive Functions consist of policy making, fiscal & financial management, regulation + crisis & risk management. Mission support facilities contain procurement, HR management, IT for official and internal finance. Service delivery functions behold tax administration, digital services, and social security administration. Attributes being the widest subcategory withholding integrity, capabilities, staff engagement, openness, inclusiveness, and innovation. (Blavatnik School of Government, 2019a, p. 14.) A picture perhaps illustrates this better:

Picture from InCiSE 2019 Results report, p. 14.

The quality of data perhaps has to be discussed more thoroughly. Most of the data for the 2019 survey has been collected at the end of 2018. Some data types can be collected by only every second year or even more seldom, while part of the data was produced specifically for this survey. Some pieces of information can contribute to more than one category; however, the study has attempted to avoid overlap. In contrast to the 2017 edition there has been advancement in which core functions have been measured, that is 11 parameters instead of eight + attributes. The authors also acknowledge that this is merely a piloting survey and occasional subjective judgements had to be made while filling in some data-gaps and making aggregations over the results. Although the number of indicators is 12, metrics used to assemble these

indicators climb even to a high number of 116. As well the number of countries has risen for the 2019 study up to 38. A majority of the countries provided high-quality data for most of the indicators. However, for four factors: crisis & risk management, digital services, capabilities, and risk management certain countries had no data in the requested form. Moreover, the TOP-5 countries in data quality were in descending order: UK, Italy, Poland, Sweden, and Norway. Missing data has been attempted to cover by statistical simulation, a method named multiple imputation. (Blavatnik School of Management, 2019a, p. 15–17.)

Figure 2 Summary of data quality by country

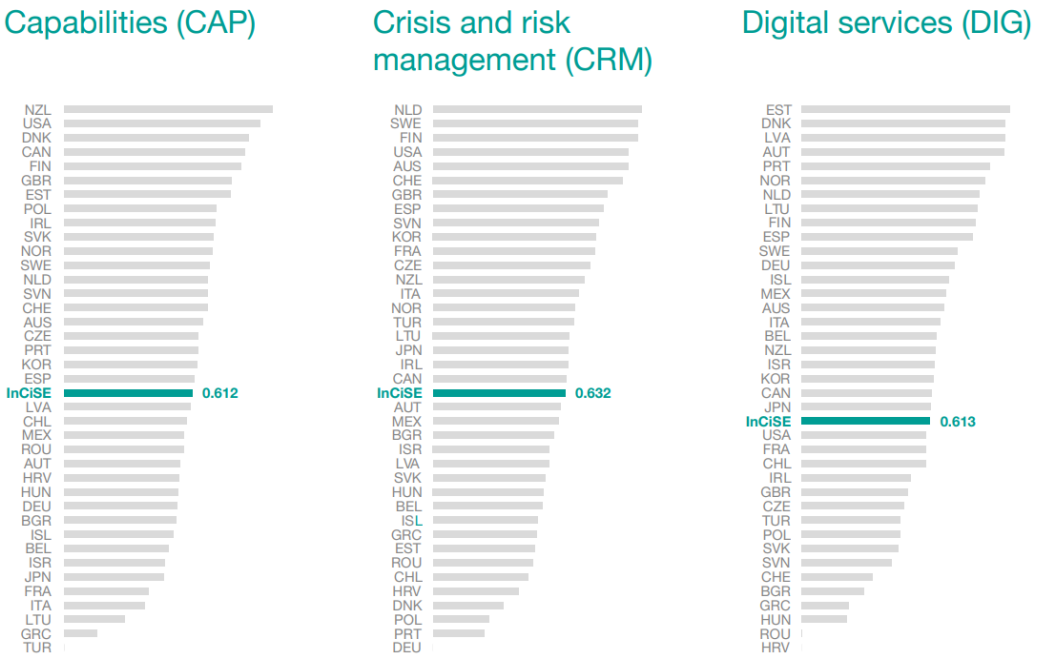
Country	Overall data quality		% of metrics available	Data quality of indicator											
	Score	Grade		CAP	CRM	DIG	FFM	HRM	INC	INT	OPN	POL	PRO	REG	TAX
GBR	0.757	A+	100%	●	●	●	●	●	●	●	●	●	●	●	●
ITA	0.755	A+	99%	◉	●	●	●	●	●	●	●	●	●	●	●
POL	0.755	A+	99%	●	●	●	●	●	●	◉	●	●	●	●	●
SWE	0.755	A+	99%	●	●	●	●	●	●	◉	●	●	●	●	●
NOR	0.752	A+	99%	●	●	●	●	●	●	◉	●	●	●	●	●
SVN	0.75	A	99%	●	●	●	●	●	●	●	◉	●	●	●	●
AUT	0.738	A	98%	●	●	●	◉	●	●	◉	●	●	●	●	●
FIN	0.736	A	97%	●	●	●	◉	●	●	○	●	●	●	●	●
ESP	0.733	A	97%	◉	●	●	◉	●	●	◉	◉	●	●	●	●
NLD	0.731	A	98%	●	●	●	◉	●	●	●	●	●	◉	●	●
FRA	0.718	A	97%	◉	●	●	●	●	●	●	●	●	○	●	●
PRT	0.716	A	85%	×	◉	●	●	●	●	◉	●	●	●	●	●
DNK	0.707	A	93%	●	◉	●	◉	●	●	○	◉	●	●	●	◉
DEU	0.701	A	96%	●	●	●	●	●	○	●	●	●	●	●	○
GRC	0.696	B	94%	●	●	●	◉	●	●	◉	●	●	◉	●	○
SVK	0.692	B	93%	●	◉	●	◉	●	●	◉	○	●	●	●	◉
HUN	0.671	B	81%	×	◉	●	●	●	○	●	◉	●	●	●	●
EST	0.669	B	90%	●	○	●	◉	●	○	●	◉	●	●	●	◉
CZE	0.659	B	91%	●	◉	●	●	●	○	◉	●	●	○	●	●
TUR	0.65	C	90%	●	●	●	◉	●	●	◉	●	●	○	●	○
MEX	0.648	C	73%	×	●	×	●	●	●	●	●	●	◉	●	◉
NZL	0.644	C	83%	●	●	×	●	◉	○	◉	●	●	◉	●	●
CHL	0.643	C	79%	●	○	×	●	●	●	●	●	●	◉	●	●
CAN	0.638	C	78%	●	○	×	●	●	●	◉	●	●	◉	●	●
KOR	0.636	C	78%	●	○	×	●	●	●	●	◉	●	◉	●	●
BEL	0.635	C	85%	●	×	●	◉	●	●	●	●	●	○	●	●
LVA	0.628	C	75%	×	○	●	○	●	●	◉	○	●	●	●	●
CHE	0.627	C	79%	×	●	●	◉	●	●	●	○	●	○	●	◉
AUS	0.618	C	71%	×	●	×	●	●	●	◉	●	●	×	●	●
LTU	0.615	C	82%	●	○	●	○	●	●	○	○	●	●	●	●
IRL	0.614	C	84%	●	○	●	◉	●	●	◉	○	●	○	●	●
JPN	0.597	D	75%	●	○	×	●	●	●	◉	●	●	○	●	○
USA	0.579	D	74%	●	○	×	●	●	●	●	●	●	×	◉	◉
ISR	0.578	D	72%	●	○	×	◉	●	●	○	○	●	◉	●	◉
ISL	0.563	D	68%	×	○	●	○	●	●	○	○	●	○	●	◉
ROU	0.529	D	66%	×	◉	●	○	○	○	○	◉	●	●	○	◉
BGR	0.511	D	66%	×	◉	●	○	○	○	○	◉	●	○	○	●
HRV	0.501	D	65%	×	◉	●	○	○	○	○	◉	●	○	○	◉

● High data quality ◉ Medium data quality ○ Low data quality × No data available

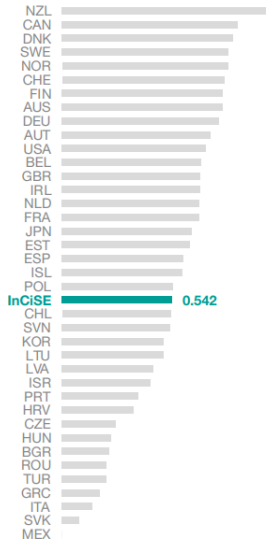
Summary of the collected data quality from InCiSE 2019 report, p. 16.

5.2.1. Analysis of the InCiSE 2019 results

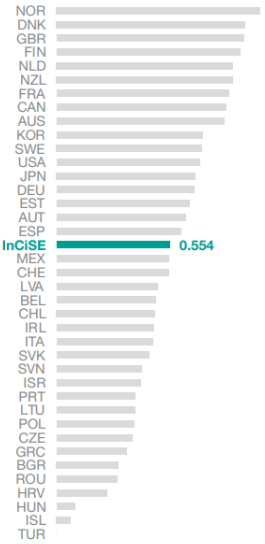
Figure 4 Country rankings across each of the 12 InCiSE indicators



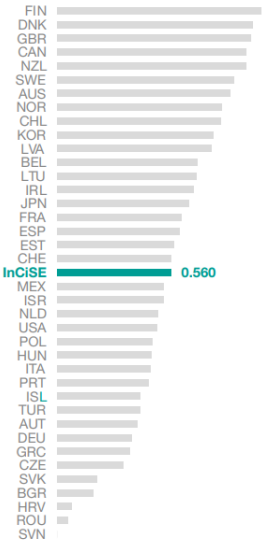
Integrity (INT)



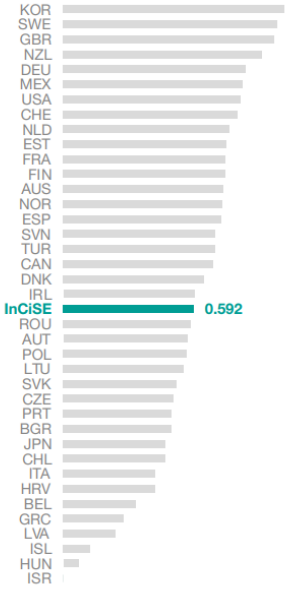
Openness (OPN)



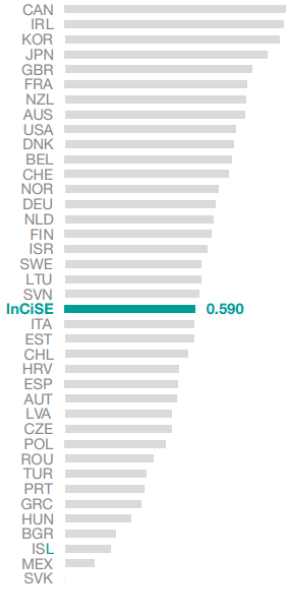
Policy making (POL)



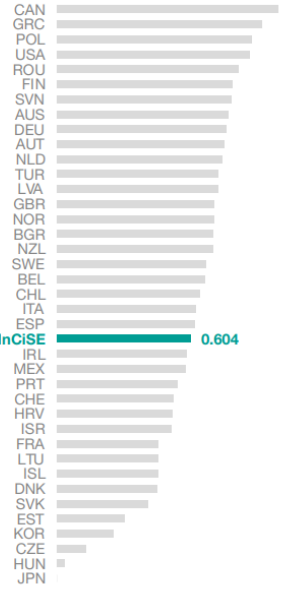
Fiscal and financial management (FFM)

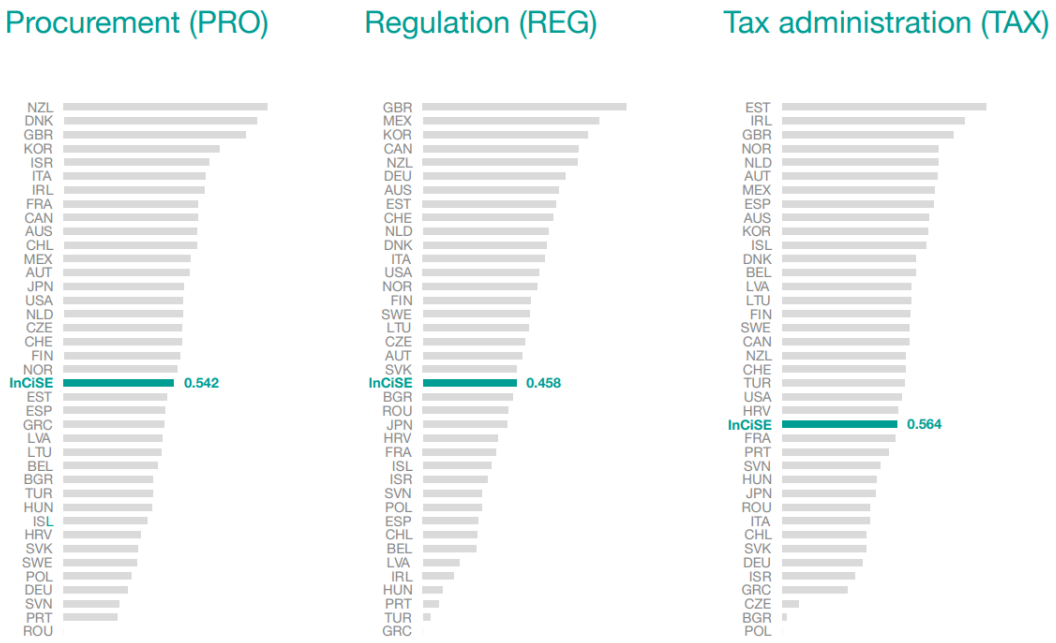


Human resources management (HRM)



Inclusiveness (INC)





All 38 studied countries' rankings per the 12 InCiSE-variables. Pictures from results InCise Results rapport p. 26-27

An analysis on the InCiSE results is appropriate at this point. Since the study is sophisticated and includes a great deal of indicators, no country is constantly in the TOP-5 of all categories even though there are some countries that stand well in many categories. On the other end a similar trend is shown, for certain countries seem to be constantly ranking relatively poorly within this methodology. Nevertheless there are 12 times 38 that is 456 variables altogether if all countries are included within. The 38 participated countries are (in order as they appear on Regulation-section): Great Britain, Mexico, Korea, Canada, New Zealand, Germany, Australia, Switzerland, Estonia, Netherlands, Denmark, Italy, United States of America, Norway, Finland, Sweden, Lithuania, Czech Republic, Austria, Slovakia, Bulgaria, Romania, Japan, Croatia, France, Iceland, Israel, Slovenia, Poland, Spain, Chile, Belgium, Latvia, Ireland, Hungary, Portugal, Turkey and Greece. Most of the countries in this survey are European, which might not be a surprise for the mandator of the survey was indeed the European Commission.

However a slightly larger nominator is needed to include USA, Canada, Australia, New Zealand, Israel and Japan with Korean Republic/ South-Korea. Countries of the western hemisphere would apply here. And Japan justifiably does include to this categorization, for it virtually benchmarked legislation, scientific thinking and industry from leading European nations during Meiji-restoration. During the Korean war and after that alliance between South Korea and the USA ensured western development in at least the southern part of the Korean peninsula. Most wealthy and western-style developing countries have also been included such as Mexico, Chile and Turkey, latter of which in a broad approach can be by geography included into Europe (at least partly) while two first being South-American countries, Mexico in Middle-America to be precise. Perhaps one original aim has been to broaden the survey towards well-developing countries as well.

It is noteworthy that the competition for top-places is stiff while the ranking scale is relative and normalized to the range of 0.00–1.00. Therefore, margins are often settled with decimals. Moreover, concerning the relativity: in this survey we have pretty much the TOP-40 ranking countries in the world, the order given here is reciprocal only to this method. The TOP-5 countries being 1) Great Britain 2) New Zealand 3) Canada 4) Finland 5) Australia. TOP-10 allowing other Nordic countries in place: 6) Denmark 7) Norway 8) Netherlands 9) Southern Korea 10) Sweden. Moreover, four commonwealth countries (UK, Canada, Australia, and New Zealand) have managed to sustain their place in TOP-5. Nordic countries have also done well by all of them being in TOP-10 except for Iceland. As this survey is focused on OECD-countries, generally countries from western and northern Europe have tended to rank higher than those of eastern or southern Europe. InCise averages are typically in numeric values between 0.5-0.6 with additional decimals, where only regulation is under that, 0.458. (Blavatnik School of Government, 2019a, p. 18–19.)

More thoroughly analyzed per factor. Within Capabilities (CAP) the TOP-5 there are Anglican-American and Nordic Countries; New Zealand as the first, USA as second, Denmark as third, Canada fourth and Finland fifth. In TOP ten same hemispheres continue but in addition two former members of the Eastern Bloc: Estonia and Poland also take place. Others being Great Britain, Ireland and Slovakia. In the middle is a relatively mixed setup of countries from Middle and Eastern Europe and South-American countries. In the bellow-5 are France, Italy, Lithuania, Greece and Turkey.

Within Crisis and Risk management (CRM) again Anglican-American and Nordic Countries elaborate the TOP-5. First however being the Netherlands. Sweden and Finland in second and third place. USA and Australia on fourth and fifth. In top also Switzerland is included as sixth, Great Britain as seventh, Spain as eight, Slovenia ninth and South Korea as tenth. In the mid-range there is again a mixed setup of countries from Middle- and Eastern Europe and South-America as in many factors onwards. Turkey however ranks relatively high in this category as 16th with an active and large army and collective will to protect the nation.

In the bellow-5 There are Croatia, Denmark, Poland, Portugal and Germany, with no pillar at all. It might be risky to make too definitive interpretations here but perhaps Germany's result is particularly perplexing as a solid and well-prepared society and with the history of prussian military doctrine. Even in modern Germany having Bundeswehr and Bundespolizei. As Demmke analyzed this most likely is a data-collection bias for Germany is a federal-state in which crisis and risk management is dealt by the constituent states rather than federal state. Therefore with a different setup of data collection or processment Germany would actually probably rank much higher than it now does.

In digital services (DIG) the TOP-5 is nearly opposite to that of crisis management. Estonia holds the first place, while Denmark being the second, Latvia third, Austria

fourth and Portugal fifth. Norway, Netherlands, Lithuania, Finland and Spain also in the top-ten. Baltic and Scandinavian countries rank very high in this category. Also Spain and Austria seem to be ICT-nations. On the contrary Romania and Croatia have not provided analyzable data on this category. Digitalization also seems to be slumbering in Southeastern-European countries such as Bulgaria, Greenland and Hungary.

Integrity (INT) seems to be an array of Anglo-Saxon and Nordic countries. New Zealand holds the first rank, Canada being the second, Denmark third, Sweden fourth and Norway fifth. Switzerland on the sixth, Finland seventh and Australia eight. Germany and Austria being ninth and tenth. The below-5 of integrity consist of Turkey, Greece, Italy, Slovenia and Mexico from which data was not available.

Openness (OPN) goes again mainly with a familiar composition of Nordic and Commonwealth Countries. However France being in 7th place breaks the pattern. First being Norway, second Denmark, third Great Britain, fourth Finland, fifth the Netherlands, sixth New Zealand, eight Canada, ninth Australia and tenth Korea. On the below-5-side then again there seems to be somewhat familiar states: Romania, Croacia, Hungary, Iceland being the exception and Turkey with no data.

Policy making (POL) is in the core essence of politics. To make decisions and outline strategic goals that is. The top-dogs here are again Nordic and Anglo-American countries. Finland being first, Denmark second, Great Britain third, Canada fourth, New Zealand fifth, Sweden sixth, Australia seventh, Norway eight, Chile ninth and Korea tenth. On the other end there are Eastern Bloc countries Slovakia, Bulgaria, Croatia and Romania on the tail end with no data available from Slovenia.

Fiscal and financial management (FFM) is the state apparatus' ability to collect taxes and take care of financial transactions. On TOP-5 places South-Korea as first, Sweden second, Great Britain third, New Zealand fourth and Germany fifth. In the top ten are also the USA, Chile, Netherlands, Estonia and France. In the "down under '' top-5 Israel

having no comparable data. Hungary and Iceland are located on the next steps. Latvia and Greece in 4th and 5th lowest places.

Human resources management (HRM) is on the employment side of states or governmental apparatus, an important indicator. That is how state employers are being taken care of, supported and managed. Most employment for the state is expert work and therefore employees are also an important resource to the state. Here we have a relatively typical line of well-performing anglo-american and Scandinavian states with Canada being first and Ireland second. Asian countries Korea and Japan are located in third and fourth place. Britain is found here in 5th place. France is in sixth place. After that New Zealand (ranking 7th) and Australia (ranking 8th). One more English-speaking country in the top-10: The USA (ranking 9th). Denmark in 10th place here is the only Scandinavian country. In the “below”-5 we encounter Eastern-Block state-names; Slovakia missing data on last place. Mexico and Iceland are next followed by Bulgaria and Hungary.

Inclusiveness (INC) means availability by physical, social and mental standards for example accessibility in public buildings with wheelchairs, availability of public services such as basic education and healthcare and legislation that takes into account UN declaration of human rights. Canada is in first place, Greece in second, Poland in third, USA in fourth and Roumania in 5th. Finland in sixth, Slovenia in seventh, Australia in 8th and in places nine and ten Germany and Austria. On the tail end of inclusiveness are Japan, Hungary, Czech, Korea and Estonia.

Procurement (PRO) means the acquisition of public purchases. Mainly this means material or immaterial goods and services of consultancy. In European countries this sphere of public activities is regulated by competition legislation. New Zealand was first, Denmark second, Great Britain third, Korea fourth, Israel fifth, Italy sixth, Ireland seventh, France eight, Canada ninth and Australia tenth. The least efficient end of

procurement according to this study are: Roumania, Portugal, Slovenia, Germany and Poland.

Regulation (REG) is in collaboration with Policy making in the core-essence of legislative public power functions. Giving orders and making rules on how to live in a society also mirrors the valuations within society. On a more juridical and technical side this category consists of not only writing the national laws but in EU countries adopting supranational directives to local legislation as well as to introduce EU statutes directly into national collection of regulation. Decrees clarifying laws but settling in legislative norm-hierarchy below laws also belong to regulation. A vast majority of top-ranking regulators are commonwealth countries with Great Britain first, Mexico in second place and Korea in third, Canada fourth, New Zealand fifth. Germany is placed 6th and Australia seventh. Estonia was 8th and Switzerland in 9th place whereas Netherlands in 10th.

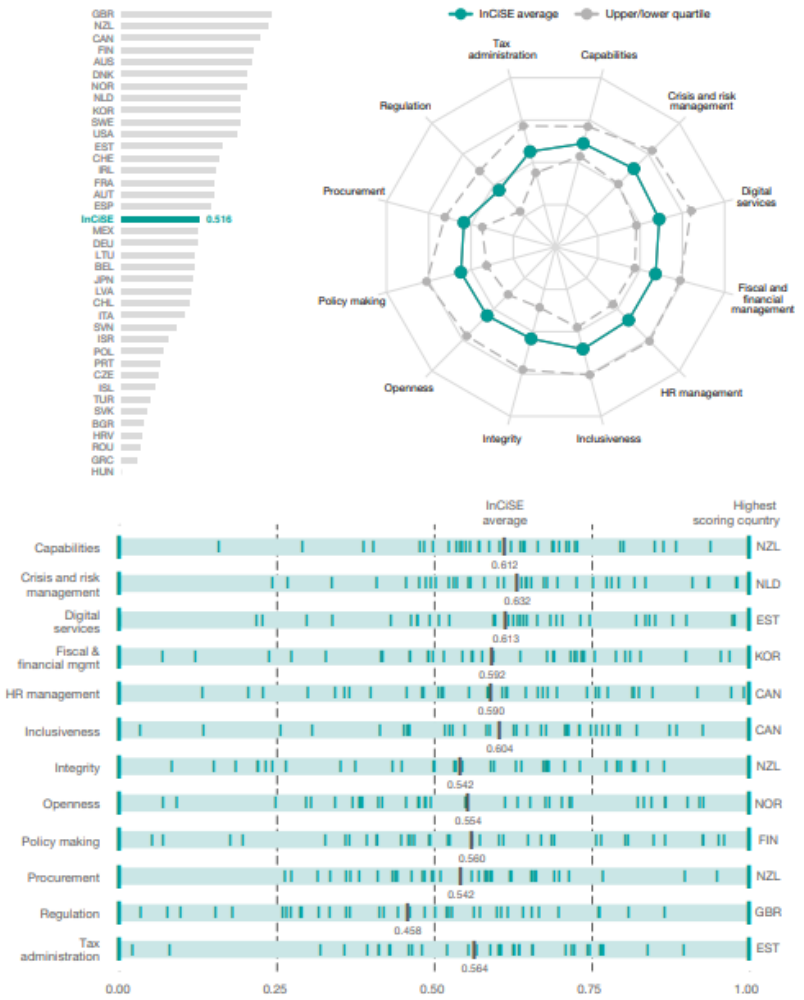
In the western hemisphere two major legal systems are the Common Law within Anglo-American countries and Roman-German written and punctual law. From the point of view of juridical philosophy, the Roman-German system is based on a written extensive collection of laws and regulations which are applied into individual justice-cases. Common law system bases more on legal praxis and case examples which are used and applied within individual similar or same-type of cases. The role of precedents and supreme courts giving those is highlighted. Representatives on both of these systems appear to be presented in the TOP-10. In the least regulative end we have Ireland, Hungary, Portugal, Turkey and Greece with no column at all.

Last but not least of the twelve public body core functions namely Tax Administration (TAX). A vast majority of public activities are financed by taxation. In the neoliberal era customer and usage fees have taken a bit stronger role in public formation of earnings. Within tax administration Estonia is the first and Ireland second. The United Kingdom is in third spot. Norway in fourth place also appears to have an effective tax

administration. The Netherlands is in fifth place. Austria is found in sixth place. Mexico is ranked in seventh place, whereas Spain is in eight place. Australia in ninth place and Korea in tenth. On the other end of the tax administration line we have a setup of Eastern Bloc countries Poland, Bulgaria, Czech whereas Greece and Israel.

For those interested a summary-picture of the *Overall results of the 2019 InCiSE Index is on display below. Picture from Blavatnik School of Government, InCiSE results report, p. 18*

Figure 3 Overall results of the 2019 InCiSE Index



5.2.1. Analysis on Finland's ranking in InCiSE 2019

Finland's overall ranking within InCiSE 2019. Picture from InCiSE results report, p. 40.

One area in which our example country, Finland actually did quite well in the InCiSE 2019 comparison was the overall category of "data quality by country". Finland ranked in the top-10 amongst 38 OECD countries that participated in the survey. To be more specific Finland was ranked eighth in data quality, grading a straight A with the score of 0.736. Finland had 97% of the requested metrics available. In the 12 categories Finland had high data quality in ten indicators: capabilities, crisis and risk management, digital services, human & resource management, inclusiveness, integrity, openness, policy making, procurement, regulation, and tax administration. One indicator was of medium data quality: fiscal and financial management. And only one indicator was of low data quality: integrity. Data of some quality was available in all of the measured indicator categories within the case of Finland. In comparison in the lowest-ranking quartile of

countries data was missing or of low quality in several or even over half of the categories. (Blavatnik school of Government, 2019a. p. 16.)

It is time to present some highlights of the indicator categories in which Finland ranked among top-5. In policy making Finland was the winner of the category, although there was a stiff competition. Thus, it seems that in this study Finland had the highest degree of strategic planning, policy coordination over government silos and the highest quality in policy monitoring during their implementation. (BSG, 2019a, p.20.) Finland was also successful in crisis and risk management in third place. In openness Finland was ranked fourth. Somewhat surprisingly the perfect pupil of meritocracy was only the fifth within capabilities. (BSG, 2019a, p. 21–28). In a more profound analysis Finland was credited for high scores of readiness to learn from job-related matters resulting in a strong ethos for development (BSG, 2019a, p. 40).

In overall results of the InCiSE 2019 index Finland was placed as the fourth. In top-3 there are Canada on the bronze, New Zealand on silver and Great Britain as the gold-medalist. That means Finland is the highest-ranking non-Anglo-American country in the study. (Blavatnik School of Government, 2019a. p. 18). And the result is somewhat perplexing: Is the conservative-bureaucratic system of regime that likely derives from the time of the British Empire still THE benchmark of the 2020's? According to the InCiSE study it seems to be. Then again we must bear in mind that origins of the ranking- and benchmarking-mindset come from the Anglican-American cultural hemisphere as Ringel&Werron kindly remind us. Therefore we cannot entirely avoid the idea that administrative cultures in the Commonwealth countries and USA have over the years co-evolved with benchmarking-tidiness in mind.

However, Finland – the world's happiest country seems to have done very well in the study, nevertheless. Perhaps the combination of market liberalism and Scandinavian welfare-state have generated a world-class civil service in Finland that is successful in policy formulation, transparency, and social rights. The Finnish system of education has

been praised world-wide as well. (Demmke, 2020b). Finnish civil servants are highly and well-trained for the formal qualification in virtually any public post requiring a diploma from a higher education institution. And for leading civil servants, teachers and senior ranking police and military officers a master's degree is required. Researchers, medical doctors, and university staff are required even to obtain a licentiate or doctoral degree. Also, the official elections in Finland are strongly based on meritocratic recruitment (Demmke, 2020b). So, the higher the public position the more skilled its holder has to be.

Also an interesting historical insight to the background structure of the modern Finnish system of governance is that it was constructed under the times of czars in the 19th century. And at that time the Russian regime was quite similar to that of the British Commonwealth's monarchy. So perhaps some silent knowledge or procedural practice originates from those times and therefore Finland has certain similarities with Anglican-American bureaucracy traditions hence increasing the overall rank in InCiSE.

5.3. Comparing the two benchmarking surveys

Now is the time to compare these two worldwide top-of-the line benchmarking surveys. They both appear to be fulfilling the criteria that Ringel and Werron have pointed out. They are comparisons of performance although there are differences in subjects. InCiSE is focused more on the OECD countries and to the functions of state apparatus whereas the World Happiness Index is worldwide and includes also developing and frontier countries and is more inclined towards softer matters such as living standards within countries. Secondly both surveys are at least in principle repeated studies. The Happiness Index is issued annually. The InCiSE has been issued periodically within the two years interval 2017 and 2019. However, during 2021 it wasn't published. So it is on the verge of being a regularly published survey. This year will tell it. Thirdly both of these surveys certainly contain quantification -- and lot's of it! The exact amounts of variables in these surveys is discussed in the next paragraph.

Fourthly they are both well visualized with accurate and informative multicolour graphs in them. In addition, the World Happiness Index contains pictures from western point of view of underrepresented groups such as ethnicities. Both surveys are available in electronic form via the Internet as we have accustomed to. The results side is on visual behalf presented in the traditional best on top and the rest in descending order.

The scope of both surveys is vast, World Happiness Index consists of 146 countries whereas InSiCE 38, mainly wealthy and highly industrialized countries of the western world as well as parts of countries that have copied or adapted elements of Western governance-tradition (for example Japan). Strictly speaking, the sampling for WHI is much larger than that of InCiSE, nearly four times. Then again, the sub-factors studied is in all manners larger within InCise, crudely 11 on 8 but if we also consider the partial parameter toll rises to 116 within InCiSE. Perhaps the latter being more subtle and on purpose focusing on the countries of the western hemisphere.

Both surveys deal with relatively abstract conceptions such as “happiness” or “civil service effectiveness”. It appears that we have no direct quantitative measures for gauging these. However, it is common in human sciences, including administrative sciences, to generate a definition-based paradigmatic universe for each specific area to study (except for certain areas of economic sciences that are heavily based on for example in statistics and econometrics). That however means that no paradigmatic system can be full-proof but rather based on the best knowledge at disposal on each timely moment of science-making. Even paradigms can change when better, more detailed knowledge accumulates. It might nevertheless be so that we are dealing with some of the most accurate benchmark-surveys for time being, even though that language and definition of concepts can cause biases to the data collected. Theoretically the risk for this is higher within the WHI as it has been answered with a greater number of languages than in InCiSE. However, it appears that the executor of the WHI, World Gallup Poll also has a greater number of resources at their disposal to translate and culturally adapt the survey.

In both surveys the main research question studied has been divided into subcategories. Happiness consists of national wealth as in the form of GDP, social support, healthy life expectancy, freedom of making life chooses generosity and perceptions of corruption. Here it must be noted that experiences of subjective and objective happiness are not the same. Subjective happiness is perhaps such a complex expression of an individual's inner world and pool of experiences that it might not be measurable with current research methods and equipment. However, the methodological approach to objective happiness in the WHI seems to be mainly focusing on life-standards from a myriad of viewpoints. Indeed, the material well being presented as the cornerstone of wellbeing; all countries in the study had at least a bit of this as an explanation in their happiness, even Afghanistan which lacked many other factors. It would thou be too inaccurate to explain happiness merely based on the material and financial wealth, otherwise Switzerland, United States and Norway would be holding the TOP-3 ranks.

Health is likely to be another fundamental of happiness. Abolition of sickness as the WHO defines health or at least having proper medication and treatments at an affordable price is one sphere of life defining ability to act and enjoy life (World Health Organization, 2022). Perhaps healthy life expectancy may even be a narrower concept since it appears not to consider last years of weaknesses and illness most common in even human living. There are many ways of arranging health services for example in Nordic Countries public healthcare is the main formula strengthened with health-enterprises whereas in the U.S. private health and doctor-companies in which treatment is paid either by cash or via health insurance is the major way. In the Middle- and Southern European countries employers, churches or various civic-society organizations mainly arrange the health services. (Esping-Anderssen, 1989.) Nevertheless, experience of fairness is as well a major part of people's experiences of being treated fairly and equally. Lack of it in either public bodies or business-life can

certainly contribute to dismay and unhappiness. But perhaps these matters are more thoroughly discussed within InCiSE.

WHI has studied many factors related to social quality of living. Social support, especially when experienced by the hardships of life and even when not, seems to be a crucial element for long-term happiness. Forms of this social support can be expressed with variance. Traditionally this means relative(s) taking care or assisting one during hardships as often is in the Catholic and Muslim countries. Also, in the urbanized west also help from friends can be of this quality. Not to forget organizational aspects of social support provided with varying degrees by the public bodies, municipalities, or 3rd sector organizations such as the Red Cross (/Crescent) or the Salvation Army. Generosity associated with charitable donation is relatively closely connected to general prosocial behavior.

Third part of the social living standards is freedom to make life choices. This perhaps two-edged sword has various aspects in a way having strict social structures and traditions can help to focus on the made choices at the crossroads-points of life. On the other hand, a too strict and conservative social atmosphere can be both limiting the individual's possibilities to make life-choices directing life to a wanted direction even to the extent of severe anxiety. Also, too little choice might limit the possibilities to make necessary choices as response to our today's global dilemmas, such as overpopulation or climate crises. In summary, in the World Happiness Index operationalization of the matter studied has been made successfully and with great care.

As for the InSiCE factors measured might be even more difficult to define into examinable form. However, a significant amount of effort has been put into formulating the model with its indicators. Even in the era of scandal-driven political theater, barbed satire and twitter-rage roaming around politics and politicians, the ultimate purpose of politics is to legislate based on the common declarations of intent and formulate policies based on these laws. In addition, supreme public budgetary

authority and overseeing the crises-organizations that maintain stability in society: army, police, and rescue services belong the entitlements of leading politicians. Oftentimes public power also acts as a certain back board of public healthcare, social services, and some sort of civilization services like kindergartens and elementary schools, however variance within these sets of tasks is already greater. Nevertheless, the four core executive functions identified within InCiSE seem to orthodoxly follow this classical constitutional-Weberian mindset over state tasks consisting indeed of policy making, regulation, fiscal and financial management added with crises and risk-management.

Summarizing the InSiCE, in the results report 2019 it has been highlighted that there are two principal categories in the study: the ‘what’ and the ‘how’-questions. The ‘what’ categories are functions: the core matters that civil services provide in every country. This category also has three subcategories which are firstly core executive functions. This category contains policy making, fiscal management, crisis & risk management, and regulation. Second category contains service delivery functions. In these areas civil service has a more direct interaction with citizens. This category includes tax administration, digital public services, and social security administration. Third category consists of functions that enable the “frontline” civil service to take care of its tasks, also a sort of “mission support”. This category includes internal finance, Human Resources Management, IT services for officials and procurement. And then the ‘how’-questions. They are behaviors and characteristics all across the civil service that are essential drivers in the way that core functions are being delivered. These are mostly in a way more abstract conceptions as well for it includes integrity, openness, capabilities, and inclusiveness. Also, staff engagement and innovation are considered within these attributes. (Blavatnik School of Government, 2019a. p. 13-14).

Concluded from this we could say that there is an enormous amount of data that has been required to collect and analyze in order to successfully execute the InCiSE study. For the time being internal finance, social security administration, staff engagement, IT

for officials and innovation are not measured in the InCiSE index (BSG, 2019a, p. 14). Presumably this is due to keeping the amount of data and variables at a feasible level. Furthermore, it has to be pinpointed that this is just the second implementation of the InCiSE survey, so it still is under process of development. So perhaps in the future surveys the number of variables will be increased. However, the principle with data is to use the latest and most accurate data given at the time of conducting the survey. However not always has this been possible. Some part of the data has been collected annually, while other parts of it biennially or with even longer intervals. In addition, the aim has been to collect data from various sources so that the material would not be too reliant on only one source. (Blavatnik School of Management, 2019a, p. 15).

Methodologically speaking there has been fewer discrepancies in the World Happiness Index. However, the subject of studying within InCiSE is also more abstract, specific, and difficult to define due to the vast amount of subtle and complex relations of interconnectedness compared to Happiness Index. Gallup World Poll seems so far to have succeeded slightly better in cultural adoptions and translation matters, despite a greater number of countries involved. However, we must remember that the Happiness Index has been well resourced and has been going on for a full decade – every year. Whereas InCiSE has so far been only carried out for twice and every second year as even its developers have claimed that it is still in test-phase honing its dynamics. So, fully halting the whole survey due to some potential methodological points of development seems somewhat disproportionate procedure. It seems that responding to this constructive critique and enhancing the InCiSE-study would be a more suitable course of action (Demmke, 2020a&2020b). Perhaps also adding some complexity thinking to the operationalizations and qualifications as well… Based on these to develop the survey into a conceptually more consistent manner, provided that the high-ranking civil servants still have the time and interest in participating in this survey. As the World Happiness Index shows, interest towards benchmarking hasn’t declined when times around us cool down a bit!

6 DISCUSSION ON THE CASE STUDIES

With some elements from social-liberalism or/and welfare-state school of thought also providing of public services are generally deemed to belong to the broadened task-set of public power operating more broadly then merely as a “night-watch”. Social security-administration is perhaps the largest and in aging countries of Europe also the most money-consuming. In the Scandinavian model, which is the most extensive realization of social support, direct social services are roughly divided into three sub-categories: primary health-care (including health centers, receptions from nurses and general practitioners, child health clinics and on-call duties), special health care (containing surgeon services and appointments to medical specialists) and social welfare services (consisting of various social benefits, sickness allowances, guardianship services, public trusteeship for legally incompetent etc.). Depending on the point of view, education could perhaps be categorized here as a function of welfare states, but then again, a vast number of countries have organized their school systems on different bases. To enable all these services and core functions effective tax administration is also needed to follow, implement, collect, and supervise tax definition and transactions of taxes. Moreover, in the contemporary ICT-era also digital services provided for citizens, internal use of the state apparatus, and international collaborative purposes are an essential part of services delivered by public bodies.

In a relatively complex modern public sector bureaucracy also supportive facilities are needed. As civil services rely heavily on intellectual properties, social and communicative skills, and substance service capabilities, effective and sharp-eyed HR management is needed to both make wise recruitments for public bodies and take care of the work-wellbeing and work stamina of the employees. Wages are a significant and perhaps even the largest budget range in developed countries. Depending on the scope of the public sector its procurements can vary from moderate to virtually

immense. So how public money is being used in investments and obtaining of goods or services has a great deal to do within public and even national economies.

Regulation in the stiffness of inviting tenders to the public procurement varies, although the EU has implemented a certain framework into it for the EU states, where not all OECD countries are included though. Transferring such significant sums of money obviously requires developed internal finance services. The predictability of public spending and especially multi-year large public investments require careful budgeting with sufficient number of sub-items and lines. Keeping this all in good order requires well-functioning services of internal finance. As mentioned before, orderly IT services are also essential for civil servants in order to enable them to keep vast amounts of data in order, process the inquiries and be accessible to citizens, cooperation between the international authorities as well as inter-authority cooperation.

The InSiCE also contains certain attributes or adjunct-categories. These are perhaps more advanced, subtle, or even partly “spiritual” parts of the public services. In this six-fold, half of them are primarily related to internal affairs of the civil service and half of them being external or towards citizen-inclined. Capabilities of the staff are perhaps the most tangible or easily measurable of these. Staff engagement is perhaps an heir of the participative bottom-up or at least bi-directional interaction-doctrine of public governance. From a HR point of view, it certainly is important to motivate and bind employees to the agency by inter-action. Although the “ocean liner” of public power stereotypically turns slowly, it is healthy to give the most old-fashioned and stiff practices an airing through innovation. On the other hand, public power is also the guarantor of stability in the society so not everything needs to be altered merely for the sake of innovation.

Openness is the first InCiSE attribute inclined towards citizens. Openness as well as transparency are important in public action as public service takes place with citizens

or more accurately taxpayers' monies. There is a varying degree of legislation defining how large part of public information and documents in each bureau type are entirely public. The premise however is that all public information is openly available for public examination unless there is a specific reason not to. These reasons can be state security in military, police, and diplomatic affairs or individuals' right to privacy when health data or matters related to minors are being processed. Also long have we come from days of sole tax-collecting military-states. Public inclusiveness has been increased with great care in the western world by many means. Democracy is the most important, but also various forms of citizen panel's, open feedback, surveys, and appointments for civil society actors to affect public decision-making. Inclusiveness also includes aspects of equality for various minorities or otherwise underrepresented groups. And as a complex sum of all this arise the ethical standards or, in one word, integrity of public governing. So, there is nothing especially alarming wrong with the operationalizations of the InCiSE per se.

Also, the question of which services belong to the public body in addition to the minimum legislative and security-services duties vary from extensive Scandinavian welfare state to Southern-European insights in which most nurture- and care services are provided by family or religious communities, such as the Catholic church as suggest by Esping-Andersen. This also highlights the degree to which religion has shaped ways of perceiving the world and arranging social norms – an influence that cannot be considered insignificant. Also, the geography, natural resources, climate, language, local temperaments and not to forget, the academic scholarly traditions have contributed to how people think of the basis of public living in certain cultural hemispheres. In these countries included in InCise at least the Anglo-American, German-Roman, French, Scandinavian, Eastern-European (Slavic) and Japanese schools of thought occur.

One essential drawback within InCiSE might be that for example the elaborateness of German-Roman "staatslehre"-spirit hasn't duly been captured with current

operationalizations. The robust independence of stately life and thinking within the Germanic hemisphere hasn't formed or co-evolved with benchmarking-mindset. This might have caused Anglican-American countries to over-perform in the InCiSE study on the cost of Germanic countries such as Germany, Austria and Switzerland. Then also the size of the state matters on its sub-division in addition to the state philosophy behind. Some large countries such as Germany and the USA have a constituent state system. In Germany it is based on the old Germanic kingdoms that were united by national spirit in the aftermath of Napoleonic Wars and "with iron and blood" as Otto von Bismark put it. In this federation-system constituent states are closer to individual states of a larger federation rather than relatively state-controlled regions under highly centralized national rule such as for example the Finnish provinces have used to be (at least before the regional administration reform).

To condense: here we bump into the first of many pitfalls which Demmke has shown up in his research and lectures (Demmke, 2020a & 2020b). Has the state-philosophical, political, or administrative schools of thought or ways of arranging been accordingly taken into consideration in operationalization of the InCiSE study? Or worse yet has this been a blind spot within the research positioning either deliberately or inadvertently making the assumption that some doctrine of state-organization is superior to others. On the other hand, benchmarking is about ranking countries. At least four out of five of the TOP-5 being countries are from the Commonwealth of independent states, representatives of British doctrine.

If we go more thoroughly through possible areas of development in data collection within InCiSE, we locate some insights to ponder. To start with, respondents of the survey have been leading civil servants with centered data collection. So this survey is a bit like the early art-rankings: a highly elitist study. In contrast to the Happiness Index with wide and decentralized obtainment of data. Those who answered or provided the data are top-dog or at least high-ranking civil servants in each of the regimes. Sampling isn't very wide but on the other hand with GDRP having taken place such information

cannot be collected or stored by any common man for national and data-safety reasons. Partly due to the highly specialized nature of this survey not everybody's expertise would be sufficient. Since most ministries are still monocratic offices it is at least partly natural that the chief answers to international matters of this proportion.

Also as mentioned before up-to-date data or occasionally even data in general or at all might not have been at disposal from certain countries or certain variables. Consequently, statistical procession and especially subjective interpretation used while filling the gaps may have caused minor overlaps and biases. Not to mention that risk of slight linguistic misunderstandings can have occurred since the survey was available only in six languages: English, German, French, Italian, Portuguese, and Spanish. Three first being official procedural languages in the EU and latter three otherwise major world-languages in the recipient-countries.

InCiSE-comparisons were not made, and a results report published with the previous two-year cycle which would have placed it to the year 2021. It may very well be that the governments and their officials have been extremely busy with their hands tied into resolving the Ukrainian and Covid-crisis over the past years. Therefore, there has been no time to think about benchmarking or to answer this kind of theoretical survey when the main focus has been to keep the citizens as much alive as possible while simultaneously attempting to keep some sort of grip on the national economies.

However, the state or at least some form of governing is an ancient institution. Political ideologies and powers as well as religions have formed how public life has been arranged in various lands. As stately life is a tradition almost as old as the military and in most parts a conservative world, this tradition originates to the twilight of written human history and continues to contribute to public administration policies even in these days. Even without going especially deep to state-philosophy, at least certain policy choices still share opinions on the fundamentals in the arrangement of the public sector. Especially the question between public – private involvement and

ownership is still much debated. One extreme is full state ownership in the form of communism, while the other extreme being minimal (if not at all) state intervention for example some of the most republican states of the U.S. midwest, such as Texas. There of course has to be a side note here about the Hobbesian social contract which exists in at least some form in all civilized countries excluding nature tribes in rural areas of the world (Lloyd & Sreedhar, 2018).

Now at the end with some words of ponder about happiness, inspired by the World Happiness Index. Since Finland has won fifth time in a row. This seems quite a breathtaking result for the country that has suffered from depression – or at least seasonally affected one – how can we suddenly be the happiest people in the world? As already presented in the summary section of case studies, the Happiness Index isn't actually about happiness but rather about having high social and living standards. Objective measurement on experiences of subjective happiness might be impossible or at least wait for several years of technological advancements to have microchips or something else to measure data from internal physical and mental processes – and yet this to be ethically questionable, at least without the consent of the test subjects. Perhaps we have to remain in the methodological approach to study objective happiness for the spirit of scientific research. Curious though it seems, that money itself cannot buy happiness. It can merely be used to buy the living standards that accumulate happiness.

How about the Finns? Perhaps we have used our guts (sisu) to turn the living standards that were unfavorable for centuries into more favorable ones. To climate or geography, we can do little. Rather utilization of technological advances has been and is a more intellectual way. We have had to build our houses, roads, and infrastructure to last both extreme cold and heat. Perhaps that's why we have so much advanced engineering knowledge and capabilities in Finland. Electric lights have brought us a victory over seasonal affective disorder. Under centuries of rule under mighty military powers a wealth of knowledge and skill concerning that sector has also accumulated to

us. Having utilized that to create even in world terms a unique military system with much-popularity enjoying general conscription and heavily armed, accurate artillery we are now as full NATO-members something of a source for international interest.

So even the dismays of geopolitical position have been managed to overcome one way or another with diligence and intellect. Even as a sparsely populated land we have managed to concentrate our resources to create a network of social support in the form of Nordic welfare-state seconded by few. Let us hope that the public economy is kept in balance so that the benefits of this welfare-Sampo can be enjoyed further in the future. So, against the odds we have managed to create one of the highest living standards in the world from a scratch.

7 CONCLUSIONS

The stage of international politics in 2020's is turmoiled by the wrangling between the USA and China. However neoliberal Anglican American hegemony seems not to be diminishing away. Therefore Benchmarking as a suitable endeavor for the zeitgeist of our time is here to stay. Some might even say that adopting best-functioning practices in economics, foreign policy and military branches are needed to stay one step ahead in this rivalry between the East and the West. Benchmarking within domestic policies and comparison between public governments or standards of living are not likely to vanish away either. At least the globally mobile intellectual and workforce elite is likely to choose very carefully the universities it studies and the countries it stays for long-time employment.

On the financial side RRO's continue to yield global benchmarking on state economies, business-friendly countries and even efficiency between major worldwide corporations. Within sciences the triumph of quantification and significance of measurable metrics is also likely to prevail. Compilation-style dissertations have taken the winning edge within academic publications while monographies appear to be disappearing folklore.

World itself within political, economical or social aspects appears to be increasingly more intertwined. Therefore understatement of complexity has become a necessity for nearly any leading politician, CEO or social scientist. Even methodologically operationalisations have become more complex when the amount of correlating factors increases. Therefore if not knowing complexity theory per se at least capability to perceive more complicated setups in research can be a great asset to a scholar.

Both case studies are good examples of this multifactorial research. The amount of correlating factors and strands of interconnectedness reach high if not even tremendous numbers when multi-national or international comparative studies are being made. Artificial intelligence or at least increased computing power can help us to do research on statistical correlation and probability-calculation. However creating the research frame is still a craftsmanship art (for the time being). Machines can do calculations better than humans, but drawing conclusions and making interpretations still require human contribution.

As for the case studies. Metrics on what to study and how can be honed. That is also the key idea in the school of thought within metric research in social sciences. The transformation from qualitative research paradigm to quantitative paradigm in humanities is not always easy, especially since we don't yet have a SI-system for social studies. Perhaps that will be created some day, who knows? On the other hand, research groups within "soft sciences" could benefit from having experts on mathematics and even physics in the composition for they could contribute a great deal to converting social phenomena into calculatable ones.

InCiSE deals with functionalities, effectiveness and transparency within state governmental apparatuses. It is a relatively new study, which could and should be developed in order to overcome certain "teething troubles". The World Happiness Index deals more with social standards of living in a wide setup of countries but from more of the angle how citizens perceive these matters. One could say that where InCiSE compares the trained professionals creating and providing social standards WHI

compares how people living in these countries experience these services. With the differentiation that WHI has a larger setup of countries although lesser number of indicators. Within the global Internet era both the rulers of countries as well as citizens living in them keep staying curious about how their home country holds its own in the increasingly accelerating global competition.

LIST OF REFERENCES/BIBLIOGRAPHY

Arizona State University. (2023). *Wealth of Nations.* Center for Political Thought and Leadership. Retrieved 3-4-2023 from address: https://cptl.asu.edu/wealth-of-nations

Balakrishnan, S. (2018). *At the edge of chaos – making sense of nonsense.* Retrieved 2022-04-18 from address: https://beeshankar.medium.com/at-the-edge-of-chaos-making-sense-of-nonsense-e2947d09ac0f

Banzhaf, W. (2009). *Self-organizing* S*ystems*. SpringerLink. https://doi.org/10.1007/978-0-387-30440-3_475. Retrieved 2022-04-18 from address: https://link.springer.com/referenceworkentry/10.1007/978-0-387-30440-3_475

Barats, C. (2020). *Dissemination of international rankings: Characteristics of the media coverage of the Shanghai Ranking in the French press.* Palgrave Communications, 6(1), 1–11. doi:10.1057/s41599-020-0440-5

Bellesguard Gaudi. (2021). *The Roman God Janus and Bellesguard.* Retrieved 2022-04-28 from address: https://bellesguardgaudi.com/en/the-roman-god-janus-and-bellesguard/

Blavatnik School of Government. (2019a). *The International Civil Service Effectiveness Index. Results Report*. Oxford: Oxford University.

Blavatnik School of Government. (2019b). *The International Civil Service Effectiveness Index. Technical Report 2019.* Oxford: Oxford University.

Blavatnik School of Government. (2017). *The International Civil Service Effectiveness Index Main Report*. Oxford: Oxford University.

Britannica. (January 19th 2023). *Taylorism.* Retrieved from address: https://www.britannica.com/science/Taylorism on 3-4-2023.

Complexity Research Group (2023.) *What is complexity research?.* Retrieved 2023-04-29 from address: https://www.uwasa.fi/en/tutkimus/ryhmat/kompleksisuustutkimus. University of Vaasa.

Cooley, A. & Snyder, J. (eds.) (2016). Ranking the World, Cambridge: Cambridge University Press

Demmke, C. (2020a). *Benchmarking – can we trust? The case of the InCiSE technical report 2019.* [Lecture at University of Vaasa 18.2.2020.] Vaasa: University of Vaasa.

Demmke, C. (2020b). *Benchmarking governments and public administrations How does it work? Does it?* [Lecture at University of Vaasa 11.2.2020.] Vaasa: University of Vaasa.

Esping-Andersen, Gøsta. (1989). *Three Worlds of Welfare Capitalism.* John Wiley & Sons. New Jersey, USA.

Future Learn. (2022). *Sensitivity to initial conditions and extreme events.* Retrieved 2022-04-18 from address:

https://www.futurelearn.com/info/courses/systems-thinking-complexity/0/steps/20399

Helliwell, J. F., Layard, R., Sachs, J. D., De Neve, J.-E., Aknin, L. B., & Wang, S. (Eds.). (2022a). *World Happiness Report 2022*. New York: Sustainable Development Solutions Network.

Helliwell, J. F., Layard, R., Sachs, J. D., De Neve, J.-E., Aknin, L. B., & Wang, S. (Eds.). (2022b). *World Happiness Report 2022*, *Statistical Appendix 1 for Chapter 2*. New York: Sustainable Development Solutions Network.

Helliwell, J. F., Layard, R., Sachs, J. D., De Neve, J.-E., Aknin, L. B., & Wang, S. (Eds.). (2022c). *World Happiness Report 2022*, *Appendix_2_Data_for_Figure_2.1*. New York: Sustainable Development Solutions Network.

Helliwell, J. F., Layard, R., Sachs, J. D., De Neve, J.-E., Aknin, L. B., & Wang, S. (Eds.). (2022d). *World Happiness Report 2022*, *Appendix_2_Data_for_Table_2.1*. New York: Sustainable Development Solutions Network.

Hyyryläinen, Esa. Sopimuksellisuus, talous ja johtaminen. *New Public Management sopimusohjauksessa ja julkisten organisaatioiden sopimustenhallinnassa.* (2004). Vaasan yliopiston julkaisusarja. Vaasa.

Institute for Government, UK. *Our work > Whitehall > International Civil Service Effectiveness Index (InCiSE)*. (2022). Retrieved 07–22–2022 from address: https://www.instituteforgovernment.org.uk/our-work/whitehall/international-civil-service-effectiveness-index-incise

Juan, D., Abascal, F., Jungreis, I., Martinez, L., Rigau, M., Rodriguez, J.M., Vazquez, J, Tress, M. L. *Nucleic Acids Research*, Volume 46, Issue 22, 14 December 2018, Page 12194, https://doi.org/10.1093/nar/gky1146.

Suomalaisen kirjallisuuden seura/ The society for Finnish bibliography. (2023). Kansallisbiografia/Finnish National Bibliography. *Oxenstierna, Axel (1583-1654).*

Retrieved 4-3-2023 from address: https://kansallisbiografia.fi/kansallisbiografia/henkilo/3236. Limited or restricted access.

Kelley, J. G., & Simmons, B. A. (2015). *Politics by number. Indicators as social pressure in international relations.* American Journal of Political Science, 59(1), 55–70. doi:10.1111/ajps.12119

Lloyd, S.A. & Sreedhar, S. (2018.) *Hobbes's Moral and Political Philosophy.* Stanford Encyclopedia of Philosophy. Stanford University. Retrieved 2021-03-07 from address: https://plato.stanford.edu/entries/hobbes-moral/

Macintosh, David. (2012). *Plato: A theory of Forms.* Retrieved 2022-04-03 from address: https://philosophynow.org/issues/90/Plato_A_Theory_of_Forms

Mau, S. (2019). *The metric society: On the quantification of the social.* Medford, MA: Polity Press**.**

Mäntylä, N. (2020.) Justice in society: concepts and discussion of constitutional state. A lecture at University of Vaasa. Restricted access.

Pernaa H-K. (2020). *Hyvinvoinnin toivottu tulevaisuus. The aspired future of welfare.* Acta Wasaensia 443.

Ringel, L., & Werron, T. (2020). *Where do rankings come from?* A historical-sociological perspective on the history of modern rankings. In A. Epple, W. Erhardt, & J. Grave (Eds.), Practices of comparing. Towards a new understanding of a fundamental human practice (pp. 137–170). Bielefeld: Bielefeld University Press. doi:10.14361/9783839451663

Stanford University. (2021). *Definition of Neoliberalism in Stanford Encyclopedia of Philosophy.* Department of Philosophy. Retrieved 3-3-2023 from address: https://plato.stanford.edu/entries/neoliberalism/

Tuominen, K. (2016). *Introducing Benchmarking*. Turku: Oy Benchmarking Ltd.

United Nations. (1948.) *Universal Declaration of Human Rights.* Retrieved 2021-03-07 from address: https://www.un.org/en/universal-declaration-human-rights/

Van Dooren, W. (2018). *Measuring Public Administration: A Feasibility study for better comparative indicators in the EU.* Brussels: European Commission. Directorate General for Employment, Social Affairs, and Inclusion.

Vartiainen, P.; Ollila, S.; Raisio, H.; Lindell, J. (2013). *Johtajana kaaoksen reunalla – kuinka selviytyä pirullisista ongelmista? / As a leader on the verge of chaos – how to cope with wicked problems?.* Gaudeamus Helsinki University Press.

Vernon, J. (2017.) *Butterfly effect.* American Scientist, vol 105, no 3. Retrieved 2022-04-04 from address: https://www.americanscientist.org/article/understanding-the-butterfly-effect

World Health Organization. *Constitution.* 2022. Retrieved 2022– 07–29 from address: https://www.who.int/about/governance/constitution

LIST OF KEY CONCEPTS DEFINED

Key concepts have been defined in the text at those points they have been more thoroughly studied. Here is a summary of them in alphabetical order.

Benchmarking is a method in which a number of factors, typically public body, or business institutions, regions, countries, or habits of doing things are operationalized, data of them is collected and based on the analyzed data factors are ranked. The idea of benchmarking is to provide information on functions of the benchmarked matters for decision makers and individual citizens based on which choices can be made. There also is a possibility to obtain information on what has been done properly in the benchmark, the one ranking highest in the preference order.

Case study is a form of study where a certain concretic example is highlighted in order to better illustrate some theory/theories. Or vice versa case studies can be used on commissions to better explain why some practical organization or matter functions in a certain way.

Complexity is a field of study within many fields of sciences but especially administrative sciences. Complexity is interconnectedness between people, matters, countries, social or natural phenomena or progressions. It is contrasted to linear, straight-word. Complexity-thinking originates from chaos theory and is used to explain and better understand complicated phenomena. Also, discussion of wicked problems and approaches to their resolving is part of complexity-thinking.

Happiness Index is a benchmarking survey written by independent experts that studies Happiness in countries around the world. It consists of eight studied factors that form the concept of objective happiness. 146 countries have been analyzed and ranked based on these factors after thorough data analysis. The report also contains articles on the physiological base of happiness, methods of tracking emotional expression with the help of artificial intelligence in social media and pondering the concept of balance and how it is related to happiness.

InCiSE is an International Civil Service Effectiveness Index commissioned by the EU and implemented by Oxford University. It has been carried out twice, 2017 and 2019 in the form of a benchmarking survey. Its aim is to analyze and compare the civil services of various OECD countries and rank them. One idea has also been to standardize administrative practices within the EU area and to share information to peers on areas that are especially well in a certain country concerned.

New Public Management or NPM is a neoliberal administrative doctrine that emphasizes productivity and responsibility for outcomes in the public sector. It has implemented certain management and leadership policies from business life into public management.

Rating is a form of survey in which a vast amount of data is operationalized, collected, analyzed and thereafter published. Ranking is a more economically oriented form of rating usually listing performance of its subject in the competitive markets.