

PALACKÝ UNIVERSITY
FACULTY OF SCIENCE
DEPARTMENT OF DEVELOPMENT STUDIES

Tereza SUCHÁNKOVÁ

Comparative Study on Indicators of Aid Quality

Bachelor Thesis

Supervisor: Mgr. Miroslav Syrovátka, Ph.D.

Olomouc, 2016

I declare in lieu of oath that I wrote this thesis myself. All information derived from the work of others has been acknowledged in the text and a list of references is given.

Olomouc, 10. 4. 2016

.....

Tereza Suchánková

Hereby, I would like to express my special thanks to my supervisor Mgr. Miroslav Syrovátka, Ph.D. for his expert guidance and valuable advice. I would also like to thank my family and friends for their support.

UNIVERZITA PALACKÉHO V OLOMOUCI
Přírodovědecká fakulta
Akademický rok: 2014/2015

ZADÁNÍ BAKALÁŘSKÉ PRÁCE
(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Tereza SUCHÁNKOVÁ**
Osobní číslo: **R12317**
Studijní program: **B1301 Geografie**
Studijní obor: **Mezinárodní rozvojová studia**
Název tématu: **Komparace ukazatelů kvality rozvojové pomoci**
Zadávací katedra: **Katedra rozvojových studií**

Z á s a d y p r o v y p r a c o v á n í :

V posledních letech panuje mezi donory shoda, že nestačí pouze navyšovat objemy rozvojové pomoci, ale důraz musí být kladen také na její kvalitu. Práce má za cíl analyzovat a zhodnotit vybrané aspekty tří ukazatelů, které měří kvalitu poskytované rozvojové pomoci u jednotlivých donorů. Jmenovitě se jedná o indexy, jejichž autory jsou Easterly & Williamson, Knack et al., Birdsall et al. (index QuODA).

Rozsah grafických prací: **dle potřeby**
Rozsah pracovní zprávy: **10 - 15 tisíc slov**
Forma zpracování bakalářské práce: **tištěná/elektronická**
Seznam odborné literatury:

Birdsall, N., Kharas, H. 2014. The Quality of Official Development Assistance (QuODA). Center for Global Development.
http://www.cgdev.org/sites/default/files/QUODA_final_revised_september.pdf
Easterly, W., Williamson, C. 2011. Rhetoric vs Reality: The Best and Worst of Aid Agency Practices. World Development 39 (11), 1930-1949.
Knack, S., Rogers, H., Eubank, N. 2011. Aid Quality and Donor Rankings. World Development 39 (11), 1907-1917.
OECD 2012. Aid Effectiveness 2011: Progress in Implementing the Paris Declaration, Better Aid. OECD Publishing, Paris.
<http://dx.doi.org/10.1787/9789264125780-en>.
3rd High Level Forum on Aid Effectiveness. 2008. Accra Agenda for Action.
<http://siteresources.worldbank.org/ACCRAEXT/Resources/4700790-1217425866038/AAA-4-SEPTEMBER-FINAL-16h00.pdf>.

Vedoucí bakalářské práce: **Mgr. Miroslav Syrovátka, DiS.**
Katedra rozvojových studií

Datum zadání bakalářské práce: **30. dubna 2015**
Termín odevzdání bakalářské práce: **15. dubna 2016**

L.S.

prof. RNDr. Ivo Frébort, CSc., Ph.D.
děkan

doc. RNDr. Pavel Nováček, CSc.
vedoucí katedry

V Olomouci dne 30. dubna 2015

ABSTRACT

In connection with foreign aid, more and more attention has been put on its effectiveness lately. The aim of this study is to compare three indexes (Knack et al. 2011, Easterly & Williamson 2011, Birdsall et al. 2010, 2011, 2014) measuring donors' approach towards effective aid. An explanation why the results of indexes change is also incorporated. Indexes are compared in term of their technical aspects of construction, as well as in term of their content. We found that the index by Knack et al. (2011) is inexcusably poorly transparent and that calculations in the index of Easterly & Williamson (2011) do not necessarily match the theoretical base. We found that QuODA is technically best constructed.

Key words: Aid effectiveness, foreign aid, Paris Declaration on Aid Effectiveness, QuODA, best practices on aid effectiveness, Composite indicators

ABSTRAKT

V souvislosti se zahraniční pomocí se kromě jejího navyšování v posledních letech zaměřuje pozornost také na její efektivnost. Cílem této práce je porovnat tři indexy (Knack et al. 2011, Easterly & Williamson 2011, Birdsall et al. 2010, 2011, 2014) měřící přístup donorů k efektivní pomoci. Součástí práce je vysvětlení, proč se výsledky indexů liší. Indexy jsou porovnávány jednak z pohledu technických aspektů konstrukce, jednak z pohledu jejich obsahové stránky. Zjistili jsme, že Knack et al. (2011) je nedostatečně transparentní, a využití kalkulace u Easterly & Williamson (2011) ne vždy odpovídají teoretické bázi. Shledali jsme, že QuODA je technicky i obsahově nejlépe řešená.

Klíčová slova: Efektivnost pomoci, Pařížská deklarace o efektivnosti pomoci, QuODA, praktiky v efektivní pomoci, složené ukazatele

TABLE OF CONTENTS

- Introduction.....11**
- 1. introduction to aid effectiveness.....13**
 - 1.1. Definition13
 - 1.2. Academic approach13
 - 1.3. Political Declarations.....15
 - 1.4. Measuring aid effectiveness.....16
 - 1.4.1. QuODA.....18
 - 1.4.2. Index by Easterly & Williamson (2011).....19
 - 1.4.3. Index by Knack, Rogers & Eubank. (2011).....20
 - 1.4.4. Brief comparison20
- 2. Index construction24**
 - 2.1. Theoretical framework.....24
 - 2.1.1. Theoretical framework: Theory24
 - 2.1.2. Theoretical framework: Analysis25
 - 2.2. Normalisation.....27
 - 2.2.1. Normalisation: Theory.....27
 - 2.2.2. Normalisation: Analysis29
 - 2.3. Weighting.....32
 - 2.3.1. Weighting: Theory32
 - 2.3.2. Weighting: Analysis33
 - 2.4. Aggregation.....35
 - 2.4.1. Aggregation: Theory35
 - 2.4.2. Aggregation: Analysis36
- 3. Good practices of aid effectiveness.....38**
 - 3.1. Key themes in aid effectiveness38

3.2. Selectivity	39
3.3. Specialization/Fragmentation.....	46
3.4. Transparency	50
Conclusion	54
List of references	56
List of Appendices	61

LIST OF TABLES AND FIGURES

Table 1. Number of overlapping indicators	21
Table 2. Correlation matrix (Spearman correlation coefficients)	22
Table 3. Comparison of normalisation methods	30
Table 4. Good practices of aid effectiveness in the indexes	39
Figure 1. Differences in donors' ranking	22
Figure 2. Z-scores for sectoral concentration (KRE)	30
Figure 3. Z-scores for sectoral concentration (KRE) without the extreme value	31
Figure 4. EW – correlation of final rankings for percentile ranks and z-scores	31
Figure 5. Implication of headcount measures (EW).....	43
Figure 6. Selectivity performance (EW).....	45
Figure 7. Selectivity performance (KRE).....	45
Figure 8. Selectivity performance (QuODA).....	46
Figure 9. Gaps between transparency ranks for QuODA and EW.....	52

LIST OF ABBREVIATIONS

AAA	Accra Agenda for Action
CDI	Commitment to Development index
CPA	Country Programmable Aid
CPIA	Country Policy and Institutional Assessment
CRS	Creditor Reporting System
CGD	Center for Global Development
DAC	Development Assistance Committee
EP	index by Easterly & Pfutze (2008)
EW	index by Easterly & Williamson (2011)
GVI	Governance Vulnerability Index
HHI	Herfindahl-Hirschman Index
IATI	International Aid Transparency Initiative
ICRG	International Country Risk Guide
KRE	index by Knack, Rogers & Eubank (2011)
MDGs	Millennium Development Goals
ODA	Official Development Assistance
OECD	Organization for Economic Cooperation and Development
QuODA	Quality of Official Development Assistance index by Birdsall et al.
PD	Paris Declaration on Aid Effectiveness
RCA	Revealed Comparative Advantage
SMPD	Survey on Monitoring Paris Declaration
UN	United Nations

INTRODUCTION

If foreign aid is to serve its purpose, it is necessary for rich countries to provide sufficient volume of financing and to ensure that these funds are used in the most effective way possible. Aid effectiveness is a topical issue, it was debated especially in connection with fulfilling Millennium Development Goals (MDGs).

“We recognise that while the volumes of aid and other development resources must increase to achieve these goals [MDGs], aid effectiveness must increase significantly as well to support partner country efforts to strengthen governance and improve development performance. This will be all more important if existing and new bilateral and multilateral initiatives lead to significant further increases in aid.”

Paris Declaration on Aid Effectiveness (OECD 2005, 1).

Aid effectiveness refers to using given resources in a way that the desired impact is maximized. Such an approach leads to savings of valuable resources. The final aid effectiveness depends on donors, recipient countries as well as on many external factors. This thesis focuses solely on the donors' role in aid effectiveness. Donors can reach aid of good quality through respecting certain practices which are considered best for aid effectiveness. The debate about these practices is still in progress.

We believe that this topic deserves attention because of several reasons. Firstly, there is little aid effectiveness in donors' practices of delivering aid. Secondly, the international aid community puts effort to increase aid quality and recently, important political pledges concerning aid effectiveness were signed. Thirdly, it seems that there is a turning point in the perception of aid effectiveness. The Paris Declaration stresses the principle of ownership saying that developing countries have leadership in development cooperation; the whole system should be newly based on collaboration and joint effort.

Assessing donors according to their aid quality is relatively young field of study. Academic literature is a source of only a handful of multidimensional indexes measuring donors' aid effectiveness. Besides, the fulfilment of declarations is monitored in connection with political pledges.

The aim of this thesis is to compare three indexes of aid effectiveness – indexes by Easterly & Williamson (2011) (EW), by Knack et al. (2011) (KRE) and by Birdsall et al.

(2010, 2011, 2014) whose index is called Quality of Official Development Assistance (QuODA). These indexes were published roughly at the same time and they should reflect a consensus on best practices of aid effectiveness but the ratings of donors differ significantly in some cases. Sweden, as an example, was ranked among the last ten countries in EW, but among the top ten countries in KRE. On top of this, QuODA assessed Sweden as an average donor.

The study will attempt to clarify following questions: 1) How are the indexes constructed regarding technical aspects? Do they match commonly accepted standards for index construction? 2) How did indexes cope with measuring individual practices? Do the formulas match the theoretical base? The analysis will also reveal strengths and weaknesses of individual indexes and what caused the different ratings of donors.

In terms of methodology, the study makes use of comparative analysis and critical assessment. In other words, the three indexes (and their components) are compared with each other to reveal the differences among them and a following discussion critically assesses them. Correlation analysis serves as a tool for checking linear relationship between indicators. In most cases, we prefer Spearman correlation coefficient which is insensitive to extreme values and data outliers. The thesis also contains a number of tables and figures for visualisation of data and a better insight for the reader.

The main contribution of this study lies in highlighting strengths and weaknesses of indexes, which can help to avoid possible misleading practice in the future. Only a little number of comparative studies was conducted in the aid effectiveness literature. The most similar study was carried out by Clist (2015) who focused on the comparison of selectivity components of five chosen indexes.

The study is divided into three chapters. A following chapter provides an insight into the issue of aid effectiveness – it presents basic academic literature as well as main political declarations regarding this topic. One part of the chapter is dedicated to the problem of measuring aid effectiveness by donors. It also contains an introduction to the three indexes which are subjects of this study, and their brief comparison. Second chapter examines technical part of the indexes by assessing how certain methodological issues match commonly accepted standards of index construction and what effects the selected methodology can have on final results. Third chapter focuses on the approach of indexes towards individual themes of aid effectiveness literature. Initially, it provides an insight into particular themes and into their importance, then, selected themes are discussed in connection with the indexes. Findings are summarized in the conclusion.

1. INTRODUCTION TO AID EFFECTIVENESS

1.1. Definition

Aid effectiveness is a self-explanatory term - not even the Paris Declaration offers definition for this concept. Nevertheless, Stern et al. (2008) extracted a definition from the principles of the Paris Declaration in their independent evaluation and defined aid effectiveness as “arrangement for the planning, management and deployment of aid that is efficient,¹ reduces transaction costs and is targeted towards development outcomes including poverty reduction” (Stern et al., 2008, vii). This definition stresses results, aid efficiency and embraces basic principles of the Paris Declaration.

1.2. Academic approach

Since the mid-1970s, the question ‘Does aid work?’ has been dominant in aid debates (Glennie & Sumner 2014, 9). Majority of studies examined foreign aid from the macroeconomic perspective. In other words, they examined the relationship between volume of aid and economic growth and the prevailing conclusion was that aid effect is insignificant or even negative (Mosley et al. 1987, Boone 1996, Burnside & Dollar 2000, Bourguignon & Sundberg 2007, Rajan & Subramanian 2008). As a response to these findings, studies on aid effectiveness emerged. Their aim is to answer the question ‘When is aid most likely to work?’ by setting conditions under which aid has a positive effect in developing countries. Thus, since late 1990s, the studies on foreign aid have moved towards examining which forms of aid are more effective than others. Radelet (2006) identifies three categories of conditions under which aid is more effective: The characteristics of the recipient economy, the characteristics of the recipient policies and the characteristics of aid.

A good many studies pursue the topic of aid allocation as the path to increase aid effectiveness. Especially one issue – policy selectivity – is discussed in the aid community. The studies examine foreign aid vis-a-vis different governance aspects and in general, they find a positive relationship between good policies and effectiveness of foreign aid (Burnside & Dollar 2000, Kosack 2003).

¹ Effectiveness and efficiency have slightly different meanings. ‘Effectiveness’ refers to the extent of accomplishing intended results, whereas ‘efficiency’ refers to accomplishing results with the least waste of resources.

Studies based on this finding are of two main types. Normative studies create models of ‘optimal’ aid allocation, it usually means such an allocation where maximal poverty reduction (through economic growth) is achieved. The authors include poverty as well as policy in recipient countries among the principles of optimal aid allocation (Collier & Dollar 2001, 2002, Cogneau & Naudet 2007). The studies of the second type make use of regression models to find what factors play role in donors’ allocation decisions and how strong they are (Alesina & Weder 2002, Nunnenkamp & Thiele 2006, Dollar & Levin 2006, Clist 2011, In’Airat 2014). Surprisingly, the conclusions of these studies differ significantly.

At the same time, the perception of foreign aid changes as its impact is not referred only to rising per capita income but also to other features. As an example, Kosack (2003) studied the impact of aid on the quality of life and Askarov & Doucouliagos (2013) examined effect of development aid on democracy and governance.

Also other practices of aid effectiveness are discussed, though to a lesser extent. Roodman (2006) identified six good practices among discussed themes: untying, selectivity, harmonization, alignment, coordination, and proliferation, but it is not a comprehensive list (e.g. transparency and aid predictability are not included). Some practices are obvious by using common sense, and some of them are rather contentious. The following text will present them in short.

Transparency and coordination are examples of good practices which can be deduced by common sense. Collin et al. (2009) attempted to quantify increased transparency by cost-benefit analysis and, as expected, he found that benefits exceed costs by far. Transparency allows to coordinate donors’ activities and it is important for mutual accountability. Coordination facilitates integrated approach by donors, allows better targeting and limits duplication of projects. A study by Bourguignon & Platteau (2015) suggests why there is too little of these two practices. They envisage coordination as a public good and the whole problem as n-player coordination game. They find that the more donor countries coordinate the more benefits they acquire. But coordination imposes certain costs and donor countries tend to free ride.²

² Free rider problem is connected to public goods. A country can free ride if it uses a public good (benefits of coordination) without paying its share of costs (e.g. collecting information, sharing data). Remaining actors have less incentives to finance such a public good if some actors do not participate and as a consequence, the good is provided in sub-optimal level.

Tied aid refers to aid whose procurement of goods and services are limited to the companies in the donor country. This practice is believed by aid community that it rather serves the interests of donors than recipients and it increases costs of development projects.

Some practices are efficient only to certain extent. For example, literature suggests that current excessive fragmentation is not optimal as it imposes too much burden on recipient countries, increases transaction costs for donors and hinders economies of scale (Annen & Kosempel 2009, Anderson 2012, Roodman 2006). Studies examining the impact on economic growth find that fragmented aid is less effective (Djankov et al. 2009, Kimura et al. 2007). But some voices emerge that neither complete specialization is optimal. They argue that more donors mean more points of view on a given problem with the possibility to find more convenient solutions (Han & Koenig 2015) and that some level of competition among donors is important (Frot & Santiso 2010).

1.3. Political Declarations

The academic discussion about good practices is followed by political declarations that attempt to revise foreign aid. They define principles of good aid relationship with the aim of making aid more efficient through donors' behaviour. These declarations came out of four High Level Fora on aid effectiveness that took place in Rome (2003), Paris (2005), Accra (2008) and Busan (2011). The Rome Declaration sketches basic principles (alignment to partner country priorities, cooperation, leadership by partner countries, transparency). Negotiations in Paris gave birth to the most cited document – the Paris Declaration (PD). The importance of this document lies in the fact that, in contrast to the Rome Declaration, it contains commitments of donors and recipient countries, establishes targets and a monitoring system for measuring progress towards these targets. The outcome of the monitoring system is the Survey on Monitoring Paris Declaration (SMPD)³. This independent assessment by OECD/DAC measures progress in aid effectiveness and highlights lagging areas.

The PD is based on five partnership commitments – ownership, alignment, harmonisation, managing for results and mutual accountability. These broad categories embrace multiple practices. Donors pledge to respect the leadership of partner countries in their development (*Ownership*), they pledge to align with partner country priorities and strategies (*Alignment*), mainly through untying aid, and using country systems (procurement

³ In total, there were three SMPDs – in 2006, 2008 and 2011

and financial systems). *Harmonisation* contains the commitment that donors will avoid activities that undermine institution building (i.e. using programme-based aid rather than project aid), they will be more transparent (through reporting to governments about aid flows and their activities) and they will coordinate to reduce duplicative missions. Furthermore, donors commit to implement aid with the aim to reach desired results (*Managing for Results*) and together with partner countries, they accept to be accountable for results (*Mutual accountability*).

The Accra Agenda for Action (AAA) and the Busan Declaration are regarded as extensions of the PD. They confirm the principles and stress particular practices. The AAA builds on the same five principles as the PD and stresses predictability of flows, ownership of development projects by partner countries and more inclusive partnership. The Busan Declaration summarizes commitments into 4 broad categories: ownership of development priorities by developing countries, focus on results, inclusive development partnership, transparency and accountability to each other. The document also contains topics corresponding with the MDGs – gender equality, sustainability and climate change. The progress is measured by the Global Partnership for Effective Development Co-operation⁴ and first report was launched in 2014.

Compared to the PD, these two extensions broadened the base of stakeholders participating in negotiations. In Accra, these new stakeholders were invited to negotiations, whereas the Busan Partnership is an agreed framework of all, new as well as old stakeholders. The Paris negotiations were held only by ministers of developed and developing countries and Heads of multilateral and bilateral development institutions (OECD 2005). The Busan Partnership is an agreement among many types of donors. It states, “We, Heads of States, Ministers and representatives of developing and developed countries, heads of multilateral and bilateral institutions, representatives of different types of public, civil society, private, parliamentary, local and regional organisations...recognise that we are united by a new partnership that is broader and more inclusive than ever before...” (OECD 2011, 1).

1.4. Measuring aid effectiveness

Aid effectiveness is in itself hardly measurable concept. Both, a donor as well as a recipient country take a share in aid effectiveness. In addition, it is hard to distinguish the real

⁴ It is a multi-stakeholder platform created at the High-level forum in Busan that was mandated to create a set of indicators and monitor the progress

effect of foreign aid as many factors (which are often hard to quantify) play role in the final outcome. This is the reason why indexes measure aid effectiveness indirectly – they try to grasp the potential of the given foreign aid and assess if the aid is more likely or less likely to be effective. This is achieved by assessing donors according to the extent to which they follow generally recognised best practices of aid effectiveness.

First attempts to capture donors' performance concerned mainly the allocative performance. A great number of studies was launched (McGillivray 1989, White & McGillivray 1995, Rao 1997, Nunnenkamp & Thiele 2006, Anderson & Clist 2011) which differ by employing different criteria according to which the donors are assessed. Not surprisingly, the final rank of donors differs as well. Measurements of other practices are rather in their infancy.

Only a limited number of studies has addressed the issue of donors' overall performance by creating a multi-dimensional index. Apart from KRE (Knack et al. 2011), EW (Easterly & Williamson 2011) and QuODA (Birdsall et al. 2010, 2011, 2014), which are the indexes of the direct interest, we will shortly present indexes by Easterly & Pfitze (2008) and by Roodman (2006b).

'What would an ideal agency look like?' is the fundamental question for Easterly & Pfitze (EP) (2008). They identify practices an ideal agency should follow and the donors are ranked based on these criteria. EP consists of five dimensions – transparency, fragmentation, selectivity, ineffective aid channels and overhead costs – which are aggregated by averaging donor's percentile ranks for each dimension (equal weights are used). The authors conclude that: "(1) the data are terrible, and (2) the patterns the data show are terrible" (Easterly & Pfitze 2008, 51). In general, they find that donors do not follow practices they say they should follow. Development banks ranked best and the UN agencies worst.

The Commitment to Development Index (CDI) by Roodman (2006b) rates donors on the 'development friendliness of their policies'. It comes with much broader concept of assessing donors as only one component addresses aid⁵ and remaining six components concern other donors' policies towards developing countries (Finance, Technology, Environment, Trade, Security, Migration). The CDI is published by the Center for Global Development (CGD) and it is updated annually.

⁵ Aid component is composed of two sub-components assessing quantity of aid and quality of aid. The latter makes use of QuODA rates in the calculation.

1.4.1. QuODA

QuODA stands for Quality of ODA index. The goal was to create a “framework that provides summary information in a quantitative fashion on donor efforts to improve aid effectiveness” (Birdsall & Kharas 2014, 2). It was created by N. Birdsall and H. Kharas from the CGD. The first edition came out in 2010 and used dataset for the year 2008. The second edition followed in 2011 and the last third edition in 2014. After the first edition, the methodology was revised and the second edition contained some marginal changes.⁶ For the comparability of data over the years, the first edition was recalculated according to the new methodology.

There are 31 indicators in total which are grouped into four dimensions. These dimensions should reflect best practices on aid effectiveness. Three of them contain eight indicators, and the remaining one (Reducing Burden) contains seven. 1) *Maximizing Efficiency* dimension assesses donors’ allocation of aid taking into account consensus about how to achieve poverty reduction. 2) *Fostering Institutions* measures to which level donors use partner country systems. The more, the better rating because this practice should shift ownership of projects towards recipient countries and strengthen their institutions. 3) *Reducing Burden* is linked to administrative costs donors impose on recipient countries. Donors with low fragmentation of aid and high degree of cooperation are ranked better. 4) *Transparency and Learning* dimension is based on assumption that “recipient countries benefit most from aid when they know what their donor partners are doing” (Birdsall et al. 2010, 21). Transparency also allows to comment on agency performance and can consequently lead to improvement within the agency in the area of aid effectiveness. In addition, it is a necessary precondition for coordination among donors.

Donor sample consists of 23 bilateral agencies (all of them are members of OECD/DAC) and 8 multilateral agencies. Five UN agencies are aggregated and treated as a single donor (QuODA 2010, 5).

Indicators in each dimension are aggregated into four sub-indexes. Nonetheless, the authors decided not to aggregate the sub-indexes into an overall index for two main reasons. Firstly, QuODA was created with the purpose of identifying donors’ strengths and weaknesses, which are not possible to read from an overall index. Secondly, low correlations

⁶ There were several changes in the methodology for the second edition. The total number of indicators increased from 30 to 31. See more changes in Birdsall et al. 2011, p. 2. The methodology for the third edition did not change.

among sub-indexes would cause high sensitivity to any choice of weights. Instead, they present separate indices for each dimension in the form of ranking and so called Quality of Aid Diamond, which is a spider diagram.

1.4.2. Index by Easterly & Williamson (2011)

EW measures donors' performance in five dimensions which are based on the practices the donors themselves suggest agencies should follow. The index distances itself from the PD or the AAA saying that "they are negotiated political process rather than an academic monitoring exercise" (Easterly & Williamson 2011, 1930), so apart from the PD, the best practices raise from aid agency documents and academic literature. The study was conducted by W. Easterly and C. Williamson from the Development Research Institute and published in *World Development* in 2011.

The 5 dimensions are following: 1) *Transparency* measures the accessibility of information about agency's management. The authors mark this component as possibly the most important. 2) *Overhead Costs* tries to penalize rampant bureaucracy. 3) *Specialization/Fragmentation* measures the concentration of donors' aid by sector and country. 4) *Delivery Channels* indicators penalize tied aid, food aid and technical assistance, which rather reflect the interests of donors than the needs of recipients, 5) *Selectivity* is linked to donors' choice of recipient countries. Donors providing aid to poor countries with good governance and little corruption are awarded most.

Each sub-index contains two or three indicators. Three of these dimensions contain indicators which award donors for high score. For example, more specialization means better rank in this dimension. Two remaining indicators (Delivery channels and Overhead costs) have the opposite logic, they award donors for low score. For example, lower score in overhead costs indicators means better rank in this sub-index.

The sample consists of 33 bilateral and multilateral donors plus nine UN agencies. The average score and ranking is provided for all three categories of donors – bilaterals, multilaterals and UN agencies. The output of this paper is the percentile rank of donors for each sub-index, as well as the overall rank for the aggregate index.

1.4.3. Index by Knack, Rogers & Eubank. (2011)

KRE is an index assessing quality of aid by donor countries. The study was first published in 2010 as a working paper under the heading of the World Bank and one year later, in the World Development. The contributions of the study lie in 1) including a more comprehensive set of indicators than have been used in the past and 2) adjusting them (by regression) for important factors that are not directly under the control of donor agencies. According to the authors, the former should more reflect the content of the PD and lower the sensitivity of weights assigned to any one indicator or inclusion of a new one. The latter should better reflect how effectively the aid agencies are managing.

The index is composed of four sub-indexes - Selectivity, Alignment, Harmonization and Specialization - which, according to the authors, capture key aspects of aid quality. They have equal weights in the overall index. 1) *Selectivity* dimension measures to which extent donors' aid flows to countries where it will be most effectively used. 2) *Alignment* measures to which extent the donors' aid is aligned with country policies and systems. 3) *Harmonization* dimension penalizes donors who impose high transaction costs to recipient countries and 4) *Specialization* measures penalize fragmentation of recipients, sectors and projects. In total, 18 variables are categorised into these sub-indexes, varying from 2 to 7 for each sub-index. Alignment and Harmonization dimensions contain only indicators borrowed from the SMPD whereas the rest of indicators have its roots in academic literature.

The sample consists of 38 donors (including non-DAC members at that time). Donors with missing data on more than half of the indicators or on more than one sub-index were excluded from the study. Data from several UN agencies are aggregated and UN is treated as a single donor.

The paper presents ranks of donors in each sub-index to demonstrate in which aspects donors are doing poorly, as well as the ranks of donors in the overall index. It was created with the intention to improve the measurement of aid quality, and to present some methodological considerations.

1.4.4. Brief comparison

When comparing the three indexes described above, their main objective seems to be very similar - all of them assess how donors stick to global standards about aid effectiveness. They all build on the PD, however, in different extent. QuODA and KRE closely follow the

PD – KRE borrows one half of its indicators from the SMPD. In case of QuODA, the number is nine out of 31 indicators. EW is the most critical towards the PD and describes it only as one of many sources on what the consensus about aid effectiveness is (Easterly & Williamson 2011, 1931) and do not directly borrow any indicators. In terms of academic literature, EW as well as KRE draws on the paper ‘Where does the money go?’ by Easterly & Pfutze (2008), which is one of the first attempts to capture donors’ performance by a multidimensional index. QuODA uses many sources in academic literature, but none of them as the main one.

Table 1 shows how the indexes overlap. The diagonal line reveals the total number of indicators in the indexes and the remaining values express the number of overlapping indicators between the pairs of indexes. From the table, it is obvious that KRE and QuODA share many of the same indicators – mostly the indicators borrowed from the SMPD. For KRE, 16 out of 18 indicators (88.9%) overlap with QuODA and 5 out of 18 indicators (27.8%) overlap with EW. For EW, 7 out of 13 indicators (53.8%) overlap with QuODA and 5 out of 13 indicators (38.5%) overlap with KRE.

Table 1. Number of overlapping indicators

	QuODA	KRE	EW
QuODA	31	16	7
KRE	16	18	5
EW	7	5	13

Source: author

The indices differ in how broadly they take the concept of aid effectiveness. In general, the scope matches the number of included indicators. QuODA employs 31 indicators and its dimensions are defined the most broadly of all the 3. For example, its Maximizing efficiency dimension overlaps with 2 dimensions in EW - Selectivity and Specialization - and also partly overlaps with the same 2 dimensions in KRE. On the contrary, dimensions in EW, which contain only 13 indicators in total, are defined very narrowly. In comparison with the remaining 2 indices, KRE does not contain any measures of transparency. EW does not have any measures of coordination nor any measures about how donors use institutions in recipient countries when delivering aid nor how donors contribute to reducing administrative burden imposed on recipient countries. The structure of the indexes can be found in Appendix 1.

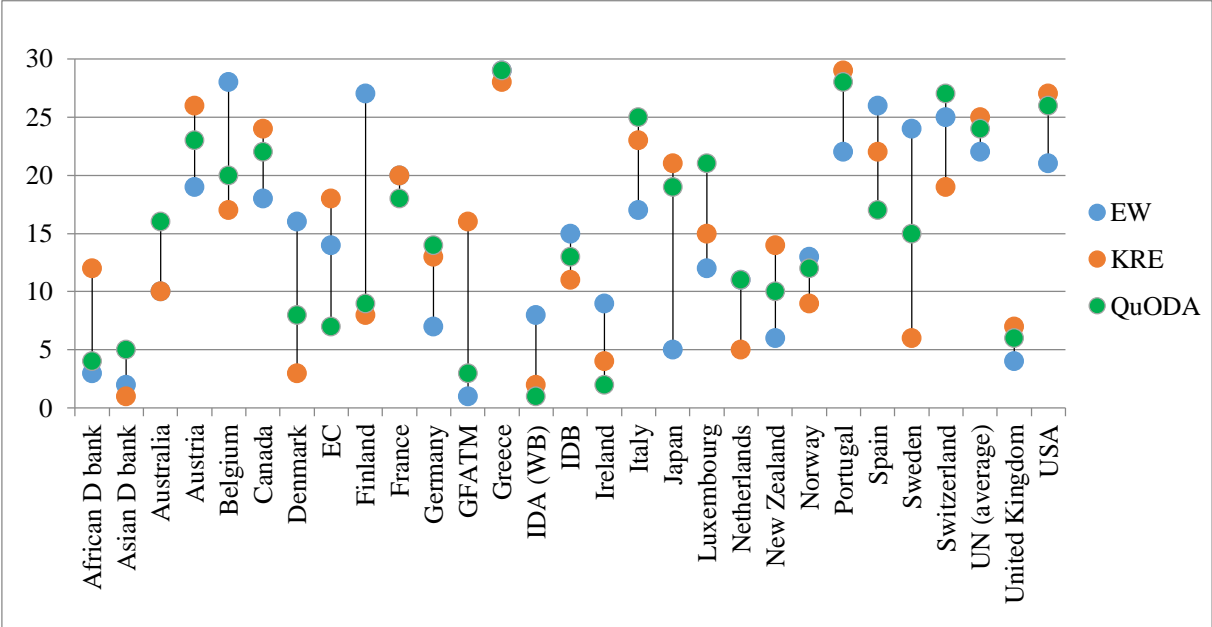
Table 2. Correlation matrix (Spearman correlation coefficients)

	EW	KRE	QuODA
EW	1	0.488	0.669
KRE	x	1	0.804
QuODA	x	x	1

Source: author, using Birdsall et al. (2010), Easterly & Williamson (2011), Knack et al. (2011)

Table 2 is a correlation matrix displaying values of correlation coefficients of the pairs of indexes. For each pair, the common set of donors was chosen, the rank of donors was adjusted for this new set and then, the correlation analysis for donors’ ranks was conducted. In case of QuODA, the dimensions were aggregated by using unweighted average to obtain the overall rank of donors. EW and QuODA (first edition) made use of 2008 dataset and KRE used 2007 dataset so that data are fairly timely comparable. The table shows that the highest correlation is between QuODA and KRE. On the other hand, only moderate correlation (0.488) was found between ranking of KRE and EW. The last pair, QuODA and EW show moderate to strong correlation.

Figure 1. Differences in donors' ranking



Source: author, using Birdsall et al. (2010), Easterly & Williamson (2011), Knack et al. (2011)

Figure 1 shows differences in donors' ranking for all three indexes. Firstly, we identified a common set of 29 donors which are contained in all three indexes. Secondly, rank of these donors was accordingly adjusted so that it can take values from one to 29. The greatest disparities in ranking are for northern donors – Finland, Sweden, Denmark – that are rated significantly worse in EW than in the remaining two indexes. On the other hand, another donor with great disparities – Japan – is assessed much better in EW than in the remaining two indexes. All three indexes place Asian development bank, IDA, Ireland and United Kingdom among the top ten donors. On the other side, Greece, Portugal, UN and USA are allocated among the last ten donors.

2. INDEX CONSTRUCTION

This section assesses selected aspects of index construction. The main focus is put on how the three indexes match commonly accepted standards about composite indicator construction, as stated in OECD Handbook on Composite Indicators (OECD, 2008) and other studies, and how the selected methodology affects final output. Does the technical part contribute to different results? Regarding the structure of this section, each theoretical part about a particular technical issue is followed by a practical part where the issue is examined on three chosen composite indicators - QuODA, EW and KRE. The study attends to following steps for developing composite indicators: theoretical framework, normalisation, weighting and aggregation.

2.1. Theoretical framework

2.1.1. Theoretical framework: Theory

What is a theoretical framework good for? It's a key element for creating a meaningful composite. When the whole system of index creation is not underpinned by a good quality and transparent theoretical framework, the output will be senseless. Academic literature (OECD, 2008, Freudenberg, 2003) feature a series of normative recommendations about how theoretical framework should look like. 1) The main concept should be well defined in the way that it is clear what exactly is being measured and why. 2) The structure of the composite must be obvious. It should be clear what dimensions are the driving forces behind the composite and what links are between the dimension and the phenomenon measured in the composite. 3) Criteria for selecting variables should be clearly defined. "The selected variables should carry relevant information about the core components and be based on a paradigm concerning the behaviour being analysed" (Freudenberg, 2003, 7).

OECD (2008) and Freudenberg (2003) both point out that not all indexes have clearly developed theoretical underpinning. As a matter of course, some phenomena are harder to measure than the others. It can be very demanding to create a strong theoretical framework for a multi-dimensional abstract phenomenon, which is defined too broadly or too vaguely, and the knowledge and empirical evidence are limited in that area, e.g. sustainability or human development. Does it mean that multidimensional indexes about these concepts should be abandoned?

Ravallion's (2010) attitude towards this kind of indexes is sceptical – he calls multidimensional indexes with unclearly defined theoretical framework ‘mashup indices’. Among others, he warns that some indexes are in reality too distant from their theories which gives them credibility. And he concludes that some aspects cannot be captured in a single-number index. Although one can say that he proceeds to extremities, it is necessary to realise that the importance of theoretical framework is unquestionable. OECD (2008, 22) aptly adds that “what is badly defined is likely to be badly measured.”

2.1.2. Theoretical framework: Analysis

All three indexes determine what they measure as follows: KRE is a measurement of aid quality, EW assesses to which extent donors follow best practices of aid effectiveness. The fundamental question is, “Do agencies perform the way they say they should?” (Easterly, 2011, 6). QuODA is a “framework that provides summary information in a quantitative fashion on donor efforts to improve aid effectiveness” (Birdsall, 2014, 2). QuODA and EW also define what they do not measure – they are not indices of aid effectiveness. Firstly, aid effectiveness results from efforts of both, donors as well as recipients. Secondly, indicators are not adjusted for matters which are out of agencies control, e.g. geographical or sectoral mandate.

In spite of different formulation of definitions, it seems that only little nuances are between the measured phenomena. When looking at the structure of the indexes, it is obvious that their dimensions reflect the same thing – best practices of aid effectiveness. Nevertheless, some indexes determine more clearly what is being measured. EW presents itself as a measure of aid best practices, which is more precise and clear definition of what is being measured than in case of KRE that presents itself as a measure of aid quality. What aid quality means is not defined in the study, and it is rather a vague concept.

As mentioned above, the individual dimensions of the indexes contain measures of donors' adherence to best practices of aid effectiveness. All three indexes are composed of 4 or 5 dimensions. Their structure can be found in Appendix 1.

QuODA's *Maximizing Efficiency* sub-component takes over mainly from the academic literature. It is a broad dimensions embracing basic practices such as selectivity, proliferation, tied aid, administrative costs etc. *Fostering Institutions* borrows many indicators from the PD, it contains indicators on alignment, harmonisation and aid predictability. *Reducing Burden* (which is one of fundamental ideas of the PD) measures how donors reduce costs imposed on

partner countries but it seems that without considering if these practices are efficient – as an example, reduction in costs may be achieved by channelling aid through less agencies and therefore allowing less agency-recipient country relationships. It can also be achieved by the preference for large projects over small ones (as every project has some fixed costs which are relatively smaller in case of large projects). In this case, the effectiveness of such a practice is disputable. *Transparency and Learning* dimension contains proxies measuring transparency and evaluation practices.

For KRE, the number of indicators in the sub-components vary from two to seven, which implies that some indicators have much higher effective weight in the composite than the others. Indicators in *Alignment* and *Harmonization* dimensions are all borrowed from the PD. *Specialization* contains seven heterogeneous indicators some of which reflect specialization only remotely (average project size, contribution to multilaterals, administrative costs). The theoretical part above (see chapter 3.1.1) suggests that construction of an index progresses from general to specific steps - the procedure starts with defining the measured phenomenon, then defining driving forces (components) beyond the phenomenon and last, defining indicators which measure specific components. At this point, it can be interesting to look at the opposite procedure in KRE. After defining the main concept, particular indicators were chosen and subsequently categorized into dimensions according to common sense. One may say that Specialization contains indicators which suited nowhere else.

EW is typical by narrowly defined sub-components containing only two to three indicators.

Aid effectiveness as a topic belongs among these where knowledge and empirical evidence are missing (see chapter 3.1.1). The designers of indexes well realized this limitation. For example, KRE (Knack, 2011, 3) notes, “Most indicators of donor performance are based on plausible but largely untested beliefs about best practices in aid management.” EW (Easterly, 2011, 12) notes in the same spirit, “All of the studies (including ours) cannot demonstrate evidence that our measures of aid quality or aid practices are directly related to aid impact, since again we have no measure of the latter.” We can draw conclusion that the three indexes are not based on strong theoretical basis, but the authors are aware of that.

Unsatisfactory theoretical basis moves indexes towards subjective assessments and subsequently towards the use of so called “common sense“. Because measures about aid impact are unavailable, “All studies in this literature have to appeal instead to a common sense consensus that very bad performance on the indicators would make a strong positive aid impact less likely” (Easterly, 2011, 13).

What is the theoretical basis for individual dimensions? The authors advocate their choice by using academic literature, international agreements on aid effectiveness, aid agency documents or simply their intuition. EW often uses the attitude „What the most agencies agree on“, whereas QuODA and KRE use primarily the PD. Some areas (e.g. selectivity) are missing in the PD but the authors considered important adding them. In general, the level of consensus is higher for some components and lower for other ones, some components are even controversial. Such an unclear effect on aid effectiveness can be observed for technical assistance. In EW, technical assistance is taken as negative, because it is mostly tied (as it requires hiring experts from donor countries). However, Knack et al. (2011) argues that technical assistance is not always tied and the final impact would be rather positive. A study by Annen & Kosempel (2009) also finds a positive relationship between aid in the form of technical assistance and economic growth.

In conclusion, it is clear what the indexes measure and what the driving forces are behind the indexes, though the concept of KRE is rather broad and vague compared to EW that clearly specifies what is being measured. Concerning the structure of KRE, some disparities in indicators were found in Specialization dimensions. We suggest to divide this dimension into two – one containing indicators on specialisation and one containing indicators on reducing burden. Such a step would balance number of indicators in the dimensions and the indicators would better match the measured phenomenon.

2.2. Normalisation

2.2.1. Normalisation: Theory

Selected indicators come in different measurement units and in different scales. For example in EW, total ODA per employee is expressed in US dollars whereas Administrative costs are expressed as a percentage of total ODA. The former varies from 0.03 to 60.97 US dollars, the latter has larger scale – from 0.24% to 129%. The aim of normalisation is to transform them to common basis, so that they can be comparable among themselves and aggregated into a composite. The following text will present the most common normalisation methods:

1) *Ranking* – this method assigns rank to each unit of the dataset, so that it allows to observe relative position of statistical units. Nevertheless, the information in absolute terms is lost. Over time, the position of one statistical unit can improve or worsen without a real

absolute change, depending on the absolute change of other statistical units. It implies that, when taking a particular statistical unit, the value of an indicator can improve in absolute terms but without improvements in the rank and consequently this improvement does not project into the composite index. The advantage of ranking is that extreme values do not distort the output.

A special form of ranking is *percentile ranking*. It expresses the percentage of scores that are the same or below a given score. If the percentile rank of a particular value is 0.75, it means that 75 % of the values from the dataset are the same or below this score. In contrast to simple ranking, different weights can be applied to indicators during aggregation process. After the conversion, variables fall within 0-100 interval, zero is assigned to the lowest value and 100 to the highest one.

2) *Z-scores* express distance of a given value from the mean, measured in the standard deviations. The data are normalised according to formula $z_i = \frac{x_i - \bar{x}}{\sigma}$ (\bar{x} – mean, σ – standard deviation). It implies that the converted indicators are assumed to have a standard normal distribution (with the mean of 0 and the standard deviation of 1). The values below mean will get negative values and vice versa. As minimal and maximal values are not fixed, every indicator can have different range. The presence of an extreme value causes that the other values in the indicator shrink more closely to the mean (zero) and as a consequence, they have lower weight in the composite. “This might be desirable if the intention is to reward exceptional behaviour, i.e., if an extremely good result on a few indicators is thought to be better than a lot of average scores” (OECD, 2006, 28).

3) *Min-Max* – the indicators are re-scaled into an identical range [0,1] according to the formula $m_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}$, so that zero is assigned to the worst result and one to the best one. Compared to the previous method, minimum and maximum are fixed after the conversion, but mean and standard deviation change for each indicator. This normalisation method may affect final results by increasing the range for indicators with very little variation. As a consequence, these will contribute more to the composite indicator than they would if un-scaled method is used (Jacobs 2004, 38).

4) *Distance to a reference* – the reference point, which can be for example a target, a mean or a maximal value, is assigned value 1 (or 100%). Data are then transformed to percentage points away from the reference point.

Each normalisation method produces different outcome and each one is suitable in different cases. Implicit weights, which are inherent part of every method, should be taken into account. When using some methods, a special attention should be given to outliers and extreme values. Freudenberg (2003, 10) suggests to log variables with highly-skewed distributions and truncate data if there are extreme outliers.

2.2.2. Normalisation: Analysis

EW transforms indicators into percentile ranks. The methodology differs for dimensions with commensurate and non-commensurate indicators.⁷ For dimensions with commensurate indicators (Transparency and Selectivity), individual indicators are first aggregated and afterwards, the average index is transformed into percentile ranks. For the latter, (Overhead costs, Selectivity, Ineffective channels), individual indicators are first transformed into percentile ranks, which is followed by aggregation. Then, the average percent ranks are retransformed into percentile ranks again. As a whole, the transformation into percentile ranks is run twice for these non-commensurate indicators.

KRE and QuODA both use *z*-scores.⁸ In some cases, they both log data – either to stress the lower spectrum of the scale or to deal with extreme values and data outliers. As an example of the latter, QuODA (Birdsall et al. 2009, 33) logs indicator RB3, Median project size, “Scores were computed based on the log of median project size, to de-emphasize outliers in the distribution.”

Table 3 with model values shows different outcomes when using percentile ranks and *z*-scores. Percentile ranking is not sensitive to distance between scores in absolute terms. This can be illustrated by indicators A and A*. Although indicator A contains an extreme value, the percentile rank will be the same for both indicators. The overall score counted as an average of A and B percentile ranks would be 50 for each country.

On the contrary, *z*-scores reflect the gaps between values well. Compared to indicator A*, the extreme value in the indicator A raised the mean so much that the remaining standardized values are below the mean. The transformed values are also more shrunk towards the zero mean as a consequence of higher standard deviation and their relative weight in the composite is lower than without the extreme value. In practice, Figure 2 shows original

⁷ Commensurate indicators are those which are measured in the same units and the same scales

⁸ KRE transforms values into *z*-scores. After the first aggregation, scores for individual dimensions are re-transformed into *z*-scores again.

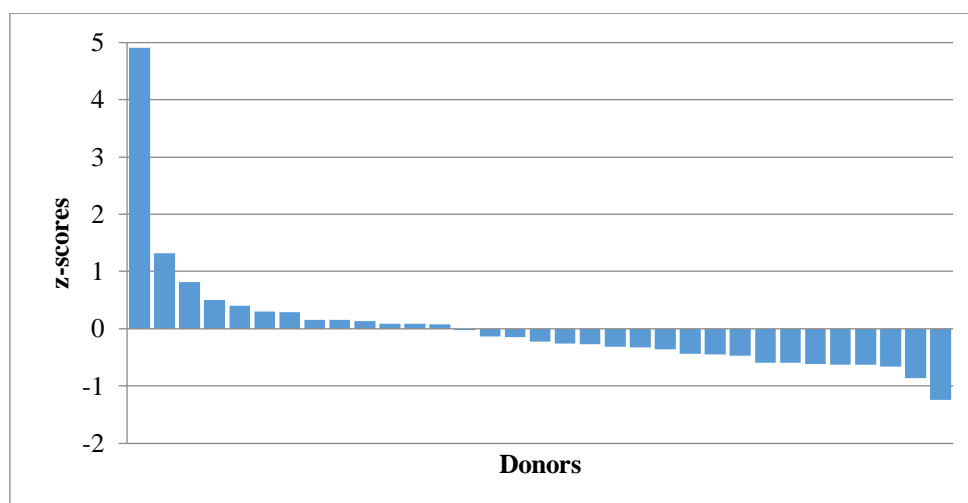
z-scores for Sectoral concentration (KRE) with one extreme value. To see how this extreme value affects other z-scores, we removed it (Figure 3). As a consequence, the value of the remaining z-scores increased significantly so that their weight in the sub-index would be higher without the extreme value.

Table 3. Comparison of normalisation methods

	Indicator A	Indicator A*	Indicator B	Percentile rank A	Percentile rank B	Z-scores A	Z-scores A*	Z-scores B
Country 1	20,00	20,00	900,00	0,00	100,00	-0,79	-1,41	1,06
Country 2	30,00	30,00	850,00	25,00	75,00	-0,64	-0,71	0,77
Country 3	50,00	50,00	800,00	50,00	50,00	-0,34	0,71	0,47
Country 4	60,00	60,00	600,00	75,00	25,00	-0,18	1,41	-0,71
Country 5	200,00	40,00	450,00	100,00	0,00	1,95	0,00	-1,60
Mean	72,00	40,00	720,00	50,00	50,00	0,00	0,00	0,00
St dev	65,54	14,14	169,12	35,36	35,36	1,00	1,00	1,00
Min	20,00	20,00	450,00	0,00	0,00	-0,79	-1,41	-1,60
Max	200,00	60,00	900,00	100,00	100,00	1,95	1,41	1,06

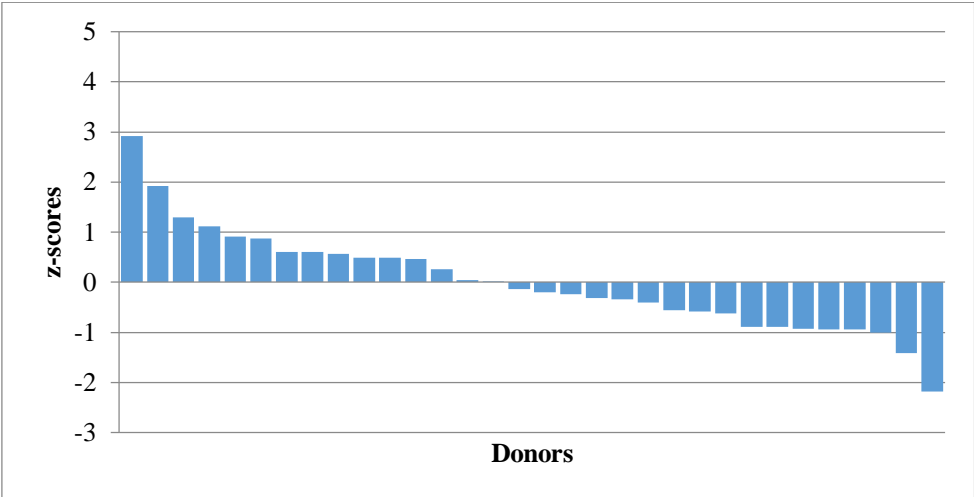
Source: author

Figure 2. Z-scores for sectoral concentration (KRE)



Source: author

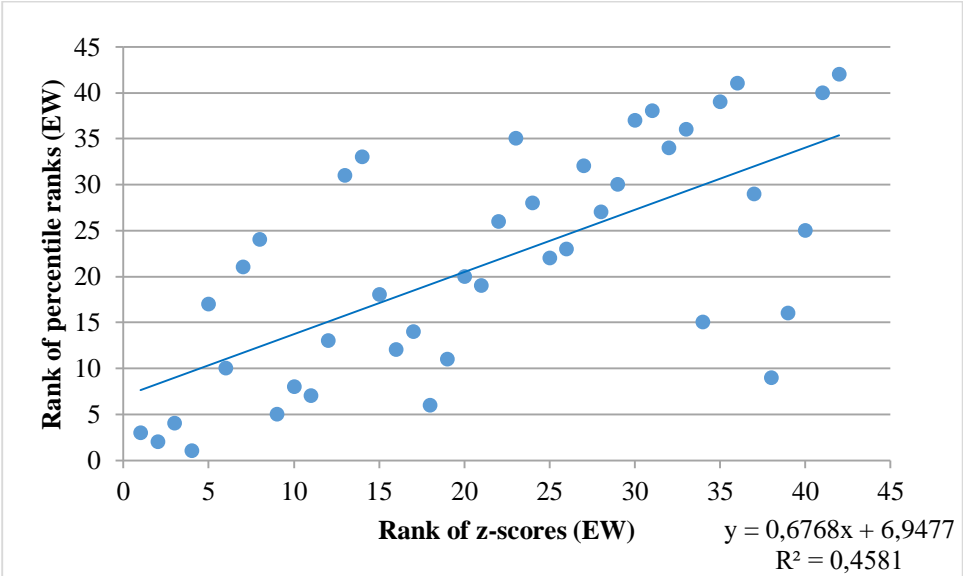
Figure 3. Z-scores for sectoral concentration (KRE) without the extreme value



Source: author

How will the rating of donors change if different normalisation method is employed? To find it out, we changed normalization method for EW from percentile ranking to z-scores, ceteris paribus (see Figure 4). Z-scores of indicators in Ineffective Channels and Overhead Costs dimensions were multiplied by minus one because of their character of inverse proportion (more is not better). We find only moderate correlation (0.68) between ranks produced by using percentile ranks and by using z-scores. On this basis, it may be inferred that normalization method significantly influences final results.

Figure 4. EW – correlation of final rankings for percentile ranks and z-scores



Source: author, using Easterly & Williamson (2011)

In conclusion, EW gives priority to the rank of donors and their relative comparison whereas KRE and QuODA keep the absolute information after the normalisation. They also markedly reward or punish extreme behaviour. Using different normalisation method may cause significant changes in the final rank as demonstrated on the example of EW. None of the three studies advocate the chosen method of normalisation.

2.3. Weighting

2.3.1. Weighting: Theory

As the next step, weights need to be assigned at the level of individual indicators, as well as at the level of sub-indexes if they are aggregated into a composite. “All variables may be given equal weight or they may be given differing weights which reflect the significance, reliability or other characteristics of the underlying indicators” (Freudenberg, 2003, 12). Therefore, the assigned weights indicate the relative importance of indicators. In the academic literature (Jacobs, 2004, Freudenberg, 2003, OECD, 2008), there is an agreement that weights can significantly influence the results and hence the assigned weights should be explained transparently. This end can be served by conducting a sensitivity analysis or simply by using alternative weighting systems and looking at changes in country rankings.

There are several approaches about how to assign weights to indicators or sub-indexes. OECD Handbook on composite indicators (2008) primarily distinguish statistical methods and participatory methods. From the statistical models, principal component analysis is the most significant. It is based on correlation between indicators and consequently on the identification of a small number of common factors which are the most responsible for the variation. Principal component analysis, as well as many other techniques based on correlations, has the disadvantage that “correlations do not necessarily represent the real influence of those sub-indicators on the phenomenon the composite indicator is measuring” (Joint Research Centre, 2002, 14).

Participatory methods consist in subjective judgements on weights by various stakeholders (e.g. politicians, citizens, experts). One of these methods – budget allocation – gives experts a budget of N points to be distributed over sub-indicators. Public opinion method polls the concerns about certain problems measured by sub-indicators.

In some cases, better quality data are assigned higher weights and vice versa. Such a practice can improve data reliability, nevertheless Jacobs (2004) warns that more emphasis

may be given to indicators which are simply easier to measure rather than to more relevant indicators where good data are not available.

Many indices, however, use the system of equal weights. As an advantage, they are simple and easy to understand for the final user. On the other hand, using equal weights implies that indicators have equal importance in the sub-component, or eventually sub-components have equal importance in the composite, which may not reflect the reality. In the academic literature, indices are often criticized for equal weights (as in the case of HDI), though there is no consensus on alternative weights. According to Jacobs (2004), equal weights may be a valid approach in some context, especially when the knowledge of impacts and casual relationships is missing.

Equal weights are connected to the problem of double counting. The point is, that if two indicators are highly correlated, they might measure the same aspect and in this case, the double weight of this aspect is counted into a composite. In this case, OECD (2008) recommends to either choose only indicators with low correlation or to give less weight to correlated indicators. However, there may be disagreement about whether the indicators measure the same phenomenon.

2.3.2. Weighting: Analysis

As weights are closely linked to aggregation, it is necessary to realize that aggregation and related weighting take place on two levels – firstly, individual indicators need to be aggregated into a sub-index and secondly, sub-indexes are aggregated into an overall index. In short, all three indexes use equal weights on the lower level of aggregation.⁹

When aggregating sub-indexes into an overall index, KRE and EW use equal weights, even though there is an intuition that the dimensions do not have the same importance. For example, EW states about transparency to be possibly the most important component, however, its weight in the final composite is the same as for the other components. QuODA uses a different approach. It does not aggregate individual dimensions, partly to avoid assigning arbitrary weights and partly because the aim is to reveal where a particular country is lagging behind.

⁹ In the selectivity component, EW assigns weight as follows: 0.5 to aid going to low-income countries, 0.25 to aid going to free countries and 0.25 to aid going to non-corrupt countries, which simply means that policy selectivity and poverty selectivity have equal weights in the selectivity component.

Effective weights of individual indicators in a composite depend on the number of dimensions and the number of indicators in them. In KRE, number of indicators vary from two to seven. Therefore, each of two indicators in the selectivity dimension weigh 12.5% whereas in the specialization dimension, the effective weight of each indicator is only 3.57%. EW has more balanced number of indicators (two to three indicators in each dimension) so the effective weights vary less - from 6.67% to 10%.

The following text will pursue the topic of transparency in explication of assigned weights because as mentioned in chapter 3.3.1, choice of weights can significantly influence results. Out of the three indices, EW shows the least transparency. It roughly informs about used weights, without presenting any deeper insight. The explication, why the authors chose given weights, is missing.

Similarly, KRE does not explain the choice of equal weights. Nonetheless, they acknowledge the importance of weights, “We show that weightings do matter to the rankings... we argue that ranking exercises need to take into account this sensitivity to weightings and should make their weighting choices explicit, to allow the reader whether he or she agrees with them” (Knack, 2011, 3). One part of the study is devoted to testing alternative weights. So called index-18 and index-20 are constructed where the former contains 18 original indicators and the latter contains two more additional indicators to test the sensitivity to weightings. In both of them, each indicator is assigned equal weight (in the sense that each indicator has the same importance in the composite and sub-indexes don't play any role). The authors then analyse changes in the rank and check for robustness of rankings to different weighting. Also one statistical method – principal component analysis – is used to check for robustness. According to the authors, all these test proved that the ranking would not change much and the authors consider weighting to be robust.

QuODA is the most transparent of the three. The choice of equal weights for indicators are advocated as follows. “Our approach gives equal weights to each indicator within each dimension – the most transparent and ‘neural’ approach, though we recognize that it does represent an implicit judgment” (Birdsall, 2008, 9). In addition, as QuODA measures how performance of donors change within years, it is not likely that weights would rest the same over time. This decision was supported by principal component analysis which did not show strong concentration of the variance. Authors also addressed the issue of double counting and concluded that it is not a concern since 1) the correlation between the most of indicators is low 2) if the indicators are highly correlated, they don't measure the same issue.

2.4. Aggregation

2.4.1. Aggregation: Theory

During the aggregation process, values are merged into a single number – a composite. It is a two-level process - individual indicators can be aggregated into a sub-index and subsequently the sub-indexes into an overall index. The utilised aggregation method may, just as in weighting, significantly change the final results. Therefore, the choice should be explained in a transparent way, including possible effects on the results.

The following text will present three principal aggregation methods:

1) *Sum of ranks* – ranks of individual indicators are summed to obtain the overall score for a country. On the one hand, this method is characterized by simplicity. On the other hand, the information in absolute terms is lost.

2) *Arithmetic mean* – in this linear additive aggregation, the normalised values are merged according to formula $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ (n – number of items in the sample). This method is simple and easy to understand, but one should be aware that it allows full compensability. Arithmetic mean is widely used during construction of a composite, as an example, we mention Commitment to Development Index.

3) *Geometric mean* – the values are aggregated according to formula $\bar{x} = \left(\prod_{i=1}^n x_i \right)^{\frac{1}{n}}$. Geometric mean is, among others, used in the construction of HDI where it replaced arithmetic mean – it was agreed that geometric mean was more suitable in the context of HDI since it allows only partial compensability.

As it can be seen from above, the issue of compensability is connected with both, arithmetic and geometric mean. It causes that the actual meaning of weights is undermined. OECD (2008, 99) defines compensability as “the possibility of offsetting a disadvantage on some variables by a sufficiently large advantage on others.” Compensability in arithmetic mean is constant whereas geometric mean shows a certain degree of non-compensability. This fact will be demonstrated by following example: indicators for a country have values 1, 1, 10. After aggregating them, the result for linear arithmetic mean will be 4, exactly the same as for a country with values 4, 4, 4. Although the characteristics of two mentioned countries would be differing, they are not projected into the mean because of compensability. Geometric mean returns score 2.15 for the first country and 4 for the second country. It means that very low

scores cannot be so easily offset and to do so, much higher scores of the other values are needed.

To illustrate how marginal utility works, we will contemplate two scenarios. In each of them, one of the indicators will increase by 1 unit. In scenario A, the score of a country increase to 1, 1, 11, in scenario B to 1, 2, 10. Arithmetic mean 4.33 will be the same for both types of means whereas geometric mean is lower in the scenario A (2.22) than in the scenario B (2.71). Therefore, in geometric mean, marginal utility is much higher for an increase in low values than in high ones.

OECD (2008, 33) writes about compensability, “To ensure that weights remain a measure of importance, other aggregation methods should be used, in particular methods that do not allow compensability. Moreover, if different goals are equally legitimate and important, a non-compensatory logic might be necessary.” One of these non-compensatory methods is for example multi-criteria decision analysis based on looking for compromise solutions.

Aggregation of various dimensions into a single number is a controversial issue. Roodman (2011) speaks on the subject that when creating such an index, incomparables are implicitly compared and he labels such indexes as “inherently crass”. Similarly, Ravallion (2010) refers to the indexes composed of many various dimensions as “mashup indices”.

Although composite indices have a lot of shortages, “The absence of an ‘objective’ way to determine weights and aggregation methods does not necessarily lead to rejection of the validity of composite indicators, as long as the entire process is transparent” (OECD, 2008, 33). Roodman (2011) refers about the issue in the same spirit and he stresses the importance of clear and well-stated purpose. He also adds that “aggregation hides at least as much as it reveals” (Roodman 2011, 483).

Another approach is less aggregation – individual dimensions are not aggregated into the composite but they are presented separately. This approach allows to avoid controversial assignment of weights and choice of aggregation method, however, these separated results can be more difficult to present to the final user.

2.4.2. Aggregation: Analysis

If the attention is focused on the lower level of aggregation, that is the aggregation of individual indicators into sub-indexes, all three indexes employ arithmetic mean. In case of missing data for some indicators, the average is simply made over the remaining ones so that

donors are not penalized for incomplete data.¹⁰ KRE and QuODA take mean over z-scores. EW treats commensurate and non-commensurate indicators differently. If they are commensurate, weighted average of raw data is employed. If not, indicators are aggregated by making an average over the percentile ranks. For these dimensions with non-commensurate indicators, absolute information of data is already lost at this lower level of aggregation.

In terms of aggregation at the level of dimensions, each of the three indexes takes different attitude towards how much to aggregate. EW aggregates dimensions into the composite and presents ranking of donor agencies as the main output. KRE aggregates dimensions as well, but it recognizes the importance of both - the overall index and the individual disaggregated dimensions. QuODA does not aggregate its dimensions at all. Instead, the result is presented in the form of 4 composite scores.

EW calculates a simple linear averages of five percentile rankings. Only relative positions are averaged so it can be said that this method is a certain form of Sum of ranks mentioned in the chapter 3.4.1. The gaps between values are lost and relative positions play the crucial role. The purpose is to create a rank of countries and organizations where one can read how well a particular country is doing with regard to other countries. Percentile rank allows to see how many donors rated the same or worse.

KRE employs arithmetic average over the z-score values of the four sub-indexes. The results keep zero as the mean value, hence above-average countries are those with a score higher than zero and vice versa. The authors put a question if aggregation of different dimensions into a single number is valid and reliable. Regarding unclear weights of dimensions it would make more sense not to aggregate into a composite. On the other hand, they also argue that an overall index should not be abandoned – it can uncover which countries are doing poorly in general. To sum it up, the purpose of KRE was to look at relative comparison of donors and beyond that, to find out where individual donors perform poorly by using disaggregated sub-indexes.

QuODA resists to aggregate various dimensions – partly because of unclear weights, partly because as it states, “Our purpose is not to rank countries and agencies on some overall abstract notion of aid quality, but to identify their strengths and weaknesses so that priority areas for change can be identified for each country or agency” (Birdsall et al. 2009, 3). Four scores presented separately for each donor may say nothing to the reader about donor’s overall performance. They only make sense if visualised in a spider diagram.

¹⁰ With the exception of KRE that treats aid with unreported status as tied aid as insufficient reporting may be motivated by hiding a high ratio of tied aid.

In terms of transparency, only QuODA explains the aggregation method and sketches its possible effects on results. The indexes allow full compensability by using arithmetic mean, but they do not comment on this issue. The author of this study assumes that regarding the character of the indices, full compensability may be intended in this case.

3. GOOD PRACTICES OF AID EFFECTIVENESS

The following chapter focuses on the best practices of aid effectiveness. At large, we look at which practices are incorporated into the indexes and how they are measured. Firstly, we identified key themes that aid community deals with and we outlined a notion which practices are considered the most important. Secondly, we examined selected themes in depth. Each subsection starts with a theoretical part which summarizes how a given theme is perceived by aid community. This part is followed by comparative analysis where we examine how KRE, QuODA and EW envisage selected themes.

3.1. Key themes in aid effectiveness

Table 4 contains identified themes and subthemes which are discussed in connection with the improvement of aid quality. We examined if the indexes contain indicators measuring individual themes in the table. Our findings are presented in the Agreement field which marks how many indexes incorporated the category into their structures – 100% indicates that the theme was incorporated by all three indexes, 66% and 33% indicates that only two or respectively one of the indexes incorporated the theme. The table suggests which practices are considered the most important for aid effectiveness.

All three indexes agree on four practices of aid effectiveness. They all reward donors for more selectivity (for recipient countries where the impact of aid may be greater), less proliferation (of small projects across countries and sectors), less administrative costs (and more efficiency of agencies), less tied aid (as it rather serves interests of donors).

Two out of three indexes suggested that aid is more effective when there is more transparency (for enhancing mutual accountability), more coordination (among donors to avoid duplication), more predictable flows of aid (so that the spending can be planned in advance), more alignment (with recipients' goals and priorities), and harmonization (of activities and requirements). Only one of the indexes incorporates indicators on evaluation and learning.

Table 4. Good practices of aid effectiveness in the indexes

THEMES	AGREEMENT
Selectivity	100%
Poverty selectivity	100%
Policy selectivity	100%
Proliferation	100%
Country proliferation	100%
Sector proliferation	100%
Project proliferation	66%
Donor agencies proliferation	33%
Administrative costs	100%
Tied aid	100%
Transparency	66%
Aid predictability	66%
Coordination	66%
Learning/evaluation	33%
Alignment	66%
Harmonization	66%

Source: author

3.2. Selectivity

Aid selectivity is one of the most discussed topics in academic literature dedicated to aid effectiveness. Two types – poverty and policy selectivity – can be distinguished. Literature suggests that aid is more effective if 1) flowing to a poorer country, 2) flowing to a country with better quality of institutions and policies (Burnside & Dollar 2000, Collier & Dollar 2001, Kosack 2003, Clist 2011, Bigsten & Tengstam 2015).

Studies concerning donors' aid allocation are of two main types – explanatory and descriptive (McGillivray 2003, 1). The former seeks to explain donors' allocation of aid through identifying its determinants whereas the latter seeks to establish criteria for assessing donors' performance and for modelling the 'optimal' allocation of aid. First descriptive studies measuring donors' performance are connected with McGillivray. His performance index (1989) considers only relative needs of recipients. Each share of aid is assigned weight

according to a recipient in a sense that aid is more discounted for wealthier recipients (measured by income per capita). The Generalised Performance Index proposed by Anderson & Clist (2011) reflects the development in debate about selectivity as the proposed criteria contain not only income, but also population and policy sensitivity.

Apart from that, several studies dealt with the issue of the ‘optimal’ aid allocation among recipient countries. In other words they seek to find such an allocation that maximizes poverty reduction (through economic growth) by mathematical modelling where certain criteria for aid allocation are incorporated. The output shows the ‘optimal’ share of total aid a recipient should ideally receive in order to maximize poverty reduction globally. This output is compared to real distributions of aid – recipients with less real aid than the model predicts are called donor orphans, and vice versa if their real share of aid is higher, they are donor darlings. Collier & Dollar (2001) create such an allocation rule where an optimal volume of aid for a recipient is a function of its policy, population, poverty level, GDP per capita and poverty elasticity with respect to income. As a result, low-income countries and among them countries with better policies are favoured. A similar model was created by Collier & Dollar (2002). This time, the optimal aid allocation is a function of a recipient’s poverty level, the elasticity of poverty and the quality policies. Cogneau & Naudet (2007) go further with their model as apart from policy and poverty criteria, they also incorporate structural disadvantages that hinder poverty reduction (climate, health situation, inequality, colonial past etc.). Compared to previous two models, they do not assume that equal effort leads to equal effect of aid; their model of optimal aid allocation is based on ‘equality of opportunities concerning the risk of poverty.’

Explanatory studies employ mostly linear regression models.¹¹ They are based on identifying determinants (explanatory variables) of aid allocation and subsequently explaining their significance in each donor’s allocation of aid; regression coefficient estimates show the degree to which a given factor determines allocation decisions. Incorporated determinants are usually of both types, donor interests and recipient needs. The majority of studies focus on one factor, controlling for other factors that may influence aid allocation. A key study by Dollar & Levin (2006) examine the extent to which each donor’s assistance is aimed at countries with sound institutions, controlling for population, per capita income and other

¹¹ Multiple linear regression analysis is used to explain the relationship between one dependent variable and two or more independent variables. The formula: $y=b_0+b_1x_1+b_2x_2+\dots+b_kx_k$, where b stands for regression coefficient

determinants of aid allocation.¹² Their policy index (a coefficient estimate) indicates the extent to which each donor favours better-governed countries among those at the same level of income, *ceteris paribus*. Clist's (2011) model suppose that amount of aid from a donor to a recipient is dependent on the recipient's poverty level, governance, population and on proximity between the two (religion, language, culture, history, commercial ties, distance etc.).

QuODA, KRE and EW all measure policy and poverty selectivity which implies that aid selectivity is considered an important factor for aid quality. Nevertheless, there is no agreement about how to measure it – different formulas and variables are employed. QuODA makes use of weighting, it simply weights the share of aid flowing from a donor to a recipient country by log GDP per capita for poverty selectivity, respectively by GVI¹³ (Governance vulnerability index which is based on the Worldwide Governance Indicators) for the latter. EW uses a headcount measure - aid must meet certain criteria to get a credit. For poverty selectivity, the criterion is that the recipient country belongs among low-income countries, as defined by the OECD. Thus, share of donor's aid meeting this criterion is counted. Policy selectivity is measured by two indicators, one measuring good-governance and one corruption. To get a credit for a share of aid flowing into a particular recipient, the recipient must score 8, 9 or 10 in the Polity IV democracy ranking¹⁴ and 0 or 1 in the corruption component of the ICRG (International Country Risk Guide) political risk index.¹⁵ The final selectivity score is counted as 0.5*percentile rank (share going to low-income countries) + 0.25*percentile rank (share going to poor countries) + 0.25* percentile rank (share going to free countries), which implies, that poverty and policy components have equal weights. KRE uses regression analysis for calculation, which is run separately for each donor. The indicators of policy and poverty selectivity are thus partial regression coefficients.¹⁶ Log of aid flowing

¹² $\log(\text{aid}_{d,r}) = b_0 + b_1 \log(\text{population}_r) + b_2 \log(\text{per capita GDP}_r) + b_3 \log(\text{governance}_r) + b_4 (\text{vector of other determinants}_r)$. 'Other determinants' are dummies that may influence aid allocation - colonial past, commercial ties, distance etc.

¹³ Share of allocation to poor countries: $\sum_r \left(\frac{\text{gross ODA}_{d,r}}{\text{gross ODA}_d} * \log \text{GDP} \right)$, share of allocation to well-governed

countries: $\sum_r \left(\frac{\text{sgrossCPA}_{d,r}}{\text{sgrossCPA}_d} * \text{GVI}_r \right)$, where GDP – GDP per capita in purchasing power parity, GVI – Governance Vulnerability Index, sgrossCPA - strict gross country programmable aid.

¹⁴ Polity IV is a data series on the level of democracy. The driving forces behind are 1) evaluation of each country's elections for competitiveness and openness, 2) political participation in general, 3) the extent of checks on executive authority. The score ranges from -10 to +10. Countries that scored 6 to 10 are considered democracies.

¹⁵ ICRG rates countries on the basis of over 30 political, economic and financial risks. It is published by the PRS Group.

¹⁶ $\log(\text{aid}_{d,r}) = b_0 + b_1 \log(\text{population}_r) + b_2 \log(\text{GDPpc}_r) + b_3 \log(\text{CPIA}_r)$

to a recipient country is adjusted for three independent variables – population, GDP per capita and Country Policy and Institutional Assessment (CPIA)¹⁷ overall score.

Clist (2015) compares five aid quality indexes in terms of aid selectivity and comments that it is ‘surprisingly contentious’ which factors make up allocative performance (2015, 808). It is not so striking in poverty selectivity, where three indexes analysed in this study use GDP per capita or GNI per capita (in case of EW) as a key factor for poverty allocation calculations.

In policy selectivity, the situation is more complicated. Clist (2015) analysed EW and KRE and concluded, “The difference in variable choice reflects a more fundamental disagreement. Both sets of authors argue that they are measuring selectivity, despite very different theoretical conceptualisations of what constitutes policy selectivity” (Clist 2015, 809). He also adds that EW emphasizes corruption and political freedom whereas KRE (by using CPIA) is rather focused on recipient countries’ performance in public sector and economic management. QuODA uses GVI which is based on the Worldwide Governance Indicators,¹⁸ the main focus is therefore put on the principles of good governance.

From the point of view of calculations and utilised formulas, QuODA logs GDP per capita of the recipient countries in order to emphasize lower part of the scale and more appreciate aid going to the poorest countries (the z-scores are then multiplied by minus one).¹⁹ Compared to EW, aid to all recipient countries is counted (in practice, aid can never be weighted by zero), but each recipient country has a different weight. Donor can get highest score when sending all aid to the recipient country with the combination of how the lowest GDP per capita and at the same time how the best GVI rate. The utilised methodology²⁰ for EW implies that each recipient country is assigned one of five possible weights: 0, 0.25, 0.5, 0.75 or 1 (similar analysis was conducted in Clist 2015). Figure 5 shows the situation graphically. Aid going to recipient countries which are out of all three sets gets zero, aid to a low-income country which is unfree and corrupted gets 0.5, aid to a country which is

¹⁷ CPIA is produced by the World Bank with the aim to assess the quality of a country’s policies and institutions. The index consists of 4 components: economic management, structural policies, policies for social inclusion and equity, public sector management and institutions.

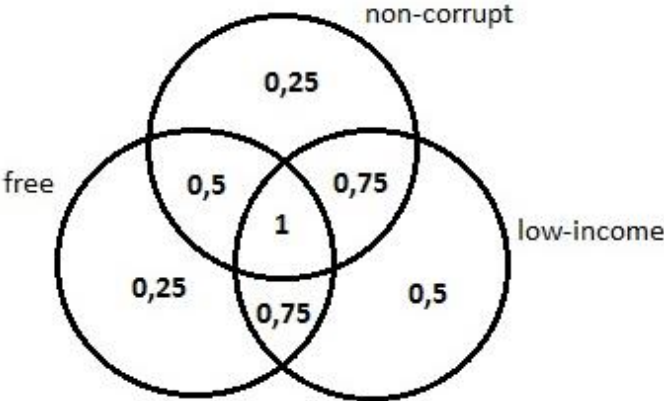
¹⁸ The WGI is a set of indices produced by K. Kaufmann, A. Kraay and M. Mastruzzi, assessing the quality of governance on six dimensions of governance: Voice and Accountability, Political Stability and Absence of Violence, Government Effectiveness, Regulatory Quality, Rule of Law, Control of Corruption)

¹⁹ The difference of reallocating a share of aid from a recipient country with GDP 1,500 to a recipient with GDP 1,000 is higher than reallocating the same share of aid from a recipient with 10,500 to a recipient with 10,000. Because of the logarithm, this difference of five hundred dollars will more project into the score for the former example.

²⁰ Donors’ selectivity score is counted as $0.5 \times \text{percentile rank (share of aid going to low-income countries)} + 0.25 \times \text{percentile rank (share going to poor countries)} + 0.25 \times \text{percentile rank (share going to free countries)}$

democratic but corrupted and not low-income gets 0.25 etc. Highest score in selectivity can be obtained by sending all aid to recipient countries belonging to the area of all three sets intersection.

Figure 5. Implication of headcount measures (EW)



Source: author

What are the drawbacks of these methodologies? Clist (2015) finds that EW is insensitive to the size of recipients’ population. The same can be applied to QuODA. The rationale for that is following: let us consider two countries with the same level of policy and poverty (expressed with regard to recipients’ total population), first country with the population of ten million and second one with one million. QuODA and EW treat them the same even though more aid is needed for the first country to reach the same effect (e.g. poverty reduction) on the population. It implies that a donor seeking to increase its score may chose a small recipient by population that meets the two criteria (policy and poverty) well and reallocate all aid to this recipient. It does not measure if aid is allocated efficiently in view of distribution among recipients and it does not incorporate diminishing returns of aid on its effectiveness.

KRE distinguishes itself from other indexes by adjusting for other factors which are not under direct control of aid agencies. Adjusting is achieved by regression which reveals weights of individual factors on the dependent variable (aid). In KRE, aid is regressed on 3 factors – population, poverty (GDP per capita), and policy (CPIA). The regression coefficients for policy and poverty are then standardized and serve as the indicators for policy and poverty selectivity. We assume that more factors should be considered in the regression analysis as the regression model may be biased if important explanatory variables are omitted. Similarly Dollar & Levin (2006) added to calculations a vector of other determinants of aid

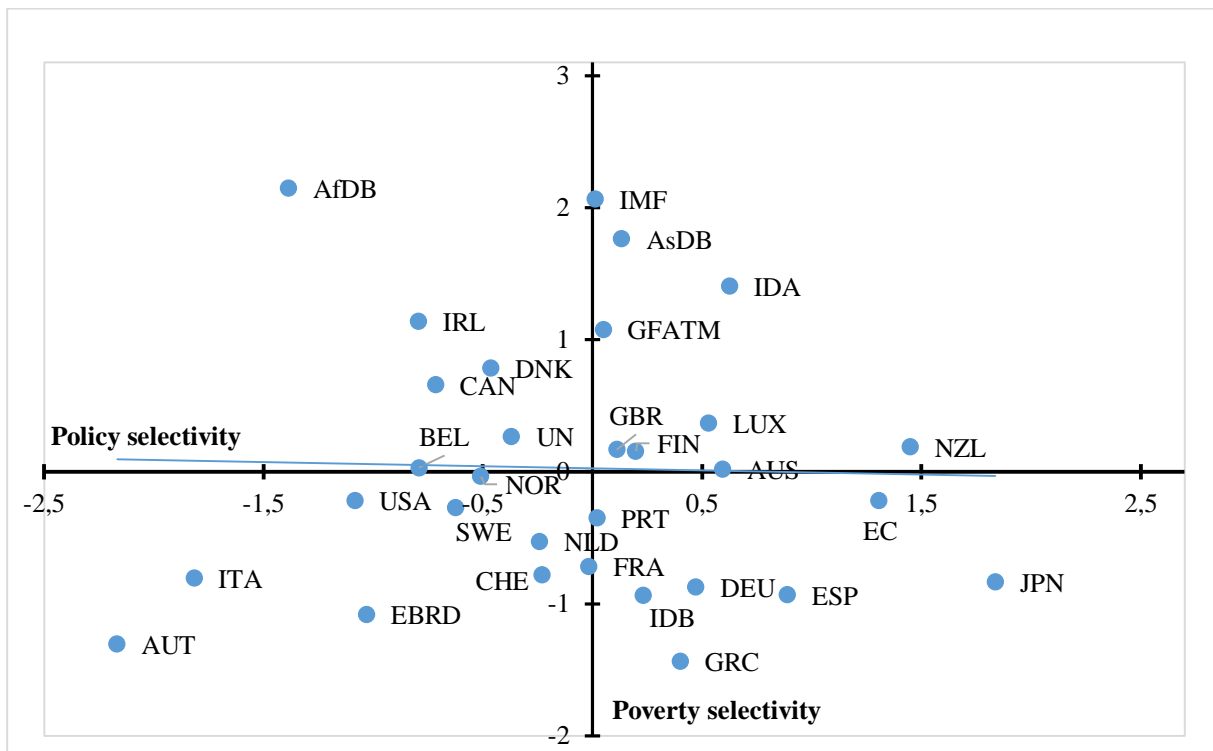
allocation – historical ties, donor’s export and distance. Also Clist (2015) adjusts for more factors and comments on KRE, “However, the approach of KRE appears flawed as they control for only some aspects of allocation but leave out important confounding factors such as historical, linguistic and commercial links between recipient and donor” (Clist 2015, 809).

Do the different understanding of concepts and different techniques of calculation project into results? Figures 6, 7 and 8 show z-scores for donors and multilaterals, data for all three indexes are covered up.²¹ The figures are divided into quadrants which reveal donors’ selectivity performance. Donors in the upper right quadrant are both, policy and poverty selective and contrariwise, donors in the lower left quadrant are neither policy nor poverty selective. In the remaining two quadrants, countries are selective in only one aspect to the detriment of the other one. The figures reveal that the results differ, and the difference is greater for policy selectivity because of the disagreement which variables stand behind policy. When looking at Asian Development Bank (AsDB), this multilateral received very good rating in KRE - it is among the most selective agencies. EW also placed AsDB to the upper right quadrant but QuODA placed this multilateral on the left of the upper left quadrant with only average poverty selectivity and one of the worst policy selectivities. For bilaterals, Austria (AUT) shows significant difference in assessment as it was placed to the lower right quadrant in QuODA and KRE, however, EW considers Austria as one of the least selective countries.

The trend lines in the figure reveals that QuODA shows certain trade-off between poverty and policy selectivity, whereas KRE shows trend that countries are either selective in both ways or not selective at all.

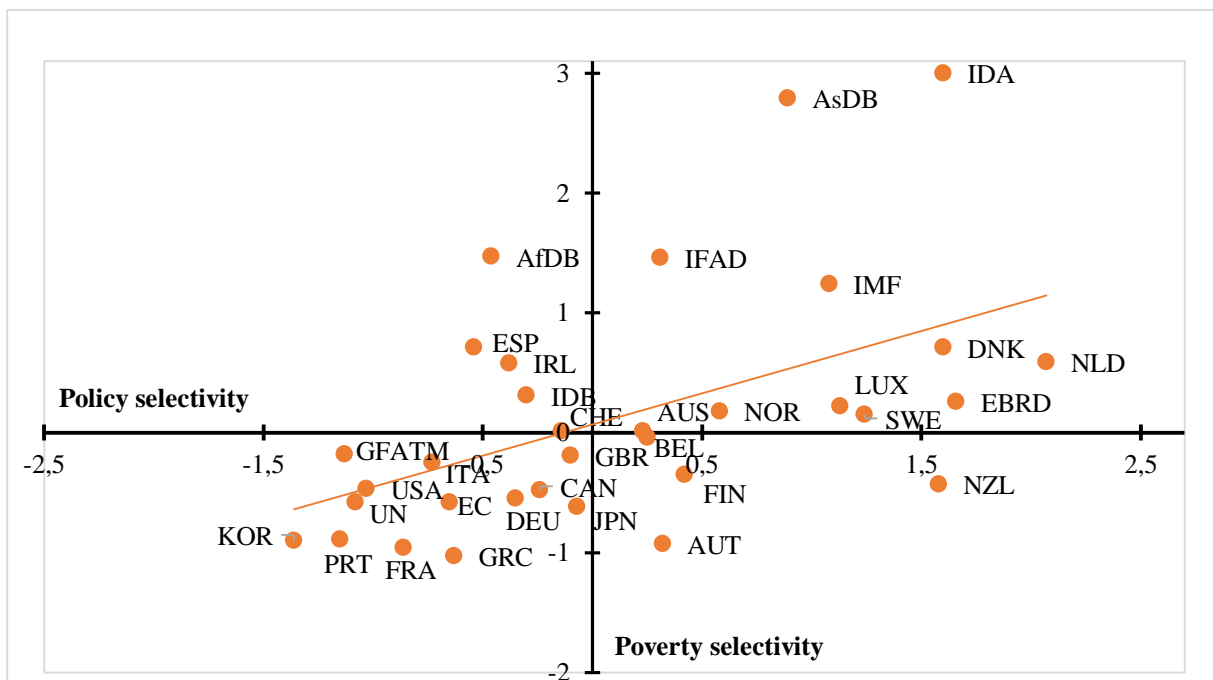
²¹ For the comparability, data in EW were transformed into z-scores and only common set of donor countries was included.

Figure 6. Selectivity performance (EW)



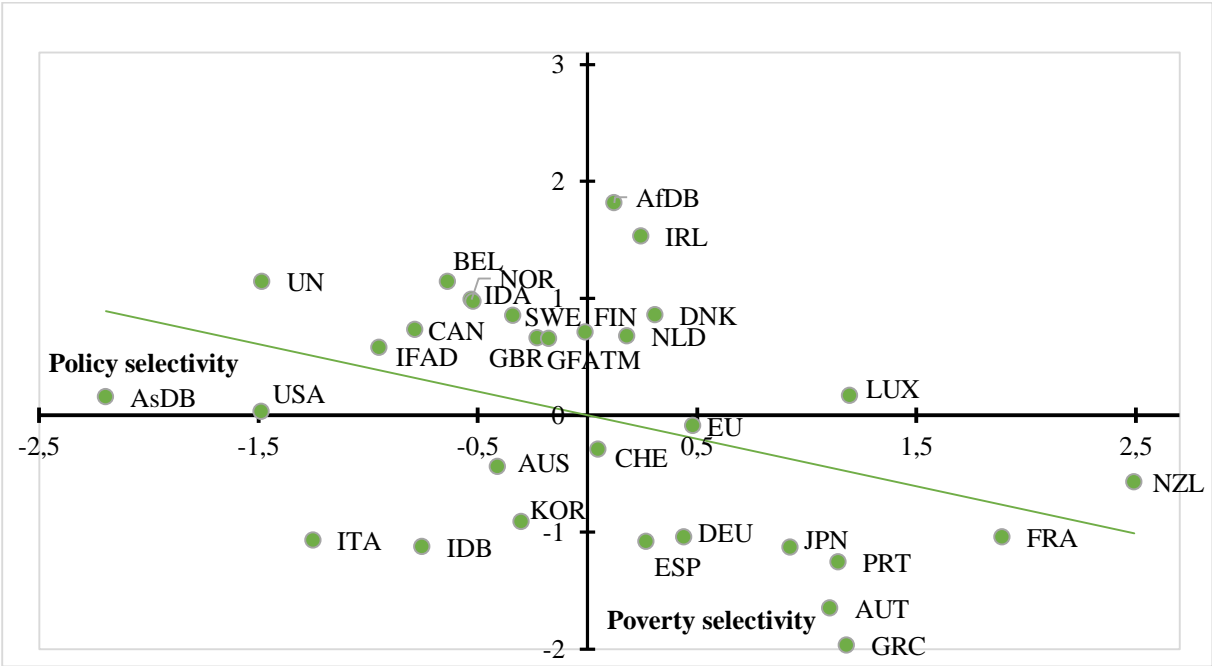
Source: author

Figure 7. Selectivity performance (KRE)



Source: author

Figure 8. Selectivity performance (QuODA)



Source: author

In conclusion, KRE recognizes only three factors which stand behind donors’ allocation of aid (population size, level of policy, level of poverty). Indeed, this issue is disputable but it seems that there are more of them as the literature suggests. Using the CPIA for measuring policy selectivity may bias the results as CPIA is produced by the World Bank which is one of assessed donors in the sample.

EW’s methodology can seem tough for donors who do not send aid to low income countries for objective reasons, e.g. donors who have some experience with economic transformation and direct their aid to post-soviet states. The criterion of poverty selectivity also implies that directing aid to other than low-income countries is not desirable. QuODA came with better solution - each aid flow is weighted according to its destination – therefore aid to all recipients is counted, but aid to poorer recipients has more weight.

3.3. Specialization/Fragmentation

There is a broad agreement that greater specialization contributes to aid effectiveness through reducing administrative burden imposed on partner countries (Knack & Rahman 2006, Roodman 2006, Anderson 2012, Bigsten & Tengstam 2015). At the same time, the prevailing view is that aid is currently fragmented too much. Williamson (2010) summarizes that “The overall picture of aid is one that is fragmented along many dimensions, forfeiting

the gains from specialization and possibly creating confusion between both donor and recipient countries” (Williamson 2010, 11).

Fragmentation is also a subject of attention at a political level. The PD (OECD 2005) states that fragmentation impairs aid effectiveness, and it appeals for more pragmatic division of labour and burden sharing with the aim to reduce transaction costs and increase complementarity. The AAA (OECD 2008) and the Busan Partnership (OECD 2011) comment the issue in the same spirit and the latter stresses the issue of aid channels proliferation.

On the contrary, only a little fraction of studies about donor proliferation are dedicated to the opposite side of spectrum – complete specialization. Some studies (Frot & Santiso 2008, Anderson 2012) warn that 100% specialization is not optimal either, though there is no consensus on how much fragmentation is optimal. Han & Koenig (2015) acknowledge the administrative burden connected with fragmentation but they point out that multiple donors bring more point of views on the issue, which can lead to finding better solution, and that the relationship between aid effectiveness and fragmentation may be curvilinear.

All three analysed indexes contain measures of aid fragmentation/specialization, specifically two basic measures: country fragmentation and sector fragmentation which will be examined in detail.

EW measures concentration of aid by using Herfindahl-Hirschman index (HHI). First, shares of aid to individual recipient countries or sectors are squared, then, the squares of these values are summed.²² The output score varies from zero to one where zero means maximal fragmentations and 1 stands for maximal specialization. HHI is also used by KRE but the score is furthermore adjusted for donors’ aid volume and geographical or sectoral mandate. Residuals from this regression are the country performance indicators.

QuODA approaches the issue of specialisation by counting revealed comparative advantage (RCA). Final score is counted as a share of donor’s aid with RCA greater than one. Whether a particular donor has the RCA in a given recipient country is calculated by comparing the share of the donor’s aid in the total volume of aid flowing into the recipient,

²² HHI is widely used to measure market concentration. It is counted according to formula: $\sum_r \left[\left(\frac{aid_{dr}}{aid_d} \right)^2 \right]$. For sector fragmentation: $\sum_r \left[\left(\frac{aid_{ds}}{aid_d} \right)^2 \right]$.

with the share of the donor’s aid in the total global aid. The sector specialization is counted similarly but instead of a particular recipient, the RCA is counted for a particular sector.²³

Which formula better reflects the theory? HHI captures fragmentation very imperfectly and we assume that the use of this formula misses the mark. Let us take two donors that disburse the same amount of aid. Donor A provides aid to 5 countries, each of them receive 20% of the total volume. Donor B disburse aid to 11 countries – one country receives 82% of the total volume and the remaining 10 recipients each receive 2%. The HHI score will be surprisingly better for Donor B (Donor A scores 0.2 and Donor B 0.644), even if this donor fragments aid more and probably imposes much higher administrative burden on recipient countries by sending relatively small amounts of aid to many countries. In addition, EW incorporates only relative shares of donors’ aid into the calculation without considering absolute volumes of aid, which implies that all donors (no matter if large or small) are treated the same. As a consequence, large donors are gratuitously penalized for fragmenting their aid more than small donors.

As for the sector fragmentation, HHI is not sensitive to number of recipient countries. Let us take two donors from the example above and suppose that both of them contribute to two similar sectors the same amount of money (50%). Their HHI score will be the same even though the aid for one sector can be disbursed among many countries. Donor B could support both sectors in each of 11 recipients and this excessive fragmentation would not be underpinned.

By contrast, QuODA deals with the issue by crediting only those shares of aid, which are significant (their $RCA > 1$). At the same time, their formula addresses the problem of aid reallocation among donor darlings and donor orphans and favours aid allocation to the latter. Let us consider two identical donors that both disburse 10 units of aid – one to a donor orphan and one to a donor darling. In this case, the variables in the formula for RCA ²⁴ are all identical for both donors with the exception of the total aid in recipient countries. It is clear

²³ QuODA - Specialization by recipient country is counted according to formula: $\sum_r \left[\left(\frac{CPA_{d,r,RCA>1}}{CPA_d} \right) \right]$ where $RCA = \frac{CPA_{d,r}}{CPA_d / CPA_w}$. Sector specialization by sector is counted similarly: $\sum_s \left[\left(\frac{CPA_{d,s,RCA>1}}{CPA_d} \right) \right]$ where $RCA = \frac{CPA_{d,s}}{CPA_d / CPA_w}$. CPA – country programable aid, d – donor, r – recipient, s – sector, w – world, RCA – revealed comparative advantage.

²⁴ $RCA = \frac{CPA_{d,r}}{CPA_d / CPA_w}$, RCA – revealed comparative advantage, CPA – country programmable aid, d – donor, r – recipient, w – world

therefore that the same share of aid will have higher RCA score when flowing to a donor orphan than to a donor darling (while HHI is completely insensitive to this issue). In conclusion, the formula used by QuODA better reflects the theoretical base about aid fragmentation.

The approach of KRE, QuODA and EW towards complete specialization is another sticking point, since a donor with 100% specialization gets the best score. Therefore, the authors build on two presumptions: either fragmentation is inversely proportional to aid effectiveness to the infinity or all donors are situated far on the left from the optimum. EW is the example of the latter, as it states, “The caveat is that complete specialization by country or sector is not necessarily optimal either... In practice, however, most of our observations are at a high level of fragmentation that plausibly correspond to suboptimal behaviour” (Easterly & Williamson, 2011, 1935).

KRE and QuODA add supplementary indicators to better capture administrative costs imposed on recipients by exceeding fragmentation. KRE adds Number of recipients aided and Average number of sectors aided per recipient country receiving aid, both indicators are adjusted for donor’s aid volume and geographic or sectoral mandate. The index also assesses the Average size of projects.²⁵ Therefore, KRE favours donors who specialize on a little number of recipients and sectors, and it holds that larger donors can more fragment aid to get the same score. Next, the implementation of large projects is favoured by the index.

QuODA also adds three indicators whose goal is to measure administrative burden imposed on recipient countries. These are Significance of aid relationships, Fragmentation across donor agencies and Median project size. First indicator is connected to administrative burden in the way that administrative costs are “inversely proportional to the concentration of aid” (Birdsall & Kharas 2010, 60). The HHI for each country is weighted by the donor’s total aid²⁶. Second indicator punish donors for channelling aid through too many agencies, and by doing this increasing the number of donor-partner relationships. The HHI²⁷ is equal to one if all donor’s aid is delivered through one agency and the number decreases with increasing number of agencies. The last indicator stands on the assumption that every project imposes some fixed costs which are relatively higher in case of a small project. It is calculated as the

²⁵ Size of projects i (logged) is regressed on full set of donor, recipient and sector dummies (which are no more specified in the paper) and on the log of total aid. Donor dummy coefficient estimates are then regressed on donor’s aid commitments and the incurred residuals are taken as donors’ performance indicators.

²⁶ This indicator measures the donor’s marginal contribution to administrative costs of recipients. It implies, among others, that small donors should focus on small recipient countries (to get a good score).

²⁷ The HHI is calculated as a sum of squared shares of total aid (gross CPA) for each agency.

log of a donor's median project size. According to QuODA, an optimal donor would have significant relationships with its recipients, it would deliver aid through few agencies and it would implement rather large projects.

We find that both indexes have some inadequacies in measuring project size. Concerning QuODA, the current, prevailing approach is that larger projects should be preferred to reduce excessive project proliferation and subsequently higher administrative burden imposed on recipients. However, the right size of an aid project is unclear as the relationship between project size and aid effectiveness is apparently not the larger the better till the infinity. QuODA stands on the presumption that the size of all projects is below optimal because the larger project, the better. In addition, the index also do not take into account that donors are a heterogeneous group so that it favours larger donors that will have more likely larger median size of projects. KRE takes into account the heterogeneity of donors but it poorly specifies the variables the average project size is regressed on.

In conclusion, EW might not choose the most suitable approach because according to the formula, giving a little proportion of aid to a new recipient country, which is the least desired type of fragmentation, changes the donor's score only slightly. In view of the fact that shares of aid enter the calculation, the size of donor is not considered and large donors are penalized for providing aid to more recipient countries than small donors. KRE takes into account donors' volume of aid and limited mandates, but the calculations and entering variables are not always transparently explained. QuODA may give too much focus on reducing administrative cost and not on achieving the desired results but its approach of counting revealed comparative advantage seems the best of all three.

3.4. Transparency

As for the political pledges, transparency run through the whole text in the Paris Agenda. The AAA states that "Transparency and accountability are essential elements for development results. They lie at the heart of the Paris Declaration, in which we agreed that countries and donors would become more accountable to each other and to their citizens" (OECD, 2008).

On the basis of the PD and the AAA, International Aid Transparency Initiative (IATI) was launched. It is a voluntary initiative with the aim to promote transparency. Members of IATI pledge to publish data in a unified format which is set by IATI Standards.

The importance of transparency comes out of the fact that it is a key element for other good practices of aid effectiveness, e.g. coordination and better aid allocation. A study by Collin et al. (2009) tries to quantify costs and benefits of increased transparency. Their finding is that benefits of increased transparency exceed total costs ²⁸to the extent that costs are likely to be recovered in about one year, even though some benefits were not counted in because of difficulties to quantify them.

Only two indexes (EW and QuODA) recognize transparency as an important component they should incorporate into their measures. The authors of KRE justify omitting transparency by problems with measuring this concept, and surprisingly also by claiming that transparency is not so important to be incorporated into the index, “Transparency and overhead costs...have not been central to the aid effectiveness debates, they are largely absent from the Paris Declaration and they are hard to measure accurately” (Knack et al. 2011, 1913). At this point, it is worth noting completely opposite view of EW (Easterly & Williamson, 2011, 1932) which refers to agency transparency as possibly the most important component of aid effectiveness and calls it “the latest buzzword among the aid community.”

EW creates two indexes of transparency. OECD Reporting Transparency Index assesses donors based on their reporting to OECD/DAC statistics and Creditor Reporting System (CRS) database. The authors monitor reporting to particular tables concerning aid disbursements, commitments and official flows.²⁹ Transparency Overhead Costs Index assesses transparency of agencies as regards their operating cost. The authors conduct a two-step survey on data availability for four categories – permanent international staff, administrative expenses, salaries and benefits, total development assistance disbursed. Firstly, websites of agencies are consulted for the pieces of information, secondly, agencies are directly contacted via e-mail with a request for data. If the required data are published online, the agency gets 1 point for each category. If the data are provided only via e-mail, the agency gets 0.5.

In QuODA, we identified six proxies measuring transparency. Out of these six, two indicators are connected with International Aid Transparency Initiative (IATI) and the remaining four assess donor’s transparency through reporting to Creditor reporting system (CRS) and to AidData. Indicators concerning IATI are the only binary variables in QuODA

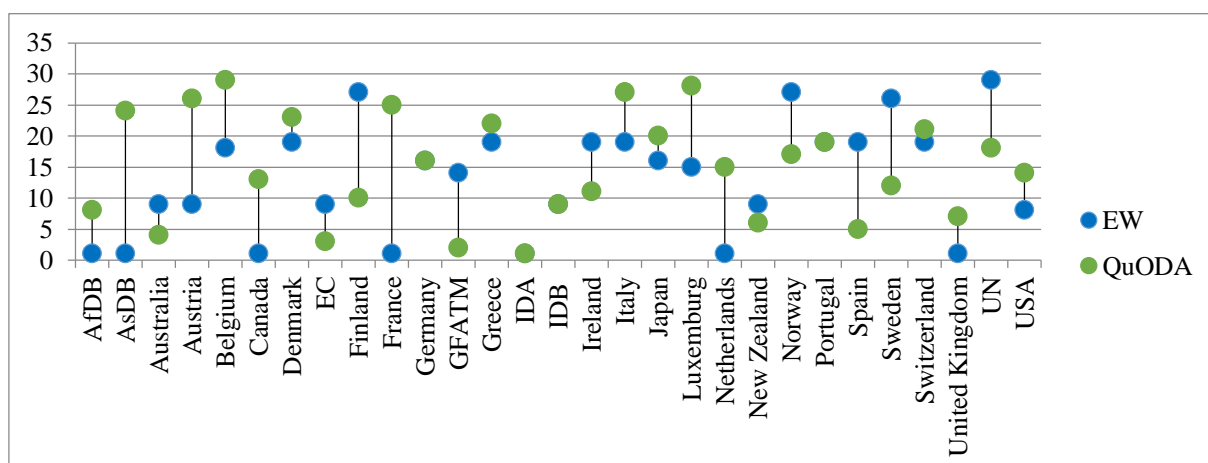
²⁸ By costs, the authors mean mainly administrative costs. Benefits can be multiple: aid predictability, accountability, coordination, facilitating research, improving allocation etc.

²⁹ Five specific tables for bilaterals and three tables for multilaterals are reviewed.

with possible answers yes or no. *Signatory of IATI* reflects donors' commitment, not really a practice. *Implementation of IATI Data reporting standards* was added in the second edition of QuODA. The remaining four indicators are following: *Reporting of project title and descriptions*, *Detail of project description*, *Reporting of aid delivery channel*, *Completeness of project-level commitment data*. In sum, the authors analyse the completeness of reported data. They look at the share of filled key fields, number of signs in the description, share of aid with sufficiently reported aid delivery channel and share of aid that was accounted for in project-level reporting.

The description above implies that QuODA and EW understand the concept of transparency differently. QuODA focuses on project transparency (how much aid is being given to whom, for what project, through which delivery agency and when) whereas EW focuses on operational costs of agencies and reporting to OECD/DAC. On the whole, EW monitors if data are reported and QuODA monitors how they are reported.

Figure 9. Gaps between transparency ranks for QuODA and EW



Source: author

To which extent do the scores differ for QuODA and EW? Very low degree of correlation (Spearman coefficient 0.21) implies that the results differ significantly. Also Figure 9 shows that the gap between rank in QuODA and rank in EW is huge, look, for example, at Asian Development Bank (AsDB) or France. This is caused by focusing at different aspects of transparency as well as by differing methodology; variables in EW can get only specific values and as a consequence, more donors can get the same rank.

The authors of both studies directly contacted agencies to collect some data and they both came across the same problem – unwillingness of some respondents to participate. EW

states, that they would ideally ask for more data, but based on previous experience, they decided for less ambitious version. Authors of QuODA launched two surveys, one concerning project transparency and one concerning evaluation. Finally, they did not incorporate this survey into QuODA because some agencies were reluctant to participate and because some concepts were understood differently so the data were not comparable among agencies.

Attempts to measure transparency face many problems. As summarized by Ghosh & Kharas (2011), “It is not straightforward to measure transparency, partly because used norms and standards are still not universally accepted and partly because transparency is an elusive and shifting concept that resists an easy definition.” EW and QuODA both admit the imperfection of measures and add that data may not be perfectly comparable across agencies.

Both indexes offered an interesting insight into the area of transparency measuring. EW monitors if data are reported, not their quality. The variables can take only limited number of values, which hinders donor grading – donors that report will have the same score no matter the detail of their reporting. On the other hand, their direct contact of agencies with requesting certain data is an original approach of measuring transparency directly. QuODA monitors not only if data are available, but also if they are timely and in international format, or how detailed the descriptions are. It is an interesting insight into how to measure transparency by proxies, though the proxies have some imperfections (e.g. number of signs in a description field may be misleading in assessing the detail of description).

CONCLUSION

The thesis compared three indexes measuring quality of aid by donors. We examined in detail how the indexes are constructed and how they measure donors' performance in particular themes.

Our analysis shows that some indexes have inadequacies in their construction. KRE and also EW to a lesser extent are insufficiently transparent considering chosen methods of construction and their possible effects on final results. KRE has a weak structure of index, which is not balanced, and indicators in the Specialization dimension do not fully match this dimension. On the other hand, QuODA is technically well constructed and it matches best the commonly accepted standards for index construction.

The content of indexes reveals that there is an agreement about what the best practices are but not about how to measure them. Number of areas included in the index depend on how broadly the concept is taken into consideration. Some topics, especially transparency here, are difficult to quantify and there are few previous attempts to capture them in an index. This is partly the reason why transparency ranks in EW and in QuODA show only moderate correlation.

We find that EW's indicators in the Specialization dimension do not really reflect the theoretical base as Herfindahl-Hirschman index is rather a measure of market concentration and therefore, excessive fragmentation of aid into little pieces does not project greatly into the score. Treating the technical assistance as an ineffective channel is also questionable.

We find that KRE is insufficiently transparent in defining what the indicators are regressed on and in unclear formulas of calculations. The explanatory variables incorporated into regressions should be a subject of further debates to reach a better model where donors' performance is adjusted to factors which are out of donors' control.

QuODA transparently explains the utilization of each indicator. The focus may be too much on reducing costs (efficiency) rather than on the fact if foreign aid leads to intended results (effectiveness). Nevertheless, QuODA seems to use the best methods of calculation of all three indexes and match the theoretical base most.

KRE and QuODA show strong correlation of donors' ranks because they both borrow a great deal of indicators from the Paris Declaration and many of their indicators overlap. EW is rather distant as it does not follow the political declarations so closely and its concept is rather narrow. Considerable amount of differing ratings may be caused by using a different

normalization method. As the analysis from the chapter 3.2.2 revealed, changing the normalisation method from percentile ranks to z-scores caused significant shifts in ranking.

Despite some inadequacies, the indexes of aid quality are indeed important as they can draw attention to the issue and encourage donors to adopt the best practices. Nevertheless, their output should be interpreted carefully as they may underestimate or overestimate some donors.

LIST OF REFERENCES

- Alesina, A., Weder, B. 2002. Do Corrupt Governments Receive Less Foreign Aid? *National Bureau of Economic Research, Working Paper 7108*.
- Anderson, E., Clist, P. 2011. Measures for Measures: Evaluating Judgements of Donor Allocative Performance. *DEV Working Paper 35*. University of East Anglia.
- Anderson, E. 2012. Aid Fragmentation and Donor Transaction Costs. *Economics Letters* 117, 799-802.
- Annen, K., Kosempel, S. 2009. Foreign Aid, Donor Fragmentation, and Economic Growth. *The B.E. Journal of Macroeconomics* 9 (1), article 33.
- Askarov, Z., Doucouliagos, H. 2013. Does Aid Improve Democracy and Governance? A Meta-regression Analysis. *Public Choice* 157 (3-4), 601-628.
- Bigsten, A., Tengstam, S. 2015. International Coordination and the Effectiveness of Aid. *World Development* 69, 75-85.
- Birdsall, N., Kharas, H. 2010. *Quality of Official Development Assistance Assessment*. Washington DC: Center for Global Development.
- Birdsall, N., Kharas, H., Perakis, R. 2011. *Is Aid Quality Improving? The Quality of Official Development Assistance Assessment 2009*. Washington, DC: Center for Global Development.
- Birdsall, N., Kharas, H. 2014. *The Quality of Official Development Assistance (QuODA): Third Edition*. Washington, DC: Center for Global Development.
- Boone, P. 1996. Politics and the Effectiveness of Foreign Aid. *European Economic Review*, 40, 289–329.

Bourguignon, F., Sundberg, M. 2007. Aid Effectiveness – Opening the Black Box. World Bank paper presented at the American Economics Association annual meeting January 5-7 2007.

Bourguignon, F., Platteau, J. 2015. The Hard Challenge of Aid Coordination. *World Development* 69, 86-97.

Burnside, C., & Dollar, D. 2000. Aid, Policies, and Growth. *American Economic Review*, 90, 847–868.

Clist, P. 2011. 25 Years of Aid Allocation Practice: Whither Selectivity? *World Development* 39 (10), 1724-1734.

Clist, P. 2015. Do Performance Measures of Donors' Aid Allocation Underperform? *The World Economy* 38 (5), 805-824.

Cogneau, D., Naudet, J. 2007. Who Deserves Aid? Equality of Opportunity, International Aid, and Poverty Reduction. *World Development* 35 (1), 104-120.

Collier, P., Dollar, D. 2001. Can the World Cut Poverty in Half? How Policy Reform and Effective Aid Can Meet International Development Goals. *World Development* 29 (11), 1787-1802.

Collier, P., Dollar, D. 2002. Aid Allocation and Poverty Reduction. *European Economic Review* 46 (2002), 1475-1500.

Collin, M., Zubairi, A., Nielson, D., Barder, O. 2009. *The Costs & Benefits of Aid Transparency*. Development initiatives.

Djankov, S., Montalvo, J. G., Reynal-Querol, M. 2009. Aid with Multiple Personalities. *Journal of Comparative Economics* 37 (2), 217-219.

Dollar, D., Levin, V. 2006. The Increasing Selectivity of Foreign Aid, 1984-2003. *World Development* 34 (12), 2034-2046.

Easterly, W., Pfutze, T. 2008. Where Does the Money Go? Best and Worst Practices in Foreign Aid. *Journal of Economic Perspectives* 22 (2), 29-52.

Easterly, W., Williamson, C. 2011. Rhetoric versus Reality: The Best and Worst of Aid Agency Practices. *World Development* 39 (11), 1930-49.

Freudenberg, M. 2003. Composite Indicators of Country Performance: A Critical Assessment. *OECD Science, Technology and Industry Working Papers*, 2003/16. Paris: OECD.

Frot, E., Santiso, J. 2008. Development Aid and Portfolio Funds: Trends, Volatility and Fragmentation. *OECD Development Centre, Working Papers* 275. OECD Publishing.

Frot, E., Santiso, J. 2010. Crushed Aid: Fragmentation in Sectoral Aid. *OECD Development centre, working paper* 284. OECD Publishing.

Ghosh, A., Kharas, H. 2011. The Money Trail: Ranking Donor Transparency in Foreign Aid. *World Development* 39 (11), 1918-1929.

Glennie, J., Sumner, A. 2014. The \$ 138.5 Billion Question: When Does Foreign Aid Work (and When Doesn't It)?. *CGD Policy Paper* 049. Washington DC: CGD.

Han, L., Koenig-Archibugi, M. 2015. Aid Fragmentation or Aid Pluralism? The Effect of Multiple Donors on Child Survival in Developing Countries, 1990-2010. *World Development* 76, 344-358.

In'airat, M. 2014. Aid Allocation, Selectivity, and the Quality of Governance. *Journal of Economics, Finance and Administrative Science* 19 (2014), 63-68.

Jacobs, R., Smith, P., Goddard, M. 2004. Measuring Performance: An Examination of Composite Performance Indicators. *CHE Technical Paper Series* 29. York: The University of York.

Joint Research Centre. 2002. *State-of-the-art Report on Current Methodologies and Practices for Composite Indicator Development*. Institute for the Protection and the Security of the Citizen. Italy: European Commission.

Kimura, H., Sawada, Y., Mori, Y. 2007. Aid proliferation and economic growth: A Crosscountry Analysis. *RIETI discussion paper series 07-E-044*. Research Institute of Economy, Trade and Industry.

Knack, S., Rahman, A. 2006. Donor Fragmentation and Bureaucratic Quality in Aid Recipients. *Journal of Development Economics* 83 (2007), 176-197.

Knack, S., Rogers, F. H., Eubank, N. 2011. Aid Quality and Donor rankings. *World Development* 39 (11), 1907-1917.

Kosack, S. 2003. Effective Aid: How Democracy Allows Development Aid to Improve the Quality of Life. *World Development* 31 (1), 1-22.

McGillivray, M. 1989. The Allocation of Aid among Developing Countries: A Multi-Donor Analysis Using a Per Capita Aid Index. *World Development* 17 (4), 561-568.

McGillivray, M. 2003. Descriptive and Prescriptive Analyses of Aid Allocation: Approaches, Issues and Consequences. *Discussion Paper No. 2003/21*. WIDER (World Institute for Development Economics Research).

Mosley, P., Hudson, J., & Horrel, S. 1987. Aid, the Public Sector and the Market in Less Developed Countries. *Economic Journal*, 97, 616-641.

Nunnenkamp, P., Thiele, R. 2006. Targeting Aid to the Needy and Deserving: Nothing But Promises? *The World Economy* 29 (9), 1177-1201.

OECD. 2003. Rome Declaration on Harmonisation. Rome, OECD.

OECD. 2005. Paris Declaration on Aid Effectiveness. Paris, OECD.

OECD 2008. Accra High Level Forum on Aid Effectiveness. Accra, OECD.

OECD. 2008. Handbook on constructing composite indicators: Methodology and user guide. Paris, OECD.

OECD 2011. Busan Partnership for Effective Development Co-operation. Busan, OECD.

Rajan, R. G., Subramanian, A. 2008. Aid and Growth: What Does the Cross-Country Evidence Really Show?. *The Review of Economics and Statistics* 90 (4), 643-665.

Radelet, S. 2006. A Primer on Foreign Aid. *CGD Working Paper* 92. Washington DC: CGD.

Rao, J. 1997. Ranking Foreign Donors: An Index Combining the Scale and Equity of Aid Giving. *World Development* 25 (6), 947-961.

Ravallion, M. 2010. Mashup Indices of Development. *Policy research working paper* 5432. Washington, DC: The World Bank.

Roodman, D. 2006. Aid Project Proliferation and Absorptive Capacity. *CGD, Working Paper* 75.

Roodman, D. 2006b. An Index of Donor Performance. *CGD, Working paper* 67.

Roodman, D. 2011. Composite indices. *Journal of Economic Inequality* 9 (3), 483-484.

Stern, E., Altinger, L., Feinstein, O., Marañón, M., Schultz, N., Nielsen, N. S. 2008. *The Paris Declaration, Aid Effectiveness and Development Effectiveness: Evaluation of the Paris Declaration*. Ministry of Foreign Affairs of Denmark, Copenhagen.

White, H., McGillivray, M. 1995. How Well is Aid Allocated? Descriptive Measures of Aid Allocation: A Survey of Methodology and Results. *Development and Change* 26, 163-183.

Williamson, C. Fixing Failed Foreign Aid: Can Agency Practices Improve? *AidData conference*. Oxford, UK.

LIST OF APPENDICES

Appendix 1: Structure of the indexes

EW				
Transparency	Overhead costs	Specialization	Selectivity	Ineffective channels
OECD reporting transparency index	admin costs/ODA	Sector fragmentation	Share to Noncorrupt	Share of tied aid
Transparency overhead costs index	salaries and benefits/ODA	Country fragmentation	Share to free (democratic)	Share of food aid
	total ODA per employee		Share to poor	Share of technical assistance

KRE			
Selectivity	Alignment	Harmonization	Specialization
Policy selectivity	Untied aid	Use of program-based approaches	Contribution to multilaterals
Poverty selectivity	Aid predictability	Coordinated missions	Administrative costs
	Use of PFM systems	Coordinated country analytic reports	Average project size
	Use of procurement systems		Geographic concentration
	Use of PIUs		Number of recipients
	Coordinated technical cooperation		Sectoral concentration
			Average number of sectors per recipient

QuODA (2nd and 3rd edition)			
Maximizing efficiency	Fostering institutions	Reducing Burden	Transparency and Learning
Share of allocation to poor countries	Share of aid to recipients' top development priorities	Significance of aid relationships	Signatory of IATI
Share of allocation to well-governed countries	Avoidance of PIUs	Fragmentation across donor agencies	Implementation of IATI data reporting standards
Low administrative unit costs	Share of aid recorded in recipients budgets	Median project size	Recording of project title and description
High country programmable aid share	Share of aid to partners with good operational strategies	Contribution to multilaterals	Detail of project description
Focus/specialization by recipient country	Use of recipient country systems	Coordinated missions	Reporting of aid delivery channel
Focus/specialization by sector	Share of scheduled aid recorded as received by recipients	Use of programmatic aid	Quality of main agency evaluation policy
Support of selected global public good facilities	Coordination of technical cooperation	Coordinated analytical work	Completeness of project-level commitment data
Share of untied aid	Coverage of forward spending plans/Aid predictability		Aid to partners with good M&E frameworks