

FRAGILE

GERMAN DEVELOPMENT COOPERATION IN FRAGILE CONTEXTS

2019



DEval

GERMAN
INSTITUTE FOR
DEVELOPMENT
EVALUATION

State fragility is an important challenge for development cooperation. Challenges to the state's monopoly on the legitimate use of force, a limited ability to make and implement binding rules and to deliver basic services throughout a state's territory, as well as lack of belief in the rightfulness of a state's authority among the population, put at risk the efforts of development cooperation. At the same time, more than half of worldwide Official Development Assistance (ODA) goes to states affected by fragility. Consequently, a key policy question is how development cooperation can operate successfully in fragile contexts.

This evaluation provides guidance for policymakers to adapt development cooperation to fragile contexts. It reviews definitions of fragility to assist them in choosing suitable indices for decision-making. An allocation analysis then critically examines whether German ODA commitments are aligned to strategic decisions and highlights where further action might be required. Based on a synthesis of evaluation reports, this report furthermore analyses the quality of evaluations in fragile contexts and offers ideas for improvements. Finally, an examination of project success across different dimensions of fragility provides evidence in respect of factors that relate to successful development cooperation in fragile contexts.

GERMAN DEVELOPMENT COOPERATION IN FRAGILE CONTEXTS

2019

Imprint

Published by

German Institute for Development
Evaluation (DEval)
Fritz-Schäffer-Straße 26
53113 Bonn, Germany
Phone: +49 (0)228 33 69 07-0
E-Mail: info@DEval.org
www.DEval.org

Authors

Dr Thomas Wencker
Ida Verspohl

Responsible

Dr Sven Harten

Design

MedienMélange:Kommunikation!, Hamburg
www.medienmelange.de

Editing

Martin Rickerd OBE MVO

Photo credits

Benjamin Nelan on Pixabay

Bibliographical reference

Wencker, T. and I. Verspohl (2019), German
Development Cooperation in Fragile Contexts,
German Institute for Development Evaluation
(DEval), Bonn.

Printing

Bonifatius, Paderborn



© German Institute for Development
Evaluation (DEval), 2019

ISBN 978-3-96126-091-1 (printed edition)
ISBN 978-3-96126-092-8 (PDF)

The German Institute for Development Evaluation (DEval) is mandated by the German Federal Ministry for Economic Cooperation and Development (BMZ) to independently analyse and assess German development interventions.

Evaluation reports contribute to the transparency of development results and provide policymakers with evidence and lessons learned, based on which they can shape and improve their development policies.

This report can be downloaded as a PDF file from the DEval website:

www.deval.org/en/evaluation-reports.html

Requests for printed copies of this report should be sent to:

info@DEval.org

BMZ's response to this evaluation is available at <http://www.bmz.de/en/ministry/evaluation/Evaluation/evaluierungsberichte-stellungnahmen/index.html>

ACKNOWLEDGEMENTS

We extend special thanks to the reference group representing BMZ, GIZ, and KfW Entwicklungsbank and the Peer Reviewers Alexander de Juan, Philipp Rotmann, Gerald Leppert, and Sebastian Schneider for their highly valuable input during all phases of this evaluation. Advice given by Manuel Bosancianu and Jens Eger has been very helpful in developing our methodological approach. The use of text-mining in the evaluation synthesis would have been impossible without the support of Andreas Niekler and Christian Kahmann. Special thanks also go to Mascha Rauschenbach, who greatly contributed to the statistical analyses of evaluation quality. Sylvia Vogt seamlessly administered the project with great competence. Roxana Bitá and Ida Seljeskog efficiently managed the layout process. Silvia Richter and Martin Rickerd provided excellent support in copy-editing.

We are deeply indebted to the interns and student assistants. Ann-Kristin Becker has greatly contributed through her precise analysis of allocation patterns. Salome Ecker supported us through her comprehensive and lucid presentation of decision-making processes. Julia Nepicks provided highly valuable assistance in systematically comparing approaches to defining and measuring fragility. Lea Smidt contributed through a comprehensive and precise assessment of strategy development. Antje Roth was crucially involved in developing the text-mining approach. Eva Bosse, Laura Pargen, Corinna Thiesen and Maria Petrova provided excellent research assistance. Without their contributions, this study would not have been possible.

EXECUTIVE SUMMARY

Fragility – a central challenge for development cooperation

State fragility is an important challenge for development cooperation. The 2030 Agenda for Sustainable Development reflects the challenge of development cooperation in fragile contexts: SDG 16 sets out the vision to “[p]romote peaceful and inclusive societies [...], provide access to justice for all and build effective, accountable and inclusive institutions” (UN, 2015, p.14). Beyond the immediate added value of functioning regulatory institutions, challenges to the state’s monopoly on the legitimate use of force, a limited ability to make and implement binding rules and to deliver basic services throughout a state’s territory, as well as lack of belief in the rightfulness of a state’s authority among the population, put at risk the efforts of development cooperation.

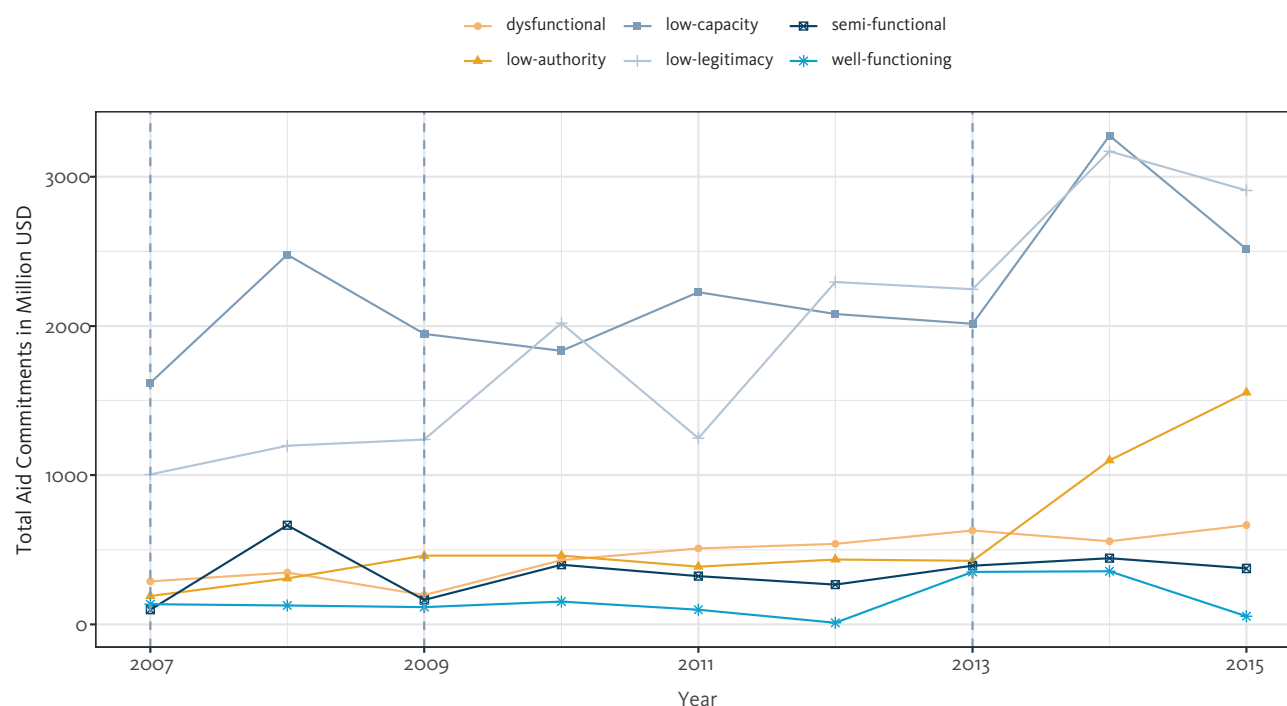
More than half of worldwide Official Development Assistance (ODA) goes to states affected by some or all of the mentioned aspects of fragility. German bilateral ODA commitments are no exception. Figure 1 shows committed grants and loans by type of fragility.¹ The largest share of German development cooperation goes to countries affected by some type of fragility – specifically low-capacity states, i.e. states with a low administrative capacity and limited supply of basic public services. It is for that reason that DEval is dedicating a Thematic Report on this topic which will compile the evidence of DEval evaluations and external contributions.

For development cooperation to be successful despite the challenging conditions in fragile contexts, the following four conditions should be satisfied:

1. A precise understanding of state fragility to align development policies to the local context
2. A political process that designs, possibly adapts, and implements strategies of development cooperation in fragile contexts effectively
3. Suitable procedures for learning what works in fragile contexts and what does not
4. Knowledge of the evidence with regard to which instruments of development cooperation work in fragile contexts.

The four self-contained chapters in this evaluation deal with the above-mentioned points. The following presents a brief summary.

¹ Grävingholt et al. (2018) distinguish six constellations of state fragility based on the state’s monopoly on the legitimate use of force (authority), the ability to provide basic services (capacity) and the recognition of state leadership (legitimacy). Dysfunctional, semi-functional and well-functioning states emerge from low, middle and high authority, capacity or legitimacy, respectively. States with low authority, capacity or legitimacy are characterized by low values in the respective dimension, while having average or high values in the other dimensions.

Figure 1 Bilateral ODA commitments (grants and loans) by type of fragility

Source: own figure based on data from OECD CRS and Grävingholt et al. (2018).

Subject of this evaluation

This evaluation provides guidance for policymakers in the German Federal Ministry for Economic Cooperation and Development (BMZ) and other donors to adapt development cooperation to the context of partner countries with the aim to make it more effective. Each chapter examines one of the four above-mentioned conditions.

Following an introduction, Chapter 2 reviews the definition and measurement, i.e. the conceptualization of state fragility. Resulting indices and typologies constitute an integral part of decision-making processes in the BMZ and the implementing agencies. This evaluation provides a basis for assessing the suitability of available indices for decision-making. Chapter 2 critically compares approaches to defining and measuring state fragility. This is essential practical information since concepts influence how policymakers perceive current affairs. Resultant typologies classify partner countries, thereby influencing decisions about the suitability of certain development policy instruments. Consequently, the choice of concepts may influence aid allocation. With different approaches resulting in different classifications of states thus having tangible consequences, Chapter 2 speaks to a core issue of strategic steering by the BMZ.

The third chapter investigates policies by analysing strategies of development cooperation and allocations. Since 2005, the BMZ has released five major strategies guiding development cooperation in fragile states. This evaluation describes continuity and change in Germany's and the BMZ's development strategy towards fragile states and critically examines whether German ODA commitments are aligned to these strategies. The allocation analysis concisely points out which strategic decisions were actually put into practice and highlights where further action might be required.

Chapter 4 analyses the quality of evaluations and thus deals with one of the pivotal procedures for learning what works in fragile contexts. The implementing agencies, KfW Development Bank (KfW) and the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH, have built up a large body of evidence on the success of development cooperation. This evidence spans all types of countries and context conditions – among them fragile settings in which a lack of security and poor infrastructure can offer

challenges in designing and conducting evaluations. Based on an earlier DEval evaluation on sustainability in German development cooperation, this evaluation analyses the quality of evaluations depending on the fragility of the context. The analysis identifies factors that impede evaluations in fragile contexts and offers ideas on how to increase the usefulness of evaluations.

The fifth chapter gives policymakers insights into which instruments of development cooperation are particularly suitable in fragile contexts. A synthesis of evaluation reports examines the correlation between context fragility and the evaluation's assessment of project success. The analysis differentiates dimensions of state fragility, sectors, and criteria to measure project success. This allows the identification, from the many possible influencing factors, of those that relate to successful development cooperation in fragile contexts.

Methods

The concept comparison and the strategy analysis are based on an evaluation of academic literature, official documents, publicly available socio-economic data, and interviews.

To delineate the structure and processes of the German bilateral development cooperation system, the evaluation team reviewed internal and public documents elaborating on the procedures for decision-making. The review placed particular emphasis on those steps in the process that relate to strategies on development cooperation in fragile contexts or fragility assessments. In order to tailor recommendations to the needs of the BMZ, GIZ, and KfW, they refer to the procedures as described in the BMZ's Joint Procedural Reform (GVR).

The analyses of evaluation quality and project success build on a synthesis of evaluation reports by the GIZ and KfW. This synthesis links characteristics of the evaluated project with data on the implementation context. To consider subnational variation of state fragility, this evaluation develops an approach to automatically retrieve geographical information from evaluation reports in order to locate projects. This data is then linked to conflict events to estimate the enforcement of the state's monopoly on the use of force at a subnational level. Moreover, the analysis of evaluation quality draws on evaluation guidelines from the GIZ and KfW and existing meta-evaluations.

The following sections summarize the four main chapters. Each section discusses the relevance, and presents the main results and recommendations.

Defining and measuring state fragility: Results and recommendations

Context-specific development cooperation requires an understanding of state fragility as a multidimensional phenomenon. Given an increased interest in the topic especially after 9/11, the number of approaches to measuring fragility has increased considerably. This resulted in more refined concepts of fragility that acknowledge its multidimensional nature. Rather than classifying states binarily or assessing fragility on a one-dimensional scale, more recent approaches measure fragility based on multiple constitutive attributes and derive typologies.

The increasing availability of sophisticated approaches to measuring state fragility presents policymakers with the challenge of assessing the strengths and weaknesses of different approaches. This is relevant, since the use of a particular concept has concrete consequences in terms of the:

- number of states classified as fragile
- ability to distinguish different types of fragility
- accuracy and timeliness of measurement, and
- number of years and countries on which information is available.

The conceptual comparison gives an overview and provides guidance. It compares four concepts: the OECD's States of Fragility (SFR), the Constellations of State Fragility (CSF) developed at the German Development Institute, the Crisis Early Warning (CEW) provided by the German Institute of Global and

Areas Studies (GIGA) and used by the BMZ,² and the fragility component of the Catalogue of Criteria (CoC) for Assessing Development Orientation overseen by the Governance division of the BMZ.

The comparison reveals significant variation between concept designs, even for similar definitions.

These conceptual differences have implications for allocation and programming decisions. The concepts differ, for example, in their empirical scope. The BMZs CEW defines fragility focusing very narrowly on violent conflict. By contrast, the OECD's SFR follows a very broad understanding which includes, e.g., environmental risks such as exposure to natural hazards. Narrow concepts leave out aspects of statehood that may influence the success of development cooperation. Broad concepts, on the other hand, conceive state fragility as an overarching term describing a broad range of political, economic, societal, and environmental deficiencies and thereby become imprecise and lose analytical value and policy relevance. Consequently, focusing on the core state functions of authority, capacity, and legitimacy appears most useful.

Since no single concept on its own fulfils all requirements for supporting strategic policy-making, a combination of existing approaches seems to be expedient. Concepts with a large scope, suitable extension, and founded on qualitative and quantitative public data and an explicit concept design are readily available but usually do not provide timely data. The latter are, however, available through the existing reporting system currently used in the BMZ. As it would be inefficient to develop yet another approach to defining and measuring fragility, the challenge lies in combining existing approaches. This requires a standardized process of combining and weighting information through mixed- and multi-method approaches. For instance, a combination of different sources could increase the precision of measurement, give insights into uncertainties, and improve the imputation of missing data.

Fragility indices are well suited to the strategic management of German bilateral development cooperation as they enable comparisons between states on a single, ideally multidimensional scale.

By contrast, a lack of guidance on what information should be considered relevant, combined with time constraints, can lead decision-makers to over-rely on volatile subjective reasoning prone to psychological biases, increasing the danger of incorrect judgements. The mandatory use of indices as currently practised in the development of country strategies thus promises to improve decision-making. At the same time, portfolio management, programming, and implementing development cooperation also requires an idiographic, in-depth understanding. In particular, highly aggregated measures that reduce dozens of indicators to a binary classification entail the danger of oversimplification.

One approach to reduce the danger of oversimplification is to present results of indices more comprehensively.

More specifically, a visual representation of concepts that includes indicators and the applied weighting mechanism would convey the wealth of information provided by multidimensional concepts and make transparent the influence of individual indicators. Another way to tackle the risks of oversimplification is to make clear the uncertainty of measurement. For instance, a country classified as non-fragile might change to being classified as fragile one year later because it barely passed a certain threshold. Since the underlying measurement is uncertain, such a change in category might be entirely due to chance. Moreover, quantitative indicators can convey a certainty that seldom exists. One way to prevent policymakers from drawing wrong conclusions from data is to report distributions rather than pseudo-exact point estimates.

Beyond possible improvements in presentation, this evaluation also discusses three ways to improve the understanding of fragility.

First, taking into account how fragility spreads across borders would allow possible spillover effects to be assessed. Second, spatial disaggregation of measures of fragility promises to improve our understanding because the level of fragility often varies within states, specifically in developing ones. We present one approach to disaggregate state authority subnationally in Chapter 5. Third, applying

² The BMZ replaced the CEW with a new system called the Escalation Potential Measurement (ESKA – Eskalationspotentialmessung) at the end of 2018.

the notion of fragility and its constitutive dimensions to the regime, administration, and government of a state could further improve the added value of fragility indices in context-sensitive programming.

In summary, we make the following recommendation:

Recommendation 1

Based on a systematic conceptual comparison of four approaches to defining and measuring fragility, this evaluation recommends that the BMZ continue to use fragility indices for strategic portfolio management. Furthermore, the BMZ should continue to constantly evaluate and possibly improve its definitions and measurement of fragility on the basis of clearly defined criteria. This should include an assessment of the extent to which existing approaches could be combined.

As part of the current revision of approaches to measuring fragility, the BMZ should ensure that the results of indices are presented to political decision-makers more comprehensively in order to make underlying concept designs transparent and to do justice to the multidimensional nature of fragility. To achieve this, the BMZ should adopt a visual representation of concepts that includes indicators and the applied weighting mechanism. Second, the BMZ should include the uncertainty of classification and measurement in presentations of fragility scores.

Moreover, with a view to possible future revisions, the BMZ should further disaggregate and contextualize its measurement of state fragility in order to further increase the value of indices for the context-sensitive management of development cooperation. Further disaggregation could be achieved by reporting information for lower-level geographical units or more regularly and by differentiating fragility between different actors within partner countries. Contextualization could be achieved by taking account of possible spillover effects of fragility across borders.

Strategies and allocation: Results and recommendations

For development cooperation to be successful, it needs to be aligned to fragile contexts. Consequently, it is no surprise that the increasing attention given to fragile contexts has resulted in a growing number of strategies on context-sensitive implementation of development cooperation.

This evaluation examines to what extent strategy papers shape German ODA commitments. Strategies provide guiding norms and principles for operations, define policy goals, set action areas, propose appropriate instruments, and explicate assumptions underlying development cooperation in fragile contexts. Beyond influencing policies and politics, strategy papers signal to the public and partner countries the German government's preferences concerning fragility, state-building, and crisis prevention. However, some practitioners and researchers argue that official policies as embodied in strategies have no effect on practice. Furness (2014), for instance, argues that, for the EU, development cooperation in fragile states reveals a "policy-operations" gap. Practitioners sometimes follow a similar reasoning when they claim that strategies are especially hard to implement in fragile contexts. The volatility and complexity of fragile contexts clearly impedes the detailed planning of interventions.

This evaluation uses three steps to investigate whether allocation patterns reflect the content of strategies. First, we describe changes and continuities in the BMZ's strategy towards fragile states. Second, we extract testable implications from strategy documents that evolved after critical junctures. And third, we test empirically to what extent ODA commitments match these testable implications.

The analysis identifies important strategic shifts during two critical junctures: the Al-Qaeda attacks in 2001 and the "Arab Spring" of 2011–12. With each critical juncture, new discourses emerged on how development cooperation can be adapted to the challenges arising from fragility. The 2004 Interministerial Action Plan "Civilian crisis prevention, conflict transformation and peace consolidation" can be considered Germany's first joint strategy targeting fragile states. Subsequently, the BMZ released two major strategies to operationalize its 161 recommendations. This resulted in two approaches to development cooperation supporting crisis prevention in fragile states: peacebuilding and institution-building. In 2012, the Federal Government released the "Interdepartmental Guidelines for a Coherent Policy of the German Federal

Government in Fragile States”. It was released at a time of disappointing results from liberal state-building efforts in Afghanistan, Iraq, and the DRC. Moreover, the five peace- and state-building goals (PSG5) concluded in the 2011 “New Deal for Engagement in Fragile States” served as a reference. The BMZ operationalized the guidelines in the strategy “Development for Peace and Security”. In 2017, the Federal Government released the interministerial guidelines “Preventing Crises, Managing Conflicts, Building Peace”. Efforts to operationalize the guidelines in the BMZ are continuing, with particular attention paid to the 2030 Agenda for Sustainable Development.

This evaluation finds that two approaches have emerged in German development cooperation towards fragile states: peacebuilding and institution-building. Although we have seen major innovations within each of these approaches and, more recently, attempts to integrate them more closely, they have both maintained a considerable degree of specialization. This specialization must be considered against the background that the foci of peace- and institution-building are distinct but complementary. They are distinct because, on the one hand, peacebuilding is about more than building strong and inclusive institutions. For instance, it also includes disarmament, demobilization, and reintegration. On the other hand, institutions affect environmental, economic, and social development and thus go beyond peacebuilding. At the same time, peace- and institution-building are themselves complementary. Institutions form the structural conditions that shape the incentives for societal, economic, and political actors and thereby influence the decision to resort to violence or play by the rules as embodied in institutions. Consequently, peacebuilding processes often centre on the question of shaping institutions. But peace and conflict also affect institutions: institutional changes due to violent conflict can render obsolete an assessment of governance and development orientation.

In the light of the fact that strategies have been implemented well, the institutional specialization within the BMZ seems fit for purpose. At the same time, the scientifically demonstrated mutual influence of institutions and political conflict illustrates that development cooperation in fragile states must always take into account peace- and institution-building together. Consequently, a close collaboration among the responsible organizational units can ensure complementarity. The joint reflection of development orientation, the level of governance, and the CEW are a good step in this direction. The current revision of the CEW and the measurement of the level of governance provides an opportunity to further deepen the cooperation, develop a uniform understanding of fragility, and eliminate possible redundancies by possibly integrating existing approaches (see Recommendation 1). On the one hand, this may include more clearly defining and demarcating the objectives of both approaches; on the other, joint data collection on shared indicators could increase efficiency and thereby improve assessments.

We thus make the following recommendation:

Recommendation 2

Based on a systematic conceptual comparison of approaches to defining and measuring fragility, an analysis of strategy documents, a review of the structure and processes of the German bilateral development cooperation system, and interviews, this evaluation recommends that the BMZ maintain its practice of dealing separately with questions of peace and security, and of governance. At the same time, the BMZ should ensure coordination and complementarity between the responsible organizational units wherever possible. The BMZ should systematically compare concept designs across both divisions with the aim of clearly demarcating approaches in order to avoid duplication and redundancy. Where concept designs complement one another, a formal mechanism for sharing indicators and data collection efforts should be established.

The investigation of patterns of commitment shows that bilateral aid allocations broadly follow the strategies. After 2005, more aid was committed to conflict transformation and peacebuilding, to countries with a German military presence, and to post-conflict countries. For the 2007 and 2009 strategies, more aid was committed to the direct delivery of basic services in countries with low governance and deteriorating development orientation, to public institutions in countries with low governance and improving development orientation, and to the support of government at local level in countries with low governance and constant development orientation. After the release of the 2013 strategy, transitional aid commitments

to non-partner countries in conflict or transition rose, and conflict-prevention programmes in fragile states saw an increase in total and relative commitments. We also find that the more general strategy of exiting from budgetary support also applies to fragile states.

Although these results show that allocation patterns are mostly in line with the policy recommendations tested, we also find some strategies to be less well implemented. First, we do not find evidence of an increase in commitments to basket funds, an instrument promising greater donor coordination and harmonization as well as increased ownership and reduced administrative burden faced by partners. Second, we find weak evidence of decreasing support to non-state actors, contrary to strategic recommendations. Both findings touch upon the question of whether state or non-state actors are suitable partners for development cooperation in fragile contexts. The low level of commitments to non-state actors may influence the efficient use of resources and the political regime of the respective country.

In summary, one of the main challenges lies in the strategic orientation of German development cooperation with regard to the question of what channels of delivery – among them the partner governments and non-state actors – offer the greatest benefits of development cooperation in fragile contexts. The strategies under analysis in this chapter already go a long way in tailoring instruments to the local context by making recommendations that are conditional on the fragility, level of governance, and development orientation of partner countries. However, in some cases implementation is falling behind.

We thus make the following recommendation:

Recommendation 3

Based on an analysis of strategy documents and commitments, and a screening of the literature, this evaluation recommends that the BMZ ensure the implementation of its strategies specifically with regard to the support of non-state actors in fragile contexts. The BMZ should also develop a concrete action plan to assess and possibly adapt guiding principles on cooperation with non-state actors in fragile contexts based on available evidence.

Moreover, the BMZ should continue the good practice of constantly adapting strategies to the ever-improving understanding of successful development cooperation in fragile contexts. The existing practice of tailoring strategic recommendations to contextual conditions is likely to do justice to the increasing analytical precision of scientific research and the validity of scientific explanations that depend on particular background conditions.

Evaluation in fragile contexts: Results and recommendations

Evaluating development interventions helps us to understand the effects of development cooperation and learn lessons for future activities. Fragile contexts can influence the quality of evaluation due, for example, to security concerns, poor infrastructure, or lack of mutual trust. Since evaluation reports are crucial for understanding what works in fragile contexts, a negative impact of fragility on evaluation quality would impede better knowledge about aid effectiveness.

A quantitative analysis shows no correlation between fragility and evaluation quality, but finds an indirect relation through reduced mobility in low-capacity states. On the one hand, we therefore conclude that evaluators generally succeed in maintaining evaluation quality – operationalized through the quality criteria developed by Noltze et al. (2018c) – at the usual level, even under difficult circumstances. On the other hand, a more detailed analysis shows that low state capacity is associated with reduced mobility and thus indirectly with a lower quality of evaluation.

In a complementary qualitative analysis, experts identified reduced mobility and poor project M&E capacities as the main challenges. The experts mentioned travel restrictions and poor infrastructure as causes for the limited mobility. Project M&E is related to various factors such as resource questions – in terms of budget as well as availability and competence of staff – leading to lacking or poor-quality indicators and other project data available for external evaluators.

In summary, the mobility of evaluators and M&E at the project level are central concerns for increasing evaluation quality in fragile contexts. Consequently, a critical review of M&E systems could be useful to identify potential to improve data quality and availability. Approaches to improve M&E systems could include reserving sufficient resources for M&E tasks at project/programme level, incentives such as M&E-capacity-related annual objectives for leadership personnel, and sensitization training specifically for leadership personnel. Moreover, the evaluation capacity of local partners could be improved by creating a sustainable pool of evaluation experts in partner countries.

Additional approaches to increase M&E capacity are budgeting and project planning. Separate budgets for project implementation on the one hand and M&E on the other should limit competition for resources. This has already been recently implemented with regard to evaluations. For instance, the GIZ has a separate budget for financing evaluations. Under the recent GIZ evaluation reform, project evaluations will now be more centralized, being commissioned by the evaluation department of the GIZ headquarters. With regard to project monitoring, module proposals (*Modulvorschläge*) for technical development cooperation do not separately report monitoring costs. However, separate reporting of costs could improve the BMZ's steering capacity in respect of the quality of monitoring and evaluation.

Mobility challenges can be addressed through further strengthening the capacity of conducting remote monitoring and evaluation. Suitable approaches to data collection include remote sensing, digital data collection using hand-held devices, and remote surveys. To make use of these new types of digital data in decision-making processes, it is moreover important to strengthen the capacity to apply and process data.

We thus make the following recommendation:

Recommendation 4

Based on qualitative interviews, observational evidence from an evaluation synthesis, and a screening of the literature, this evaluation recommends that the BMZ, GIZ, and KfW address mobility restrictions on evaluators in fragile contexts and improve M&E at the project level.

First, we recommend that the BMZ require the GIZ and the KfW to further separate out budgets for project implementation on the one hand, and project monitoring and evaluation on the other, to further strengthen the prevention of fungibility and competition for resources.

Second, this evaluation recommends that the GIZ and the KfW address mobility challenges in fragile contexts through the increased use of innovative data-collection methods. This includes strengthening capacity in conducting remote evaluation and monitoring using, e.g., geospatial methods, remote surveys, and digital data collection.

Development interventions in fragile contexts: Results and recommendation

This evaluation sheds new light on the nexus between fragility and project success by synthesizing the results of 471 digitized evaluation reports on GIZ and KfW projects in 85 countries between 2006 and 2016. As a rather broad phenomenon, state fragility is linked to many factors that have been discussed in terms of the success of development cooperation.

Two innovative features of this evaluation promise to improve our understanding of the role of state fragility. First, a multidimensional yet state-based concept of fragility measures the authority, capacity, and legitimacy of partner countries. Second, this evaluation presents a subnational measure of fragility based on violent events in the vicinity of georeferenced locations. Since the level of fragility often varies within states, the use of national fragility scores might be insufficient. Consequently, we retrieve information on project locations from evaluation reports through natural language processing and assign coordinates to each extracted location. This assessment of challenges to the state's monopoly on the legitimate use of force provides a more realistic picture of the context in which development cooperation takes place.

The synthesis shows no evidence that fragility of context correlates with project success as measured by evaluation ratings. We find no evidence of project success differing among most types of national fragility, or of varying project success conditional on conflict events near project sites. The causal

mechanism behind this perhaps surprising result is not clear. It might be that projects are well adapted to fragility in terms of programming and design. The fact that this evaluation finds strategies on fragility to be well implemented in German development cooperation seems to point in this direction. It could also be that the goals of interventions in fragile contexts are set lower, which makes them easier to achieve. However, interventions in highly fragile contexts usually have a dual objective in that they additionally aim at stabilization. A second possible explanation is that any negative effect of fragility is offset by greater benefits of projects due to the possibly low initial level of development. It could also be that evaluators take into account the fragility of the context in their ratings, which would lead to spurious results.

A closer look at the dimensions of fragility shows that projects in countries with higher levels of capacity are, on average, more successful. The results indicate that the extent to which development interventions achieve their objectives inducing sustainable positive change depends on the capacity of partner countries. By contrast, whether interventions do so in cost- and time-efficient ways and whether they are well adapted to the target group seems to depend less on capacity. This finding supports the results of an earlier evaluation by Noltze et al. (2018b), which finds the role of partner countries to be particularly important in rating sustainability.

Correlations between state capacity and project ratings are particularly noticeable in the health sector. This resonates with existing findings that interventions aimed at improving health require efficient and independent bureaucracies as well as an efficient infrastructure of service provision across the whole country (Briebe, 2018). However, we should be cautious with these results as health outcomes are also included in the measurement of state capacity.

The relation between fragility and project success turns out to be context-, sector-, and outcome-specific. Tailoring development cooperation to fragile contexts requires a deeper understanding of the complex mechanisms that determine project success in different sectors across different contexts. Although further research and better data are needed, strengthening state capacity seems to be crucial in order to create a context for successful development cooperation. Consequently, it seems advisable to continue, and possibly to extend, the good practice of assessing state capacity in the management of bilateral development cooperation. Differentiating fragility profiles and a detailed Peace and Conflict Assessment are important steps in this direction. The capacity-related risk assessment could also draw on the successor to the Governance Criteria Catalogue as well as monitoring and evaluations of similar interventions in similar contexts.

Moreover, measures of state capacity could be further improved by disaggregating capacity among different actors and by taking into account subnational variations in state capacity (also see Recommendation 1).

Furthermore, a higher standardization of evaluations should improve the comparability of evaluation reports. From an evaluator's point of view, the validity of indicators with regard to project goals as well as comparability of indicators across different projects are of primary importance. Current efforts by the BMZ, GIZ, and KfW in the joint working group "AG Evaluierung" are aimed in this direction and hence should improve the quality of observational data analysis across different contexts.

We thus make the following recommendation:

Recommendation 5

Based on weak observational evidence from an evaluation synthesis, a screening of research and evaluations, and a review of the structure and processes of the German bilateral development cooperation system, this evaluation recommends that the BMZ systematically assess the risks emanating from a lack of capacity of partner countries for the success of development projects even more than has hitherto been the case. This should include a detailed assessment in module proposals of implementation risks in low-capacity states.

Moreover, the analysis points to the added value of improving the quality of, and standardizing, monitoring and evaluation. Consequently, the joint BMZ-GIZ-KfW working group “AG Evaluierung” should strive to increase comparability across evaluations with a view to enhancing the validity of evaluation syntheses and thereby facilitate joint learning.

ZUSAMMENFASSUNG

Fragile Staatlichkeit als Herausforderung für die Entwicklungszusammenarbeit

Fragile Staatlichkeit ist eine der zentralen Herausforderungen für die Entwicklungspolitik. Die Agenda 2030 für nachhaltige Entwicklung spiegelt dies wider. So lautet Ziel 16: „Friedliche und inklusive Gesellschaften im Sinne einer nachhaltigen Entwicklung fördern, allen Menschen Zugang zur Justiz ermöglichen und effektive, rechenschaftspflichtige und inklusive Institutionen auf allen Ebenen aufbauen“. Funktionierende regulierende staatliche Institutionen haben nicht nur einen direkten Mehrwert wie die Bereitstellung von Kollektivgütern; wenn das legitime staatliche Gewaltmonopol angefochten wird, wenn die gesamtgesellschaftlich verbindliche Regelsetzungs- und -durchsetzungskompetenz eingeschränkt ist und/oder wenn die Bevölkerung die staatliche Herrschaft nicht anerkennt, gefährdet dies ebenfalls den Erfolg von Entwicklungszusammenarbeit.

Der größte Teil der öffentlichen EZ kommt Staaten zugute, die von einigen oder allen der genannten Merkmale fragiler Staatlichkeit betroffen sind. Die deutschen bilateralen Zusagen bilden dabei keine Ausnahme. Abbildung 1 differenziert die zugesagten Zuschüsse und Darlehen nach Art der Fragilität der betroffenen Länder.³ Es wird deutlich, dass der größte Teil der deutschen EZ mit und in fragilen Ländern durchgeführt wird. Dies betrifft insbesondere Staaten mit eingeschränkter Kapazität, d. h. solche, in denen die Verwaltung schwach ist und Grundversorgungsleistungen nur eingeschränkt bereitgestellt werden. Aus diesem Grund widmet das DEval dem Thema Fragilität einen Schwerpunkt, der Ergebnisse unterschiedlicher DEval Evaluierungen zusammenfasst und mit externen Beiträgen zur EZ in fragilen Kontexten kombiniert.

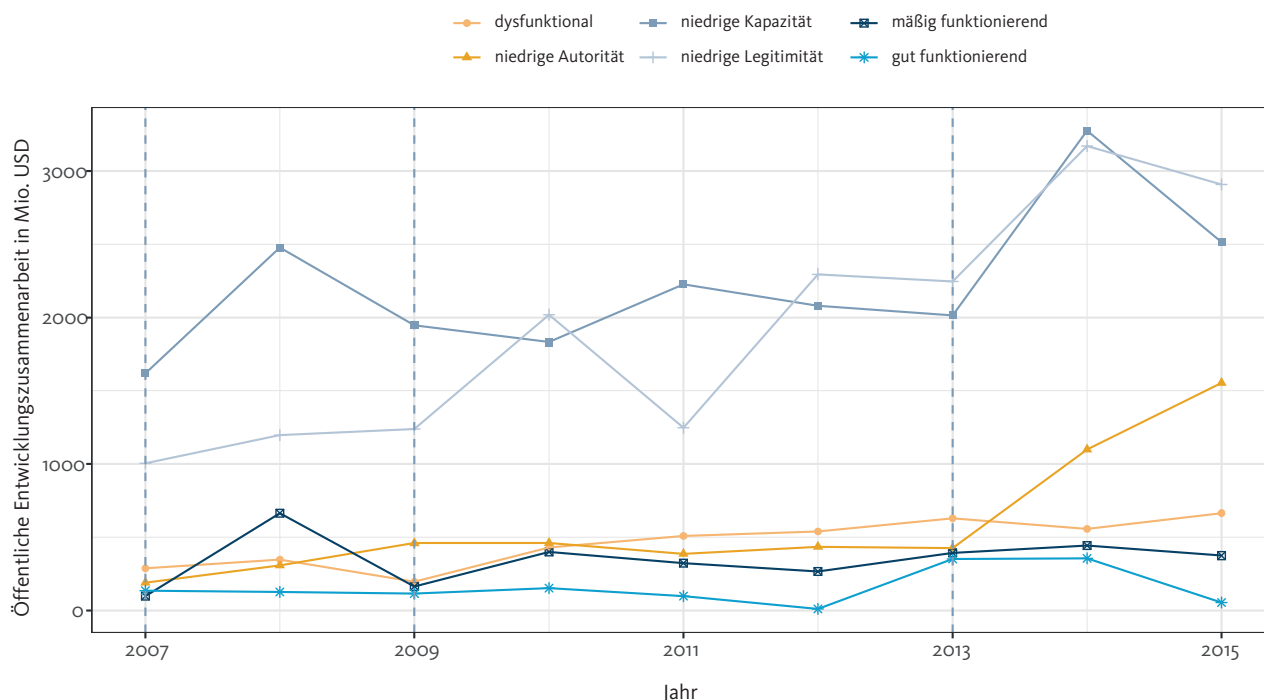
Damit die EZ trotz widriger Rahmenbedingungen in fragilen Staaten erfolgreich arbeiten kann, sollten vier Voraussetzungen erfüllt sein:

1. Ein geeignetes Verständnis des Begriffs staatlicher Fragilität, um die EZ gut auf den jeweiligen lokalen Kontext abzustimmen;
2. Prozesse der Politikformulierung und -durchsetzung, in deren Rahmen die entwicklungspolitischen Strategien zum Umgang mit fragiler Staatlichkeit effektiv konzipiert, ggf. angepasst und umgesetzt werden;
3. geeignete Verfahren, um aus vergangenen Erfahrungen zu lernen und zu ermitteln, welche Ansätze und Lösungen in fragilen Kontexten funktionieren und welche nicht;
4. Kenntnis der Evidenz bezüglich des Erfolgs bestimmter entwicklungspolitischer Ansätze in fragilen Kontexten.

Diese vier Voraussetzungen bilden den Untersuchungsgegenstand des vorliegenden Evaluierungsberichts und werden im Folgenden kurz vorgestellt.

³ Grävingholt et al. (2018) unterscheiden anhand des Gewaltmonopols (Autorität), der Fähigkeit zur Bereitstellung von Grundversorgungsleistungen (Kapazität) und der Anerkennung staatlicher Herrschaft durch die Bevölkerung (Legitimität) sechs Konstellationen staatlicher Fragilität. Dysfunktionale, mäßig funktionierende und gut funktionierende Staaten zeichnen sich durch niedrige, mittlere bzw. hohe Autorität, Kapazität oder Legitimität aus. Staaten niedriger Autorität, Kapazität und Legitimität sind durch niedrige Werte der jeweils benannten Dimension charakterisiert, während sie in den jeweils anderen Dimensionen mittlere oder hohe Werte aufweisen.

Abbildung 1 Bilaterale ODA-Verpflichtungen (Zuschüsse und Darlehen) nach Art der Fragilität



Quelle: Selbst erstellte Abbildung auf der Grundlage von Daten des OECD-Creditor Reporting System und Grävingholt et al. (2019).

Gegenstand der vorliegenden Evaluierung

Diese Evaluierung soll Entscheidungsträger aus Politik und Durchführungspraxis dabei unterstützen, die EZ an die Ausgangsbedingungen in fragilen Ländern anzupassen, um die beabsichtigten Wirkungen zu erzielen. Dabei wird in den Hauptkapiteln (Kapitel 2-5) jeweils eine der vier genannten Voraussetzungen für das Gelingen von EZ in fragilen Staaten untersucht.

Das erste Hauptkapitel widmet sich der Definition und Messung fragiler Staatlichkeit, d.h. der Konzeptualisierung des Phänomens. Hieraus resultierende Indizes und Typologien sind mittlerweile ein fester Bestandteil der Entscheidungsprozesse des Bundesministeriums für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ) und seiner Durchführungsorganisationen (DO). Dieser Teil der Evaluierung bietet eine Grundlage, um die Eignung verwendeter und verfügbarer Indizes für Prozesse der Entscheidungsfindung zu bewerten. Dies ist wesentlich, da die jeweiligen Indizes durch die Wahl und die Art der Kombination bestimmter Indikatoren die Wahrnehmung der Wirklichkeit strukturieren. Resultierende Typologien gruppieren Länder in Klassen und beeinflussen so Entscheidungen über die Eignung bestimmter entwicklungspolitischer Instrumente. Da verschiedene Konzeptionen die Fragilität von Ländern sehr unterschiedlich einschätzen können, hat die Wahl eines bestimmten Ansatzes gemeinhin auch konkrete Auswirkungen auf Allokationsentscheidungen. Insofern ist der konzeptionelle Vergleich in diesem Kapitel von großer Bedeutung für die strategische Steuerung des BMZ-Portfolios.

Das zweite Hauptkapitel behandelt Prozesse der Politikformulierung und -umsetzung. Es basiert auf einer Analyse entwicklungspolitischer Strategien zum Umgang mit fragiler Staatlichkeit sowie einer Allokationsanalyse. Das BMZ hat seit 2005 fünf zentrale Strategien zur Steuerung der EZ in fragilen Staaten aufgelegt. Die Evaluierung zeichnet entsprechende Kontinuitäten sowie den Wandel der deutschen Entwicklungsstrategien im Allgemeinen und des BMZ im Besonderen nach. Darüber hinaus zeigt die Allokationsanalyse anhand zugesagter Mittel der EZ (ODA-Verpflichtungen) auf, welche der strategischen Empfehlungen des BMZ tatsächlich umgesetzt wurden und wo weiterer Handlungsbedarf besteht.

Das dritte Hauptkapitel widmet sich der Qualität von Evaluierungen. Damit behandelt es eines der zentralen Verfahren, um aus Erfahrungen für zukünftige entwicklungspolitische Vorhaben in fragilen Kontexten zu lernen. Die Evaluierungsberichte der beiden größten deutschen

Durchführungsorganisationen – der Deutschen Gesellschaft für Internationale Zusammenarbeit (GIZ) und der KfW Entwicklungsbank (KfW) – bergen eine Vielzahl von Erkenntnissen über Faktoren, die den Erfolg von EZ beeinflussen. Diese Erkenntnisse decken Länder mit Ausgangsbedingungen aller Art ab – darunter auch fragile Staaten, in denen schlechte Sicherheitsbedingungen und mangelnde Infrastruktur die Arbeit von Evaluierenden erschweren können. Die vorliegende Evaluierung analysiert in Anknüpfung an eine frühere DEval-Evaluierung zur Nachhaltigkeit in der deutschen EZ die Qualität von Evaluierungen in Abhängigkeit der Fragilität des Kontextes. Dabei werden Faktoren identifiziert, welche die Durchführung von Evaluierungen erschweren, und daraus Ansatzpunkte abgeleitet, um die Nützlichkeit von Evaluierungen zu erhöhen.

Das vierte Hauptkapitel soll politischen Entscheidungsträgern Einblicke geben, welche entwicklungspolitischen Ansätze in fragilen Kontexten besonders geeignet sind. Eine Synthese von Evaluierungsberichten zeigt, inwiefern die Fragilität des Kontextes mit der Bewertung des Projekterfolgs korreliert. Die Analyse differenziert dabei nach Dimensionen staatlicher Fragilität, unterschiedlichen Sektoren sowie unterschiedlichen Bewertungskriterien für den Projekterfolg. Dies erlaubt es, aus dem komplexen Zusammenspiel zahlreicher Faktoren spezifische Einsichten in mögliche Erfolgsfaktoren für EZ in fragilen Kontexten herauszuarbeiten.

Angewandte Methoden

Der Konzeptvergleich sowie die Strategieanalyse fußen auf einer Auswertung wissenschaftlicher Fachliteratur, offizieller Dokumente, öffentlicher verfügbarer sozioökonomischer Daten und von Interviews.

Strukturen und Prozesse der deutschen bilateralen EZ wurden auf Basis interner und öffentlicher Dokumente mit Blick auf Verfahren der Entscheidungsfindung analysiert. Besonderes Augenmerk galt dabei jenen Verfahren, in denen Strategien zur EZ in fragilen Kontexten oder Fragilitätsindizes eine Rolle spielen. Um die Nützlichkeit der Empfehlungen für BMZ, GIZ und KfW zu erhöhen, beziehen sich die jeweiligen Empfehlungen auf Prozesse nach der Gemeinsamen Verfahrensreform (GVR).

Die Analysen der Evaluierungsqualität und des Projekterfolgs in fragilen Kontexten basieren auf einer Synthese von Evaluierungsberichten der GIZ und der KfW. Diese Synthese verknüpft Merkmale des evaluierten Vorhabens und der Evaluierung mit Daten zum Implementierungskontext. Darüber hinaus wurde im Rahmen dieser Evaluierung ein Verfahren entwickelt, das automatisiert geographische Informationen aus den Evaluierungsberichten extrahiert und so die evaluierten Vorhaben verortet. Diese Daten können mit Informationen über das lokale Konfliktgeschehen verknüpft werden, um die Durchsetzung des staatlichen Gewaltmonopols geographisch disaggregiert zu erfassen. Für die Analyse der Evaluierungsqualität wurden zudem Evaluierungsleitfäden der GIZ und der KfW sowie existierende Meta-Evaluierungen herangezogen.

Die folgenden Abschnitte fassen die vier Hauptkapitel zusammen. Sie diskutieren jeweils die Relevanz des Gegenstands und stellen die zentralen Ergebnisse sowie die hieraus abgeleiteten Empfehlungen dar.

Definition und Messung staatlicher Fragilität: Ergebnisse und Empfehlungen

Kontextspezifische EZ setzt voraus, dass staatliche Fragilität als multidimensionales Phänomen verstanden wird. Insbesondere nach den Anschlägen vom 11. September 2001 hat sich die Anzahl der Ansätze zur Definition und Messung staatlicher Fragilität erhöht. Dabei hat sich die Erkenntnis durchgesetzt, dass Fragilität zahlreiche Dimensionen hat. Anstatt Staaten binär als fragil oder nicht fragil zu klassifizieren oder auf einer eindimensionalen Skala zu bewerten, messen jüngere Ansätze Fragilität anhand von mehreren konstitutiven Eigenschaften und leiten hieraus Typologien ab.

Das breite Angebot komplexer Ansätze der Fragilitätsmessung stellt politische Entscheidungsträger vor die Herausforderung, die Stärken und Schwächen unterschiedlicher Ansätze zu bewerten. Dies ist relevant, da die Nutzung eines bestimmten Konzepts konkrete Folgen hat bezüglich

- der Anzahl der als fragil klassifizierten Länder,
- der Möglichkeit, unterschiedliche Typen von Fragilität zu differenzieren,
- der Präzision und Aktualität der Messung und
- der Anzahl und Jahre, für die Informationen zur Verfügung stehen.

Der Konzeptvergleich dient als Überblick und Orientierungshilfe. Er umfasst vier Ansätze: *States of Fragility* (SFR) der OECD, den am Deutschen Institut für Entwicklungspolitik erarbeiteten Ansatz *Constellations of State Fragility* (CSF), die Krisenfrühwarnung (vom Referat Frieden und Sicherheit des BMZ betreut⁴) und die Fragilitätskomponente des Kriterienkatalogs zur Beurteilung der Governance-Niveaus und der Entwicklungsorientierung, der von der Governance-Abteilung des BMZ betreut wird.

Der Vergleich zeigt, dass ähnliche Definitionen von Fragilität auf sehr unterschiedlichen Konzepten beruhen können. Diese konzeptionellen Unterschiede haben konkrete Implikationen für Allokationsentscheidungen und die Programmierung von EZ. Ein Beispiel: Die Krisenfrühwarnung des BMZ folgt einem engen, auf das Gewaltmonopol bezogenen Verständnis von Fragilität. Die OECD hingegen definiert Fragilität so breit, dass beispielsweise auch Umweltrisiken wie die Gefährdung durch Naturkatastrophen darunterfallen. Zu enge Definitionen blenden bedeutende Faktoren aus, die den Erfolg von EZ beeinflussen können. Wird Fragilität hingegen als Oberbegriff für ein breites Spektrum unterschiedlicher Defizite in den Bereichen Politik, Wirtschaft, Gesellschaft und Umwelt verstanden, verliert der Begriff seine Trennschärfe und seinen analytischen und politikpraktischen Nutzen. Es erscheint daher ratsam, einen Mittelweg zu verfolgen. Dieser sollte sich auf die staatlichen Kernfunktionen Autorität, Kapazität und Legitimität konzentrieren.

Da keiner der verglichenen Ansätze für sich genommen alle Anforderungen für die strategische Steuerung im BMZ erfüllt, erscheint eine Kombination geeigneter Ansätze zielführend. So sind Konzepte verfügbar, die auf methodisch nachvollziehbare Weise qualitative und quantitative Daten auf Basis einer geeigneten Definition für eine große Anzahl von Länderjahren zur Verfügung stellen. Allerdings sind diese Daten meist nicht so aktuell wie diejenigen, die über die Berichterstattung des BMZ und der Durchführungsorganisationen bereitgestellt werden. Da es ineffizient wäre, einen gänzlich neuen Ansatz zu entwickeln, liegt die Herausforderung in der Verknüpfung bestehender Ansätze. Dies wiederum verlangt nach einem standardisierten Prozess, um die vorhandenen Informationen sinnvoll zu kombinieren und zu gewichten. Durch eine solche Kombination unterschiedlicher Quellen kann die Präzision von Einschätzungen erhöht werden. Zudem können mögliche Unsicherheiten abgebildet und fehlende Daten vervollständigt werden.

Fragilitätsindizes eignen sich für die strategische Steuerung bilateraler EZ, da sie Ländervergleiche auf Basis einer einzigen, idealerweise multidimensionalen Skala erlauben. Fehlen in Situationen mit hohem Zeitdruck klare Kriterien zur Beurteilung, welche Informationen als relevant anzusehen sind, kann dies dazu führen, dass Entscheidungsträger gänzlich auf volatile subjektive Kriterien zurückgreifen. Dies erhöht die Anfälligkeit für kognitive Verzerrungen und fehlerhafte Urteile. Mit dem verbindlichen Einsatz von Indizes und Typologien in der strategischen Steuerung, wie er momentan beispielsweise bei der Erstellung von Länderstrategien praktiziert wird, können Entscheidungsprozesse verbessert werden. Zugleich ist für das Portfoliomanagement sowie für die Programmierung und die Implementierung entwicklungspolitischer Vorhaben ein tiefgehendes Verständnis der Gegebenheiten nötig. Insbesondere Konzepte, die eine Vielzahl von Indikatoren auf eine binäre Klassifizierung von Fragilität reduzieren, bergen die Gefahr übermäßiger Vereinfachung.

Um der Gefahr übermäßiger Vereinfachung zu begegnen, bietet sich eine umfängliche Visualisierung der genutzten Konzeptionen an. Vor allem eine Darstellung der zugrunde liegenden Indikatoren und ihrer Gewichtung würde die Mehrebenenstruktur von Konzepten transparent machen, die Fülle der Informationen multidimensionaler Konzepte verdeutlichen und den Einfluss einzelner Indikatoren

⁴ Das BMZ hat die Krisenfrühwarnung Ende 2018 durch ein neues System namens Eskalationspotenzialmessung (ESKA) ersetzt.

nachvollziehbar darstellen. Auch die transparente Darstellung von Unsicherheit hilft, übermäßiger Vereinfachung entgegenzuwirken. Oft ist die Zuweisung von Fällen zu Typen nicht eindeutig. So kann etwa ein Land durch das knappe Überschreiten eines Schwellenwertes im Gegensatz zum Vorjahr als „fragil“ eingestuft werden. Da die zugrunde liegenden Messungen immer mit Unsicherheit behaftet sind, kann ein solcher Kategorienwechsel allein durch Zufall begründet sein. Zudem vermitteln quantitative Indikatoren oft eine Exaktheit, die in der Realität nicht gegeben ist. Werden statt vermeintlich exakter konkreter Werte beispielsweise Verteilungen dargestellt, kann Fehlschlüssen entgegengewirkt werden.

Über mögliche Verbesserungen in der Darstellung hinaus diskutiert diese Evaluierung auch drei mögliche Weiterentwicklungen, mit denen das Verständnis von Fragilität verbessert werden könnte. So kann eine Betrachtung der Ausbreitung von Fragilität über Ländergrenzen hinweg mögliche Ansteckungseffekte besser erfassen. Auch die räumliche Disaggregation der Fragilitätsmessung erlaubt ein besseres Verständnis des Phänomens, da insbesondere in Entwicklungsländern Fragilität innerhalb von Staaten variiert. In Kapitel 5 dieser Evaluierung wird eine Möglichkeit vorgestellt, die Durchsetzung des staatlichen Gewaltmonopols räumlich disaggregiert zu erfassen. Um den Nutzen von Indizes in der kontextsensitiven Programmierung von EZ zu vergrößern, erscheint es zudem sinnvoll, das Phänomen der Fragilität und seiner konstitutiven Attribute auf das politische Regime, die Administration und die Regierung eines Staates zu beziehen (Mazzuca, 2017).

Aus diesen Schlussfolgerungen leiten wir folgende Empfehlung ab:

Empfehlung 1

Basierend auf einem systematischen Vergleich von vier Ansätzen der Fragilitätsdefinition und -messung empfiehlt diese Evaluierung dem BMZ, im Rahmen der strategischen Portfoliosteuerung weiterhin Fragilitätsindizes zu verwenden. Darüber hinaus sollte das BMZ die verwendete Definition und Messung von Fragilität auf der Basis klar definierter Kriterien weiterhin kontinuierlich evaluieren und ggf. verbessern. Dies sollte auch eine Einschätzung dazu beinhalten, inwiefern bestehende Ansätze miteinander kombiniert werden können.

Im Zuge der aktuellen Überarbeitung der Ansätze zur Fragilitätsmessung sollte das BMZ intern eine umfassendere Darstellung der Indizes sicherstellen, um die zugrunde liegenden Konzeptualisierungen transparent zu machen und der Multidimensionalität des Phänomens gerecht zu werden. Dies kann durch eine visuelle Darstellung der verwendeten Indikatoren sowie ihrer Gewichtung erreicht werden. Unsicherheiten in der Klassifikation und der Messung sollten zudem im Zuge der Präsentation von Fragilitätswerten verdeutlicht werden.

Mit Blick auf mögliche zukünftige Überarbeitungen sollte das BMZ die Messung von Fragilität stärker disaggregieren und kontextualisieren, um den Nutzen der Indizes für eine kontextsensitive Steuerung der Entwicklungszusammenarbeit weiter zu steigern. Eine weitere Disaggregation kann dadurch erreicht werden, dass einerseits Werte öfter und für kleinere geographische Einheiten erfasst werden und andererseits die Fragilität unterschiedlicher Akteure in Partnerländern differenziert betrachtet wird. Eine Kontextualisierung kann erreicht werden, indem die mögliche Verbreitung von Fragilität über Ländergrenzen hinweg stärker in den Blick genommen wird.

Strategien und Allokationen: Ergebnisse und Empfehlungen

Um den Erfolg von EZ sicherzustellen, muss diese an fragile Kontexte angepasst werden. Folglich hat sich die vermehrte Beschäftigung entwicklungspolitischer Akteure mit staatlicher Fragilität in einer zunehmenden Zahl von Strategien niedergeschlagen, welche die kontextsensitive Implementierung von EZ zum Gegenstand haben.

Diese Evaluierung untersucht, inwiefern sich Strategiepapiere in Zusagen für EZ niederschlagen. Strategien geben handlungsleitende Normen und Prinzipien für Verfahren vor, definieren Politikziele und Handlungsfelder, empfehlen angemessene Instrumente und machen Annahmen, die der EZ in fragilen Staaten zugrunde liegen, deutlich. Über die Beeinflussung von Politikinhalten und -prozessen hinaus signalisieren Strategiepapiere der Öffentlichkeit und den betreffenden Ländern die Präferenzen deutscher

Politik mit Blick auf Fragilität, den Aufbau und die Förderung staatlicher Strukturen sowie die Krisenprävention. Es ist jedoch unklar, ob strategische Politikempfehlungen Effekte auf die Entwicklungspolitik haben. So attestiert beispielsweise Furness (2014) der EU mit Blick auf die EZ eine Lücke zwischen den in Strategiepapieren ausgedrückten Politikpräferenzen und der Umsetzung vor Ort. Auch Praktiker in der EZ argumentieren nicht selten, dass die Volatilität und Komplexität fragiler Kontexte die Umsetzung strategischer Empfehlungen erschwert.

Diese Evaluierung untersucht den Zusammenhang zwischen Allokationsmustern und Strategien in drei Schritten. Zunächst arbeitet sie die Kontinuität und den Wandel strategischer Empfehlungen heraus. In einem zweiten Schritt werden aus den Strategien überprüfbare Implikationen abgeleitet. Abschließend werden diese Implikationen anhand von Entwicklungszusagen empirisch überprüft.

Die Analyse hat bedeutende strategische Änderungen in Zusammenhang mit zwei einschneidenden Situationen (*critical junctures*) herausgearbeitet: den Anschlägen durch Al-Qaida im Jahr 2001 und dem „Arabischen Frühling“ im Jahr 2012. Insbesondere im Zuge dieser Ereignisse ist ein Diskurs darüber entstanden, wie EZ an die Herausforderungen in fragilen Kontexten angepasst werden kann. Der ressortübergreifende Aktionsplan 2004 „Zivile Krisenprävention, Konfliktlösung und Friedenskonsolidierung“ mit seinen 161 Empfehlungen kann als erste gemeinsame Strategie Deutschlands für fragile Staaten angesehen werden. Das BMZ hat auf den Aktionsplan mit Umsetzungsstrategien reagiert: dem „Übersektoralen Konzept zur Krisenprävention, Konfliktbearbeitung und Friedensförderung in der deutschen Entwicklungszusammenarbeit“ (2005), dem Konzept „Entwicklungsorientierte Transformation bei fragiler Staatlichkeit und schlechter Regierungsführung“ (2007) und dem Strategiepapier „Die Förderung konstruktiver Staat-Gesellschaft-Beziehungen – Legitimität, Transparenz, Rechenschaft“ (2010). Im Jahr 2012 veröffentlichte die Bundesregierung die ressortübergreifenden Leitlinien „Für eine kohärente Politik der Bundesregierung gegenüber fragilen Staaten“. Die Veröffentlichung fiel in einen Zeitraum, in dem die Bemühungen um den Aufbau liberaler Staaten in Afghanistan, dem Irak und der Demokratischen Republik Kongo ernüchternde Ergebnisse zeigten. Darüber hinaus dienten der Bundesregierung die im „New Deal for Engagement in Fragile States“ vereinbarten fünf Ziele zur Friedenskonsolidierung und Staatsbildung (PSG5) als Referenz.⁵ Die Leitlinien hat das BMZ in seinen Strategiepapieren „Entwicklung für Frieden und Sicherheit“ (2013) und „Strategie der entwicklungsfördernden und strukturbildenden Übergangshilfe (ESÜH)“ operationalisiert. Schließlich veröffentlichte die Bundesregierung 2017 die Leitlinien „Krisen verhindern, Konflikte bewältigen, Frieden fördern“. Diese werden derzeit im BMZ operationalisiert, wobei die Agenda 2030 für nachhaltige Entwicklung den Rahmen setzt.

Die vorliegende Evaluierung hat zwei Ansätze in der deutschen EZ mit fragilen Staaten identifiziert: die Friedensförderung und den Aufbau staatlicher Institutionen. Beide Ansätze haben – trotz zahlreicher Neuerungen und jüngerer Bestrebungen nach einer engeren Verzahnung – ein gewisses Maß an Spezialisierung beibehalten und verfolgen unterschiedliche, jedoch komplementäre Ziele. Einerseits geht Friedensförderung über die Bildung starker und inklusiver Institutionen hinaus und umfasst beispielsweise die Entwaffnung, Demobilisierung und Wiedereingliederung von Kämpferinnen und Kämpfern. Andererseits beeinflussen staatliche Institutionen die ökologische, ökonomische und soziale Entwicklung und gehen somit über Friedensförderung hinaus. Zugleich stehen Friedensförderung und Institutionenaufbau komplementär zueinander. Institutionen prägen als Teil der strukturellen Rahmenbedingungen die Anreizstruktur für soziale, ökonomische und politische Akteure. Damit beeinflussen sie deren Entscheidung, im Sinne der institutionalisierten Regeln oder gewaltsam zu handeln. Die Gestaltung staatlicher Institutionen ist somit ein zentrales Element der Friedensförderung. Andererseits kann ein durch politische Konflikte hervorgerufener institutioneller Wandel die Einschätzungen der Regierungsführung und der Entwicklungsorientierung schnell obsolet werden lassen.

⁵ Der New Deal wurde im Jahr 2011 von 44 Staaten und internationalen Organisationen im Rahmen des „4th High Level Forum on Aid Effectiveness“ in Busan unterzeichnet. Er fordert legitime Politik, die Herstellung individueller Sicherheit, Gerechtigkeit, die Verbesserung wirtschaftlicher Grundlagen und die Verwaltung staatlicher Einnahmen sowie den Kapazitätsaufbau für die Erbringung von Dienstleistungen.

Die institutionelle Spezialisierung im BMZ erscheint auch vor dem Hintergrund einer guten Implementierung von Strategien angemessen. Zugleich zeigt die wissenschaftlich nachgewiesene wechselseitige Beeinflussung von Institutionen und politischen Konflikten, dass Friedensförderung und Institutionenbildung immer zusammen zu berücksichtigen sind. Eine enge Kooperation der zuständigen Organisationseinheiten kann die Komplementarität beider Ansätze sicherstellen. Die gemeinsame Betrachtung der Entwicklungsorientierung, des Governance-Niveaus und des Krisenpräventionsbedarfs ist ein Schritt in diese Richtung. Die momentane Überarbeitung der Krisenfrühwarnung sowie der Messung des Governance-Niveaus stellt eine Möglichkeit dar, die Kooperation weiter zu vertiefen, ein einheitliches Verständnis von Fragilität zu entwickeln und eventuelle Redundanzen durch die Kombination von Ansätzen zu beseitigen (siehe Empfehlung 1). Dies ermöglicht einerseits, die Zielsetzung beider Ansätze genau voneinander abzugrenzen; andererseits kann eine gemeinsame Datenerhebung zu möglicherweise geteilten Indikatoren die Effizienz und Qualität der Einschätzungen erhöhen.

Zusammenfassend leiten wir somit die folgende Empfehlung ab:

Empfehlung 2

Auf Basis eines systematischen Konzeptvergleichs, einer Analyse von Strategiedokumenten und Interviews sowie einer Betrachtung der Strukturen und Prozesse der deutschen bilateralen Entwicklungszusammenarbeit empfiehlt diese Evaluierung dem BMZ, die Spezialisierung auf die Themen Frieden und Sicherheit einerseits und auf Governance andererseits beizubehalten. Zugleich sollte das BMZ die Koordination und Komplementarität der mit diesen Themen befassten Organisationseinheiten sicherstellen. Dabei sollten die in den betrauten Organisationseinheiten verwendeten Definitionen und Indikatoren miteinander verglichen werden, um Doppelungen und Redundanzen zu vermeiden. Für den Fall, dass sich Definitionen und Indikatoren überschneiden, sollte ein formaler Mechanismus entwickelt werden, der die gemeinsame Nutzung von Indikatoren und eine gemeinsame Datenerhebung sicherstellt.

Die Analyse bilateraler Zusagen hat ergeben, dass die Mittelallokation weitgehend den Strategien folgt. Bilaterale Zusagen für die Konflikttransformation und Friedensförderung an Länder mit deutscher Militärpräsenz sowie an Länder in der Übergangsphase nach einem Konflikt haben nach 2005 zugenommen. Im Hinblick auf die Strategien von 2007 und 2009 lässt sich ein Anstieg der Mittel für die Erbringung von Basisdienstleistungen in Ländern mit geringem Governance-Niveau und einer sich verschlechternden Entwicklungsorientierung nachweisen. Zudem steigt die Unterstützung öffentlicher Institutionen in Ländern mit schwach entwickelter Governance und einer sich verbessernden Entwicklungsorientierung sowie von Lokalregierungen in Ländern mit geringem Governance-Niveau und konstanter Entwicklungsorientierung. Nach der Verabschiedung der Strategie 2013 stiegen die Verpflichtungen an Nicht-Partnerländer in Konflikt- oder Übergangssituationen. Auch bei Programmen zur Konfliktprävention in fragilen Staaten nahmen die Verpflichtungen zu. Darüber hinaus wurde der Ausstieg aus der Budgethilfe auch in fragilen Staaten umgesetzt.

Auch wenn Allokationsmuster größtenteils den getesteten Politikempfehlungen folgen, wurden einige Strategien weniger gut implementiert. So findet sich keine Evidenz für eine Zunahme der Hilfszusagen für Korbfinanzierungen. Dieses Instrument verspricht eine bessere Koordination und Harmonisierung der Geberaktivitäten sowie eine größere Eigenverantwortung bei zugleich geringerem Administrationsaufwand der Empfängerländer. Für einen Rückgang der Unterstützung nichtstaatlicher Akteure – entgegen strategischer Empfehlungen – gab es eine geringe Evidenz. Dies berührt die Frage, inwiefern staatliche oder nichtstaatliche Akteure geeignete Partner für die EZ in fragilen Kontexten sind. Die vergleichsweise geringen Zusagen an nicht-staatliche Akteure in fragilen Kontexten haben vermutlich negative Auswirkungen auf die effiziente Nutzung der verausgabten Mittel und beeinflussen das politische Regime des betroffenen Landes.

Zusammenfassend lässt sich festhalten, dass eine der zentralen Herausforderungen der strategischen Ausrichtung deutscher EZ darin liegt, geeignete lokale Partner – darunter Regierungen und nichtstaatliche Akteure – für eine erfolgreiche EZ in fragilen Kontexten zu identifizieren. Die Strategien des BMZ bilden bereits eine geeignete Grundlage kontextsensitiver EZ, da sie Politikempfehlungen in Abhängigkeit der Fragilität, der Regierungsführung und der

Entwicklungsorientierung von Partnerländern formulieren. Die Implementierung in Form einer Anpassung der Allokationen hinkt diesen Empfehlungen in wenigen Fällen jedoch hinterher.

Zusammenfassend leiten wir somit die folgende Empfehlung ab:

Empfehlung 3

Auf Basis einer Analyse von Strategiedokumenten und Allokationszusagen sowie einer Literaturrecherche, empfiehlt die Evaluierung dem BMZ die Implementierung der Strategien insbesondere bezüglich der Unterstützung nichtstaatlicher Akteure in fragilen Kontexten sicherzustellen. Auch sollte ein konkreter Maßnahmenplan erstellt werden, um bestehende handlungsleitende Prinzipien für den Umgang mit nichtstaatlichen Akteuren in fragilen Kontexten zu prüfen und auf Basis vorliegender Erkenntnisse ggf. anzupassen.

Darüber hinaus sollte das BMZ seine gute Praxis fortführen, Strategien kontinuierlich an das zunehmende Wissen zu erfolgreicher Entwicklungszusammenarbeit in fragilen Kontexten anzupassen. Die bestehende Praxis, strategische Empfehlungen auf Kontextbedingungen zuzuschneiden, ist geeignet, der zunehmenden analytischen Präzision wissenschaftlicher Untersuchungen und der Tatsache, dass die Gültigkeit wissenschaftlicher Erklärungen je nach geltenden Hintergrundbedingungen variiert, gerecht zu werden.

Evaluierung in fragilen Kontexten: Ergebnisse und Empfehlungen

Die Evaluierung von Entwicklungsmaßnahmen soll dazu beitragen, die Auswirkungen der EZ besser zu verstehen, um daraus Schlussfolgerungen für die Gestaltung künftiger Vorhaben zu ziehen. Evaluierungsberichte stellen somit eine wesentliche Grundlage für das Verständnis geeigneter entwicklungspolitischer Ansätze dar. Fragile Kontexte können Evaluierungen erschweren, beispielsweise aufgrund von Sicherheitsbedenken, schlechter Infrastruktur oder eines geringen Vertrauens in der Bevölkerung. Ein entsprechender negativer Einfluss von Fragilität auf die Evaluierungsqualität würde auch ein besseres Verständnis der Wirksamkeit von EZ erschweren.

Eine quantitative Analyse hat keine direkte Korrelation zwischen Fragilität und Evaluierungsqualität ergeben; allerdings wurde ein indirekter Zusammenhang deutlich. So hat sich gezeigt, dass es den Evaluierenden generell auch unter schwierigen Umständen gelingt, die Evaluierungsqualität – gemessen anhand der von Noltze et al. (2018b) entwickelten Qualitätskriterien – auf üblichem Niveau zu halten. Andererseits zeigte eine detailliertere Analyse, dass sich eine geringe staatliche Kapazität von Partnerländern negativ auf die Mobilität der Evaluatoren auswirkt. Somit hängt die Fragilität indirekt mit einer geringeren Evaluierungsqualität zusammen.

Im Rahmen der qualitativen Analyse identifizierten Expertinnen und Experten als Herausforderungen insbesondere eine eingeschränkte Mobilität sowie schwache Monitoring- und Entwicklungs(M&E)-Systeme auf Ebene der Vorhaben. Als zentrale Ursachen fehlender Mobilität wurden Reisebeschränkungen und eine schlechte Infrastruktur genannt. Zur Erklärung schwacher M&E-Systeme auf Projektebene wurden mehrere mögliche Ursachen angegeben, darunter begrenzte Budgets und die eingeschränkte Verfügbarkeit und Kompetenz des Personals. Beides kann dazu führen, dass Indikatoren und Projektdaten nicht oder nur eingeschränkt verwertet werden können.

In der Gesamtbetrachtung erweisen sich die Mobilität von Evaluatoren sowie die Qualität der M&E-Systeme auf Ebene der Vorhaben als zentrale Ansatzpunkte, um die Evaluierungsqualität in fragilen Kontexten weiter zu verbessern. Es erscheint somit geboten, bestehende Monitoring-Systeme kritisch zu beleuchten, um Verbesserungspotenzial bei Datenqualität und -verfügbarkeit zu identifizieren. Denkbare Verbesserungen auf Modul- und Programmebene umfassen die Bereitstellung hinreichender Mittel für die Durchführung von M&E, die Formulierung konkreter M&E-bezogener Zielsetzungen für das Führungspersonal sowie Sensibilisierungstrainings insbesondere für das Führungspersonal. Ferner kann die Evaluierungskapazität lokaler Partner gestärkt werden, indem etwa ein Pool lokaler Expertinnen und Experten in den Partnerländern aufgebaut wird.

Weitere mögliche Ansätze zur Stärkung der M&E-Kapazität betreffen die Budgetierung und die Projektplanung. So könnte einer Konkurrenz um Ressourcen durch eine getrennte Budgetierung von Implementierung und M&E vorgebeugt werden. Die GIZ hat beispielsweise ein eigenes Budget für die Finanzierung von Evaluierungen. Projektevaluierungen werden seit der jüngsten Reform des Evaluierungssystems zentralisiert durch die Evaluierungsabteilung der GIZ in Auftrag gegeben. Beim Projektmonitoring hingegen verlangen Modulvorschläge für die technische Zusammenarbeit bisher keine getrennte Budgetierung. Eine separate Aufstellung der Kosten könnte jedoch die Steuerungskapazität des BMZ bezüglich der zu erwartenden Qualität von M&E stärken.

Den Herausforderungen eingeschränkter Mobilität kann durch eine Stärkung der Kapazitäten in der Durchführung von *Remote Monitoring and Evaluation* begegnet werden. Geeignete Ansätze umfassen die Fernerkundung, die digitale Datenerfassung durch tragbare Geräte und *remote surveys*. Die Nutzung der in diesem Zusammenhang gesammelten Daten in Entscheidungsprozessen erfordert ferner eine Stärkung der Kapazität in der Datenverarbeitung.

Zusammenfassend leiten wir somit die folgende Empfehlung ab:

Empfehlung 4

Auf der Grundlage qualitativer Interviews, einer Auswertung von Beobachtungsdaten aus einer Evaluierungssynthese und einer Literaturrecherche empfiehlt die Evaluierung dem BMZ, der GIZ und der KfW, insbesondere Mobilitätsbeschränkungen von Evaluierenden in fragilen Kontexten anzugehen. Zudem sollten die M&E-Systeme auf Ebene der Vorhaben verbessert werden.

Das BMZ sollte von der GIZ und der KfW eine noch stärker gesonderte Budgetierung für die (a) Implementierung und (b) das Monitoring und die Evaluierung verlangen. So kann einer Konkurrenz um die Finanzierung vorgebeugt werden. Auch sollten GIZ und KfW vermehrt innovative Methoden der Datenerhebung nutzen, um den Herausforderungen zu begegnen, die sich in fragilen Kontexten durch Mobilitätsbeschränkungen ergeben. Dies betrifft etwa die Stärkung der Kapazitäten in der Durchführung von *Remote Monitoring and Evaluation*, beispielsweise durch Fernerkundung, *remote surveys* und digitale Methoden der Datenerhebung.

Diese Evaluierung hat den Zusammenhang zwischen Fragilität und dem Erfolg entwicklungspolitischer Vorhaben untersucht. Hierfür wurden die Ergebnisse von 471 digitalisierten Evaluierungsberichten von GIZ und KfW zu Vorhaben in 85 Ländern, die zwischen 2006 und 2016 durchgeführt wurden, **synthetisiert**. Als relativ umfassendes Phänomen betrifft Fragilität viele der in der Literatur diskutierten möglichen Erfolgsfaktoren entwicklungspolitischer Vorhaben. Ein Einfluss des Phänomens auf den Erfolg entwicklungspolitischer Vorhaben ist daher wahrscheinlich.

Zwei innovative Merkmale der Untersuchung ermöglichen ein verbessertes Verständnis von Fragilität. Zum einen greift die Evaluierung auf ein multidimensionales, aber zugleich staatsbezogenes Verständnis von Fragilität zurück, um die Autorität, Kapazität und Legitimität von Partnerländern zu unterscheiden. Zum anderen wurde im Rahmen der Evaluierung ein geographisch disaggregiertes Fragilitätsmaß entwickelt, das auf Konfliktereignissen in der Nähe von Vorhaben beruht. Da das Ausmaß von Fragilität insbesondere in Entwicklungsländern innerhalb von Staaten variiert, ist die alleinige Messung auf Ebene des Staates unzureichend. Im Rahmen der Evaluierung wurden daher Ortsangaben durch Verfahren der linguistischen Datenverarbeitung aus vorliegenden Evaluierungsberichten extrahiert, qualitätsgeprüft und automatisiert mit Koordinaten versehen. Die hierdurch ermöglichte Erfassung lokaler Einschränkungen in der Durchsetzung des staatlichen Gewaltmonopols vermittelt ein realistischeres Bild des Kontextes, in dem entwicklungspolitische Vorhaben implementiert werden.

Die Synthese liefert auf den ersten Blick keine Hinweise dafür, dass die Fragilität des Kontextes mit dem Erfolg entwicklungspolitischer Vorhaben – gemessen anhand der Bewertungen aus Evaluierungsberichten – korreliert. Dies gilt sowohl für eine Betrachtung unterschiedlicher Typen von Fragilität als auch für das neu entwickelte Messinstrument der subnationalen staatlichen Autorität. Der Grund für diesen überraschenden Befund ist unklar. Es könnte sein, dass Vorhaben gut an fragile Kontexte angepasst werden. Darauf deutet die Tatsache hin, dass sich strategische Empfehlungen in Allokationen

niederschlagen – wie in dieser Evaluierung ermittelt. Weiterhin könnten Ziele von Vorhaben in fragilen Kontexten weniger ambitioniert formuliert und daher leichter zu erreichen sein. Allerdings haben Vorhaben in fragilen Kontexten teilweise eine duale Zielsetzung, da sie zugleich auf Stabilisierung zielen. Eine weitere mögliche Erklärung lautet, dass ein möglicher negativer Einfluss des fragilen Kontextes auf den Projekterfolg durch einen höheren Nutzen von Vorhaben bei niedrigem Ausgangsniveau ausgeglichen wird. Ferner könnten Evaluierende die Fragilität des Kontextes bei ihrer Bewertung berücksichtigen, was allerdings die Vergleichbarkeit der Berichte einschränken und Ergebnisse verfälschen würde.

Bei genauerer Betrachtung der einzelnen Dimensionen von Fragilität zeigt sich, dass Projekte in Staaten mit höherer staatlicher Kapazität durchschnittlich erfolgreicher sind. Insbesondere das Ausmaß, in dem Vorhaben ihre Ziele erreichen und dabei nachhaltig positive Veränderungen erwirken, korreliert mit der Kapazität der Länder, in denen die Vorhaben implementiert werden. Im Gegensatz dazu scheinen die Effizienz und die Relevanz von Projekten nicht mit der staatlichen Kapazität zu korrelieren. Diese Ergebnisse stützen die Erkenntnisse einer früheren Evaluierung durch Noltze et al. (2018a). Diese hatte die Kapazität von Partnerländern als besonders bedeutend für die Bewertung der Nachhaltigkeit von Vorhaben identifiziert.

Die Korrelation zwischen staatlicher Kapazität und Projekterfolg ist im Gesundheitssektor besonders ausgeprägt. Dies steht im Einklang mit dem Befund, dass Interventionen, die eine Verbesserung der Gesundheit zum Ziel haben, effiziente und unabhängige Bürokratien sowie eine effiziente Infrastruktur zur Dienstleistungserbringung im gesamten Staatsgebiet voraussetzen (Briebea, 2018). Dabei muss jedoch berücksichtigt werden, dass Kennzahlen der Gesundheitsversorgung zugleich einen Teilindikator staatlicher Kapazität darstellen.

Die Beziehung zwischen Fragilität und Projekterfolg erweist sich somit als kontext-, sektor- und ergebnisspezifisch. Die Anpassung entwicklungspolitischer Vorhaben an fragile Kontexte erfordert ein tieferes Verständnis der Mechanismen, die unterschiedliche Arten des Projekterfolgs in unterschiedlichen Sektoren und Kontexten erklären. Auch wenn es weiterer Forschung und einer besseren Datenlage bedarf, erscheint die Stärkung staatlicher Kapazität als wesentlich, um den Erfolg entwicklungspolitischer Vorhaben sicherzustellen. Es erscheint daher sinnvoll, im Rahmen der strategischen Planung und Steuerung die gute Praxis einer detaillierten Einschätzung staatlicher Kapazität durch das BMZ und die DO beizubehalten und ggf. auszubauen. Die Differenzierung von Fragilitätsprofilen und ein detailliertes *Peace and Conflict Assessment* sind wichtige Schritte in diese Richtung. Die kapazitätsbezogene Risikoeinschätzung könnte ebenfalls auf die Ergebnisse des Nachfolgers des Governance-Kriterienkatalogs sowie auf Monitoring- und Evaluierungsberichte ähnlicher Vorhaben in ähnlichen Kontexten zurückgreifen. Durch die Aufschlüsselung der Daten zu staatlicher Kapazität nach unterschiedlichen Akteuren sowie bezüglich der räumlichen Verteilung innerhalb von Staaten kann die Präzision der getroffenen Einschätzungen weiter erhöht werden (siehe auch Empfehlung 1).

Eine höhere Standardisierung von Evaluierungsberichten kann für eine bessere Vergleichbarkeit sorgen. Aus Evaluierungsperspektive ist neben der Validität insbesondere die Vergleichbarkeit von Indikatoren zwischen Vorhaben von Bedeutung. Laufende Anstrengungen des BMZ, der GIZ und der KfW im Rahmen der AG Evaluierung zielen auf eine höhere Standardisierung ab. Diese kann für eine bessere Belastbarkeit von Analysen auf Basis von Beobachtungsdaten über unterschiedliche Kontexte hinweg sorgen.

Zusammenfassend leiten wir somit folgende Empfehlung ab:

Empfehlung 5

Auf Grundlage einer Auswertung von Beobachtungsdaten aus einer Evaluierungssynthese, einer Durchsicht der Fachliteratur und von Evaluierungen sowie einer Betrachtung der Strukturen und Prozesse der deutschen bilateralen Entwicklungszusammenarbeit empfiehlt diese Evaluierung dem BMZ, die durch schwache staatliche Kapazität von Partnerländern bedingten Risiken für den Erfolg entwicklungspolitischer Vorhaben noch stärker als bisher in Entscheidungsprozessen zu berücksichtigen. Hierfür sollte unter anderem eine detaillierte Analyse der Implementationsrisiken in Partnerländern mit niedriger staatlicher Kapazität im Rahmen der Modulvorschläge durchgeführt werden.

Darüber hinaus weist die Analyse auf den Mehrwert einer Qualitätssteigerung und Standardisierung von Monitoring und Evaluierung hin. 2017 wurde unter Federführung des BMZ die „Arbeitsgruppe Evaluierung“ ins Leben gerufen. Die daran beteiligten Parteien BMZ, GIZ und KfW sollten den Austausch nutzen, um eine bessere Vergleichbarkeit von Evaluierungen herzustellen. Dadurch kann die Validität von Evaluierungssynthesen verbessert und somit gemeinsames Lernen erleichtert werden.

CONTENTS

Acknowledgements	v
Executive Summary	vi
Zusammenfassung	xvi
Abbreviations and Acronyms	xxix
1. Introduction.....	1
1.1 Evaluation questions	3
1.2 Methods.....	3
2. Defining and measuring fragility.....	5
2.1 Introduction.....	6
2.2 Statehood and fragility	8
2.3 Method.....	12
2.4 Concept comparison.....	13
2.5 Results and recommendations.....	20
3. Strategies	24
3.1 Introduction.....	25
3.2 Method.....	26
3.3 Germany's strategy towards fragile states	28
3.3.1 The post-9/11 juncture: 2004 Action Plan	28
3.3.2 Responding to the Arab Spring: a new approach to stabilization	34
3.3.3 At the crossroads: the 2017 interministerial guidelines "Preventing crises, managing conflicts, building peace"	37
3.4 Empirical results	38
3.5 Results and recommendations.....	41
4. Evaluation.....	44
4.1 Introduction.....	45
4.2 Theory: quality determinants	45
4.3 Method.....	46
4.4 Empirical results	50
4.5 Results and recommendations.....	55
5. Development interventions in fragile contexts.....	58
5.1 Introduction	59
5.2 What we know about the success of development cooperation	59
5.3 Data and method	61
5.4 Empirical results	63

5.5	Results and recommendations.....	68
6.	References.....	70

Figures

Figure 1	Bilateral ODA commitments (grants and loans) by type of fragility	vii
Figure 2	ODA excluding debt relief by level of fragility according to OECD States of Fragility	2
Figure 3	Concept design.....	12
Figure 4	State fragility and evaluation quality, and possible mechanisms linking them.....	48
Figure 5	Fulfilled quality criteria by high or low fragility context	50
Figure 6	Boxplots depicting the evaluation quality index differentiated by type of fragility	51
Figure 7	Subnational fragility by CSF authority component.....	62
Figure 8	Project ratings by type of fragility.....	63
Figure 9	Project ratings by fragility component	64
Figure 10	Project rating by subnational fragility.....	64
Figure 11	Project rating by state capacity in different sectors.....	66

Tables

Table 1	Methods by evaluation question	4
Table 2	Fragility classification for different concepts.....	7
Table 3	Criteria for conceptual comparison	12
Table 4	Overview of concepts of fragility	15
Table 5	Methods of aggregation	17
Table 6	Illustration of weakest link approach to aggregation.....	18
Table 7	Types of fragility	19
Table 8	Key strategy papers relating to state fragility published by the German Federal Government and the BMZ, 2004–2017.....	26
Table 9	Evidence aggregation	28
Table 10	Results of empirical analysis.....	38
Table 11	Summary results from multivariate regression analysis of the effect of fragility on evaluation quality.....	52
Table 12	Summary results of the mediation analysis	53

ABBREVIATIONS AND ACRONYMS

AA	Auswärtiges Amt (Federal German foreign ministry)
BMZ	Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung (Federal Ministry for Economic Cooperation and Development)
BMZ Governance division	Division responsible for sectoral and thematic policies, governance, democracy, and rule of law
BMZ Peace and Security division	Division responsible for peace, security, and disaster risk management
BTI	Bertelsmann Transformation Index
CEW	Crisis Early Warning (BMZ), provided by GIGA
CoC	Catalogue of Criteria for Assessing Development Orientation (BMZ)
CRS	Creditor Reporting System (OECD DAC)
CSF	Constellations of State Fragility (DIE)
DAC	Development Assistance Committee (OECD)
DIE	Deutsches Institut für Entwicklungspolitik (German Development Institute)
EU	European Union
FC	Financial Cooperation
FS-marker	Peace and Security marker (BMZ)
GIGA	German Institute of Global and Area Studies
GIZ	Deutsche Gesellschaft für Internationale Zusammenarbeit GmbH (German federal enterprise for international cooperation)
IATI	International Aid Transparency Initiative
KfW	German government-owned development bank
NATO	North Atlantic Treaty Organization
ODA	Official development assistance
OECD	Organisation for Economic Co-operation and Development
SDG	Sustainable Development Goal (UN)
SFR	States of Fragility (OECD)
TC	Technical Cooperation
UN	United Nations
USD	United States dollar

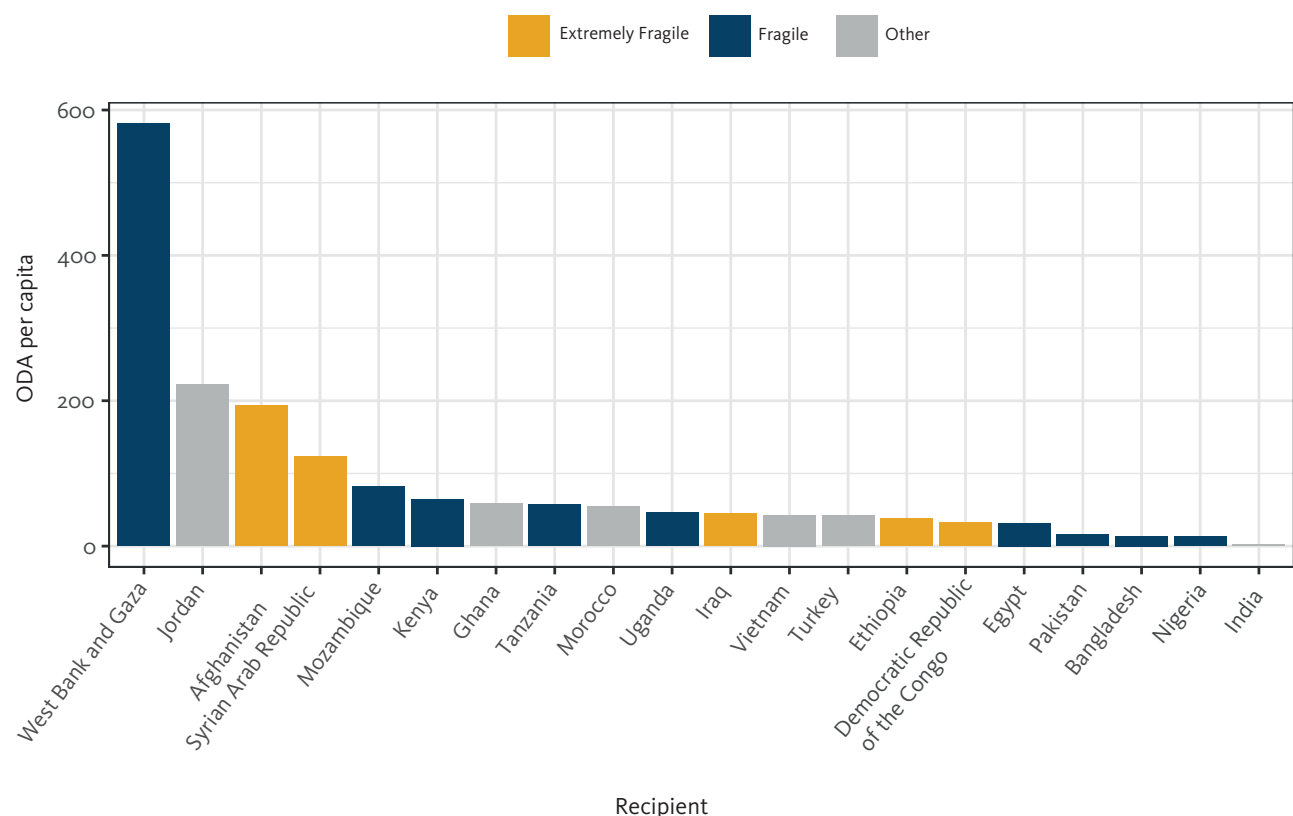
1. INTRODUCTION

Peace- and statebuilding and conflict prevention are widely acknowledged goals of development cooperation. In the light of the challenges posed by fragility, in 2011 over 40 states and organizations signed the New Deal for Engagement in Fragile States. This acknowledges stable political, judicial, and economic institutions as foundations for development, calls for whole-of-government approaches and nationally owned policies towards clearly defined goals, and commits members to effective and aligned use of development cooperation (see Hearn, 2016). The g7+, a voluntary association of self-declared fragile states, the OECD's International Network on Conflict and Fragility, and the Civil Society Platform for Peacebuilding and Statebuilding together advance the implementation of the New Deal. The 2016 Stockholm Declaration recommitted members to the New Deal principles (OECD, 2016).

State fragility puts at risk the efforts of development cooperation in achieving the Sustainable Development Goals (SDGs) (UN, 2015). The 2030 Agenda for Sustainable Development reflects the challenge of development cooperation in fragile contexts: SDG 16 sets out the vision to “[p]romote peaceful and inclusive societies [...], provide access to justice for all and build effective, accountable and inclusive institutions” (UN, 2015, p.14).

Spending patterns of development cooperation underline the importance of state fragility. Between 2011 and 2014, 64% of worldwide official development assistance (ODA) excluding debt relief went to contexts considered fragile, according to the OECD (2016, p.111). Figure 2 shows the top 20 cooperation recipients in terms of ODA per capita in constant 2014 USD and their level of fragility according to the OECD. Only six recipients are considered non-fragile states.

Figure 2 ODA excluding debt relief by level of fragility according to OECD States of Fragility



Source: own figure.

German development cooperation is no exception. According to the German Federal Ministry for Economic Cooperation and Development (BMZ) (2013a), more than half the countries in which German development cooperation is active are affected by conflict, fragility, and violence. In 2015, Germany spent some 4.2 billion euro of ODA in fragile states (BMZ, 2017a, p.109). Consequently, a key policy question is how development cooperation can operate successfully in fragile contexts.

The high prevalence of fragility, the extent of its impact, and the large amounts of money spent by Germany and other national and international donors in fragile contexts make evaluations necessary to provide a sound foundation for future decision-making. While the New Deal provides evidence that the international community is clear about the broader aims and core strategies, there is no comparative assessment of development cooperation in fragile contexts.

This evaluation contributes to a better understanding of German development cooperation in fragile contexts. It does so by reviewing the concepts and strategies underlying the German approach to fragile partner countries (portfolio analysis) and by examining the impact of fragility on project success and project evaluation (evaluation synthesis).

1.1 Evaluation questions

This evaluation consists of two complementary studies:

- A portfolio analysis examines the approach of the BMZ to dealing with fragile contexts. It combines a comparative assessment of the German approach to defining and measuring state fragility, an analysis of continuity and change in German strategy towards fragile states, and a description of allocation patterns towards fragile states.
- An evaluation synthesis systematically integrates evaluation reports by the GIZ and the KfW to identify the influence of fragility on the evaluation and success of development cooperation interventions. It draws on evaluation reports from 2006 to 2016. A first step examines the influence of fragility on the quality of evaluation reports; a second investigates the success of German bilateral development cooperation interventions according to the fragility of the contexts.

The two parts of the evaluation address four questions.

Portfolio analysis

1. Which concepts of state fragility are suitable for the management of German development cooperation?
2. What are the continuities of German strategy towards fragile contexts and where were major changes seen? To what extent do allocation patterns reflect strategic recommendations?

Evaluation synthesis

3. How do fragile contexts affect the quality of evaluation?
4. How do project ratings differ according to the fragility of contexts?

1.2 Methods

The portfolio review draws on academic literature, official documents, publicly available data, and interviews to analyse concepts, strategies, and allocation patterns. The evaluation synthesis combines multidimensional data on state fragility, geographically disaggregated data on state authority, and qualitative data to assess the quality of evaluations and project ratings from evaluation reports. Table 1 gives a summary of the methods applied.

Table 1 Methods by evaluation question

Aspect of analysis	Method of analysis	Data
Concept of state fragility	Literature review, comparative conceptual analysis	Academic literature, publications and directives of aid agencies, interviews
Strategies and allocation structure	Qualitative analysis of critical junctures, quantitative description of allocation	Academic literature, publications and directives of aid agencies, interviews, OECD DAC CRS
Quality of evaluation reports	Literature review, statistical analysis, interviews	Publicly available secondary data, sample of KfW and GIZ evaluation reports
Success of development cooperation interventions	Literature review, statistical analysis, text mining	Publicly available secondary data, sample of KfW and GIZ evaluation reports

Development cooperation interventions are often confined to specific regions at subnational level. Consequently, valid measurement of fragility and success should focus on subnational dynamics. Data on German development cooperation, however, is mostly confined to the national level. Although satellite imagery and the increased availability of geographically disaggregated data have recently improved data quality, there remains a substantial gap with regard to precise data.

This is even more true in respect of the present evaluation, since data collection in fragile contexts is more complicated than in non-fragile ones. In particular, rigorous impact evaluations are rare (for an exception, see Blattman et al., 2014).

To tackle issues of data quality and availability, we use text mining to derive precise data evaluation reports provided by the GIZ and the KfW and link them to publicly available geographically disaggregated data on fragility and other contextual factors of interest.

As this is an evaluation synthesis, the OECD DAC evaluation criteria were only of limited applicability: while the question of relevance was crucial for the chapters on concepts (2) and strategies (3), we looked at effectiveness in the chapters on evaluation quality (4) and project success (5). Analysing in detail efficiency, sustainability, and impact was neither intended nor feasible within the framework of this study.

2. DEFINING AND MEASURING FRAGILITY

2.1 Introduction

The concept of state fragility is high on the international development agenda. At the same time, its meaning is disputed among policymakers, think-tanks, and academics (Bueger and Bethke, 2014; Faust et al., 2013). There is no international agreement on which definition should be used (BMZ, no date), which is the source of some confusion among international actors (Putzel and Di John, 2012).

It thus comes as no surprise that we can also find different understandings of fragility within the BMZ. On the one hand, the BMZ follows a narrow understanding of fragility as including those states that have a high or acute potential for the violent escalation of conflict (BMZ, no date). Countries at risk of violent escalation are derived from the Crisis Early Warning (CEW). On the other hand, the BMZ puts forward a wider definition of fragility, including deficiencies in core attributes of statehood but also low levels of governance and development orientation, as well as political conflict (ibid.). Since 2016, indicators in the Catalogue of Criteria for Assessing Development Orientation (CoC) have been rearranged to measure fragility based on a state's capacity, legitimacy, and authority.

To be of use, concepts and derived typologies should adequately describe, structure, and classify real-world cases. However, even semantically similar concepts can vary fundamentally in their design. For instance, the measurement of fragility within the CoC and the Deutsches Institut für Entwicklungspolitik's (DIE) Constellations of State Fragility (CSF) builds on the same definition of fragility, based on the attributes of authority, capacity, and legitimacy. At the same time, the concept designs differ fundamentally in respect of the indicators and the way indicator values are aggregated to the overall measure of fragility.

Different concept designs, however, convey a different understanding of real-world phenomena and can thus lead to different lists of fragile states. Table 2 exemplifies the consequences of selecting a certain definition of fragility by comparing classifications by the CSF project and the OECD's States of Fragility (SFR) for the year 2015.⁶ As the table illustrates, the classification of states differs between concepts.

Concepts of state fragility are widely used in policy-making processes (BMZ, 2017a). It is thus reasonable to assume that the choice of concepts in decision-making processes influences development policies. Policymakers and practitioners should thus be aware of concept designs.

Concepts of state fragility are under continuous development. In 2016, the OECD introduced a new methodology to measure state fragility (OECD, 2016). In 2018, the DIE presented the data for a new concept of state fragility (Grävingholt et al., 2018). Moreover, the BMZ updated its concepts for Crisis Early Warning at the end of 2018 through the Escalation Potential Analysis (ESKA), which focuses on consideration of the security aspect of fragility. The BMZ also plans to update its concepts for assessing the level of governance and development orientation.

This chapter compares four concepts of state fragility. We show that a lack of agreement on definitions has very real empirical consequences. By comparing concepts and demonstrating their empirical implications, this chapter helps policymakers to choose suitable definitions and data.

⁶ Data from Crisis Early Warning and the CoC had to be excluded because they are confidential.

Table 2 Fragility classification for different concepts

	CSF	SFR		CSF	SFR
Afghanistan			Madagascar		
Albania			Malawi		
Algeria			Mali		
Armenia			Mauritania		
Azerbaijan			Mexico		
Bangladesh			Moldova		
Belarus			Mongolia		
Benin			Montenegro		
Bolivia			Morocco		
Bosnia and Herzegovina			Mozambique		
Brazil			Myanmar		
Burkina Faso			Namibia		
Burundi			Nepal		
Cambodia			Nicaragua		
Cameroon			Niger		
Central African Republic			Nigeria		
Colombia			Pakistan		
Congo, DR			Palestinian Territories		
Costa Rica			Paraguay		
Cote d'Ivoire			Peru		
Cuba			Philippines		
Dominican Republic			Rwanda		
Ecuador			Senegal		
Egypt			Serbia		
El Salvador			Sierra Leone		
Ethiopia			Somalia		



Source: own table based on CSF and SFR data for 2015.

2.2 Statehood and fragility

Given its wide use, great empirical extension, and considerable variability of forms, a straightforward definition of the concept of fragility is elusive. The following aims to clarify the concept of fragility by relating it to other concepts. We first introduce our concept of fragility before proceeding to a comparison with other concepts.



Fragility and the continuum between state collapse and consolidated statehood

A dominant approach is to relate fragility to statehood. One can think of fragility as lying on the spectrum between state collapse and consolidated statehood (see Lambach et al., 2015). The continuum systematically maps and structures a conceptual space with ideal types at the positive and negative ends of the continuum (Goertz, 2006).⁷

Ideal types are not well suited to inform policy-making. Clearly, a policy that uses an ideal type as a measure of success is doomed to fail. For example, German engagement in Afghanistan has clearly illustrated the limits of an intervention that set out to “fix a failed state” (Rotmann et al., 2014). Transferring this idea to the domain of concepts, Eriksen (2011, p.235) argues that “[i]nstead of using the extreme case as the standard, we need concepts which are closer to empirical reality”.

However, it is not advisable to get rid of ideal types in developing a useful concept of statehood and fragility. On the contrary, from a conceptual point of view it makes sense to construct ideal types as extreme cases that rarely or even never reflect empirical cases (Goertz, 2006). Ideal types merely form the ends of a continuum. Put simply, we should use not the ends but the continuum as a yardstick to set realistic expectations of policies.

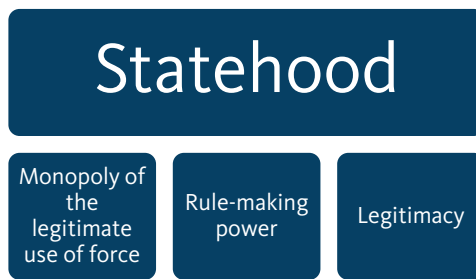
Core attributes of statehood

As a point of departure for defining statehood we draw on Weber’s definition of “state” as a “human community that (successfully) claims the *monopoly of the legitimate use of physical force* within a given territory” (Weber, 1946, p.78, emphasis original). The Weberian concept of state remains the most influential definition and many current disparate strands of the literature draw on it (Hay and Lister, 2006). The definition stands out in that it recurs to the means rather than the ends of the state (Weber, 1978). Although Weber acknowledges that states do have aims, he does not include in his definition a justification for why individuals should consent to the authority of the state or a certain purpose of the state. Rather than putting forward a moral or contractualist argument, Weber provides a description of the state free of normative implications. This makes his approach particularly useful in the empirical-analytical tradition.

Broadly following Weber and the recent discussion on the conceptualization of state fragility, statehood can be defined via the monopoly of the legitimate use of force, legitimacy, and rule-making powers. The following paragraphs briefly introduce these three attributes.

⁷ Ideal types are mental constructions that isolate the essential defining characteristics of a phenomenon without necessarily having a counterpart in reality. Empirical types, by contrast, are existing instances of a phenomenon such as specific states at specific points in time. Ideal types play an important role in typologies. For instance, if we defined “consolidated statehood” based on two attributes and measured each attribute on an ordinal binary scale, we could differentiate and rank four types of statehood. An empirical case fulfilling both attributes would be located near the positive end of the continuum. The definitional attributes of ideal types thus determine how we categorize cases. Ideal types are also theoretically important. Explicating the defining characteristics of statehood allows theoretical expectations to be derived about how it affects development cooperation. If, for example, we define consolidated statehood in terms of a well-developed infrastructure, we could hypothesize how accessibility affects the success of development cooperation. Theoretical arguments are ultimately founded on well-defined concepts.

Monopoly of the legitimate use of force



At its core, the possession of the monopoly of the legitimate use of physical force within a certain territory characterizes statehood. This includes the exercise of violence through state agents, the exclusive right to authorize the use of violence, and the protection of individuals and populations from illegitimate violence. In contrast to other types of authority, state authority is sovereign, i.e. the supreme and sole authority at the state level (Schliesky, 2004).⁸

If a state's monopoly of the legitimate use of force is challenged by non-state actors – such as secessionist movements, tribal authorities, rebel groups, drug cartels, or criminal gangs – statehood is challenged. Clearly, such challenges may vary in terms of their severity (from civil disobedience to criminal violence and civil war) and their spatial distribution.

Rule-making

The second element of statehood is the ability to make and implement rules. In this context, Migdal (1988, p.18) defines a state as an organization “that has the ability or authority to make and implement the binding rules for all the people as well as the parameters of rule making for other social organizations in a given territory, using force if necessary to have its way”. This requires the ability to draft rules as well as the administrative capacity to enforce them. States unable to protect property rights, uphold the rule of law, and enforce contracts are thus considered fragile.

Legitimacy

As a third definitional attribute, states are characterized by empirical legitimacy, i.e. a shared belief in the right of the state to use violence and to make and implement rules. The powers of a state have their roots in the population's belief in the rightfulness of a given authority and willingness to abide by state orders.⁹ If (parts of) the population cease to consent to the state's sovereignty and instead is loyal to non-state actors, statehood is deficient.

Alternative approaches to defining fragility

Before turning to the conceptual comparison, the following briefly reviews current debates on the definition of statehood and fragility.

States as service providers

The Weberian concept of statehood focuses on institutions. Another approach is to define states as service providers (Brinkerhoff et al., 2018). Rotberg (2004), for instance, categorizes states based on their ability to deliver basic public services such as healthcare, education, and the rule of law.

Including service delivery as a defining attribute of statehood broadens the view beyond the provision of protection through monopolizing the use of violence. In this view, states should contribute to liberties, opportunities, and human well-being. If the public perceives the state as effective and fair, this contributes to public consent for the state and increases stability.

However, defining statehood via service delivery also has three drawbacks. First, it is unclear what constitutes an adequate level of basic services. For instance, proponents of a minimal state argue that

⁸ This discussion focuses on internal sovereignty and does not discuss external sovereignty, i.e. freedom from external interference (see Krasner, 2007).

⁹ Lambach et al. (2015) argue against using legitimacy as an attribute of statehood because it is difficult to measure and hardly any empirical cases could be considered fully legitimate. Legitimacy is indeed hard to measure. However, from a methodological perspective, ideal types need no empirical counterpart to be useful.

property rights are so central that the state should not interfere with them (e.g. Nozick, 1974). In this view, taxation amounts to forced labour. Others attribute to the state a much stronger role. Rawls (2005), for instance, argues that states should ensure equal opportunities and redistribute services such as education, income, and healthcare. Making service delivery an attribute of statehood thus leads to (normative) discussions on distributive justice.

Second, including service delivery as a concept can lead to “conceptual stretching” (Sartori, 1970). Including the provision of services in a definition of statehood brings the concept of statehood closer to questions of governance (Börzel and Risse, 2018).

Third, indicators of good service delivery are at the same time indicators of development. Hence, making service delivery a constitutive attribute of statehood can lead to endogeneity in analyses of the relationship between fragility and development. Underdevelopment would be linked to fragility by definition. Empirical evidence shows, however, that even states with high poverty levels can be relatively peaceful (Putzel et al., 2012).

Elites, settlements, and the state

Another strand of literature takes a more dynamic and actor-oriented view of the phenomenon of state fragility. One approach in this strand of research has investigated how political elites shape institutions in order to govern. North et al. (2012) develop a theoretical framework to explain how the interplay between political, economic, and social actors can be organized as a limited access order to discourage the use of violence and foster cooperation in a society.

In a similar vein, Putzel and Di John (2012) regard the state as a political settlement. Conflict, contention, and bargaining between and among elite as well as non-elite groups form a set of power relations that is argued to be more important for development outcomes than formal institutions.

A related, more abstract approach by Bueno de Mesquita et al. (2003) differentiates political regimes according to the size of the selectorate, i.e. those who are formally entitled to choose the political leader(s), and the size of the winning coalition, i.e. those whose support is necessary for a leader to survive in office. They argue that leaders with larger winning coalitions concentrate on the provision of public rather than private services. The theory allows implications to be drawn on how interventions might foster political stability, economic growth, and democratization as well as on the effects of military intervention.

All these approaches share the view that there is a need to look beyond formal institutions to understand the drivers of fragility. They do not constitute a contradiction to the Weberian concepts of the state but, rather, widen the view taking into account societal groups and elites and their power relations.

Fragility as vulnerability

The understanding of fragility in development increasingly broadens the view to take into account multifaceted causes and symptoms of fragility. This includes non-political factors such as remoteness, natural disasters, and lack of market access. Moreover, the understanding has been extended to go beyond domestic factors and take into account the neighbourhood of countries (McCloughlin and Idris, 2016). This broadening has increasingly detached fragility from the notion of state as described above. Rather, it takes into account a large range of factors that might negatively affect the socio-economic, environmental, and political functioning of a country.

It is certainly true that non-political and non-domestic factors can affect the ability of a government to make and enforce rules on its territory. It is also true that a malfunctioning state has consequences that go beyond the political. However, acknowledging this causal embeddedness is different from including it in a definition. When discussing approaches that go beyond a narrow Weberian understanding of state fragility, it remains crucial to distinguish clearly between the causes, attributes, and consequences of fragility (Cammack et al., 2006).

Summary

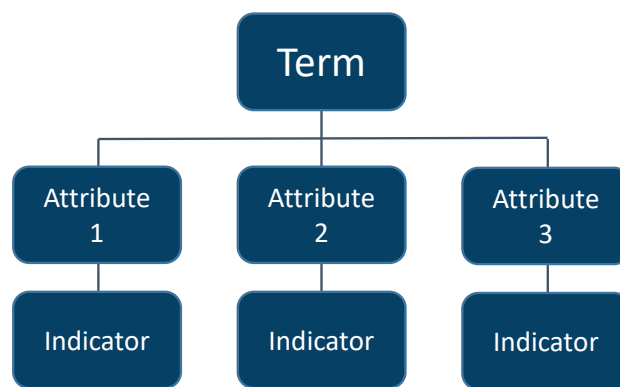
Fragility manifests itself in many ways, making a multidimensional concept of fragility necessary (Grävingholt et al., 2015; Gould and Pate, 2016; for an overview see Ferreira (2016)). We can consider state fragility as lying on the continuum between state collapse, i.e. the absence of a state, and consolidated statehood. Broadly following the Weberian understanding of state, fragile states can thus be characterized by a lack of upholding the monopoly of the legitimate use of force, limited rule-making ability, and restricted legitimacy.

Wider concepts of fragility also take into account the ability of the state to provide basic services as well as broadening the view to look beyond domestic factors. Attempts to further disaggregate the concept of fragility take into account subnational variations as well as power relations between elites and non-elites.

2.3 Method

The conceptual comparison involves two steps. First, we retrieve the design of the concepts under analysis. As illustrated in Figure 3, a concept involves three distinct levels (Goertz, 2006): the term describes the phenomenon to be defined. The attributes describe the constitutive essential properties of the object. The indicators operationalize the attributes. Appendix A includes graphic representations of the concepts.

Figure 3 Concept design



Source: author's illustration based on Goertz (2006)

In a second step, we evaluate the concepts against seven criteria described in Table 3 (Goertz, 2006; Mata and Ziaja, 2009).

Table 3 Criteria for conceptual comparison

Criterion	Description
Scope	the population of cases to which the concept can be applied
Extension	the range of empirical phenomena to be included in the concept
Multidimensionality	the number of attributes used to define fragility
Aggregation	the procedure to aggregate from the level of measurement to the overall assessment
Typologies	the number of empirical types derived from the concept
Predictive vs. status quo	whether a concept is predictive or captures the status quo
Data sources	the type and number of sources used for measurement

Source: own table partly based on Goertz (2006).

2.4 Concept comparison

This section compares four approaches to measuring state fragility using systematic criteria to facilitate the selection of suitable concepts:

- States of Fragility (SFR) provided by the OECD
- Constellations of State Fragility (CSF) provided by the DIE
- Crisis Early Warning (CEW) overseen by the Peace and Security division of the BMZ
- the fragility concept in the Catalogue of Criteria for Assessing Development Orientation (CoC) overseen by the Governance division of the BMZ.¹⁰

Concepts of fragility

States of Fragility

The OECD (2016, p.21) defines fragility as “the combination of exposure to risk and insufficient coping capacity of the state, system and/or communities to manage, absorb or mitigate those risks”. It identifies five dimensions of fragility: economic, environmental, political, security, and societal (OECD, 2016, p.22–25, 73, 152–159). Each dimension is defined by the two attributes “risks” and “coping capacities”. There are 51 indicators (OECD, 2016, p.265–270), which operationalize the attributes and are based on publicly available data.

In each dimension, the indicator scores are aggregated to two non-correlated principal components.¹¹ Aggregating to principal components leads to a data-driven weighting procedure. Each indicator is weighted according to the amount of new information it adds to the data. The two principal components in each dimension form a two-dimensional space. Each country in the sample is situated in this space, with countries similarly affected by fragility having similar positions. Emerging clusters of countries constitute specific types of fragility and are described qualitatively.¹²

The ten principal components (two in each of the five dimensions) are then aggregated by means of a second principal component analysis leading to another two-dimensional space. This second step identifies the most fragile countries and classifies them as “extremely fragile” or “fragile” based on two arbitrary thresholds.

Constellations of State Fragility

While the OECD’s concept of fragility is comparatively wide, Grävingholt et al. (2018) focus more closely on statehood. They distinguish three dimensions of fragility, based on types of state-society relations: authority, capacity, and legitimacy. The dimensions are operationalized with a total of ten indicators based on publicly available data and indices.

Grävingholt et al. (2015) employ finite mixture modelling to derive types from typical empirical constellations of the attributes of state fragility.¹³ To aggregate the indicators in each of the dimensions,

¹⁰ The conceptual comparison draws on the most recent methodology available. While the overall methodological approaches of the concepts under comparison were developed a few years ago, all concepts under analysis have since then been slightly adjusted. We incorporated all recent changes to the best of our knowledge.

¹¹ Principal component analysis is a procedure to reduce a set of correlated variables to a smaller number of non-correlated components. For instance, in the political dimension, the first principal component is mainly determined by the following indicators: voice and accountability, judicial constraints on executive power, perception of corruption, legislative constraints on executive power, and political terror. The second principal component of the political dimension is mostly influenced by the indicators of regime persistence and decentralized elections.

¹² For example, there are five types of fragility in the the political dimension: extreme political fragility, centralized state leadership fragility, high political fragility, moderate political fragility, and low political fragility (OECD, 2016, p.168).

¹³ Finite mixture modelling is a procedure to identify unobserved subgroups from a mixed distribution.

they apply a weakest link approach (see Goertz, 2006). This procedure results in scores for authority, capacity, and legitimacy in any given country-year within the sample.

After pooling all country-years in the sample, the authors apply mixture model clustering to derive constellations of state fragility from the data. The model specifications result in six constellations of state fragility: dysfunctional, low-authority, low-capacity, low-legitimacy, semi-functional, and well-functioning.

Crisis Early Warning

The BMZ's CEW, provided by GIGA, aims to detect emerging crises in order to allow the BMZ to take early action and thereby to mitigate the risk of further escalation of violence. It should make predictions of the tendency for violence and the phase of violent conflict, and is thus a predictive concept. The approach was updated at the end of 2018 through the Escalation Potential Analysis (ESKA) and we do not yet have precise information on the revised methodology.

The tendency for violence is defined using three dimensions: structural conflict factors, conflict-enhancing processes, and strategies of conflict resolution and the use of force. These three dimensions are each differentiated into three analysis sectors.¹⁴ These nine sectors each draw on between three and six multiple-choice questions to give an overall total of 35 indicators. Besides these nine analysis sectors, the violent conflict phase constitutes a tenth, and two more questions assess fragility in post-conflict situations.

Country experts answer the multiple-choice questions. The responses are mostly binary and totalled in each of the nine analysis sectors. These nine values are then weighted and again added up to derive the overall violence tendency. The potential for violent conflict, by contrast, is directly retrieved. Three broader categories for the violence tendency and four categories of potential conflict lead to 12 possible constellations, of which ten are eventually aggregated, resulting in a ternary typology of countries.

The results of CEW have concrete implications for the process of policy-making in the BMZ. If states are classified as “yellow” or “red” in CEW, peace and conflict assessments become compulsory for all bilateral assistance. Following this assessment, country portfolios must be adjusted to the security situation in a partner country. This includes an analysis of conflict dynamics, and an assessment of relevance and possible risks, as well as of the potential of development cooperation to do harm in fragile states.

Fragility index in the Catalogue of Criteria for Assessing Development Orientation

The CoC is used to assess the level of governance and development orientation. However, since 2016 some of the indicators have been used to measure state fragility. Since our focus is on state fragility rather than governance or development orientation, we focus on this newly developed measure.

The fragility index in the CoC conceptualises statehood, not fragility. The three dimensions of statehood are, in accordance with Grävingholt et al. (2015), authority, capacity and legitimacy. Each statehood-dimension draws on four to eight questions answered by the ministry employee responsible for the respective country (*LänderbearbeiterIn*). In total, there are 35 questions serving as indicators. The approach is currently being revised, and we do not yet have information on the revised methodology.

The questions are answered on a scale from 1 (worst) to 5 (best). For every attribute, the mean of the answers' values is calculated to achieve one aggregated score per dimension. To answer the questions as accurately as possible, the person responding is asked to take into account a list of public sources provided by the ministry. However, the use of the indices is neither compulsory nor systematically tracked.

¹⁴ The three analysis sectors defining structural conflict factors are: structural disparities, forms of settling conflict, and efficiency and legitimacy of political institutions. The three analysis sectors defining conflict-enhancing processes are: transformation and modernization processes, external influences, and collective perception of threat and experiences of violence. The three analysis sectors defining strategies of conflict solution and use of force are: tendencies to internal social polarisation, changes in the political strategies of single actors, and increasing use of force and violence.

If a country's mean is below 2.5 within any dimension, this dimension is considered a country's weakness. If all three dimensions score below 2.5 on average, the country is considered "a most fragile country". In internal documents, ratings by CEW and the level of governance and development orientation are included in an integrated display that allow the classifications provided by the two approaches to be compared.

Table 4 Overview of concepts of fragility

	Indices of fragility			
	DIE – Constellations of State Fragility	OECD – States of Fragility	BMZ – Fragility index in Catalogue of Criteria for Assessing Development Orientation	BMZ – Crisis Early Warning
Scope	Nation-states	Nation-states	Nation-states	Nation-states
Population	World 2005–2015	World 2016–2018	Partner countries 2016–2018	Partner countries 2001–2018
Attributes	3	5	3	3 (+2)
Indicators	10	50 (2016) 51 (2018)	17 ¹⁵	35 (Violence tendency) 2 (Potential for violent conflict) 2 (Fragility in post-conflict situations)
Aggregation	(1) Weakest link (2) Finite mixture modelling	(1 & 2) Two-stage principal component analysis (3) Threshold	(1) Mean of indicators (2) Threshold	(1) Sum (2) Weighted sum (3) Threshold
Sources	Publicly available data	Publicly available data	Experts	Experts
Weighting	Predefined	Determined by statistical model	Predefined	Predefined
Classification	Data driven (finite mixture modelling)	Data driven (Principal Component Analysis)	Previously determined	Previously determined
Number of resulting types of fragility	6	2 on aggregate level 3 to 5 in each dimension	1 on aggregate level	2

Source: own table.

¹⁵ The table shows only the indicators in the CoC relevant to measuring the statehood dimension, not all indicators of the CoC.

Conceptual comparison

The following compares the concepts under analysis in terms of their scope, extension, multidimensionality, aggregation, typologies, data sources, and focus (Goertz, 2006). The comparison systematically carves out similarities and differences among concepts. It also includes a discussion of advantages and disadvantages. Table 4 summarizes the results.

Scope of application

The scope describes the population of observations to which the concept can be applied. Ideally, all cases to which a concept applies are also coded with data (Goertz, 2006).

CSF and SFR observe around 170 countries worldwide. The fragility measure in the CoC and CEW only look at the BMZ's partner countries in bilateral development cooperation. In addition, CEW takes into account countries of special interest for the BMZ when asked to do so. CSF is available for the period 2005 to 2015, SFR for 2016–2018, the fragility component in the CoC for 2016–2018, and CEW from 2001 to 2018.

In all concepts, the unit of observation is the nation-state. The concepts do not explicitly assess subnational variance in respect of fragility or spillover effects. The OECD (2018, p.87), for instance, explicitly acknowledges that “the framework is unable to capture [...] subnational and micro-level drivers that indicate pockets of fragility within borders”. However, the degree of fragility can vary greatly within countries.

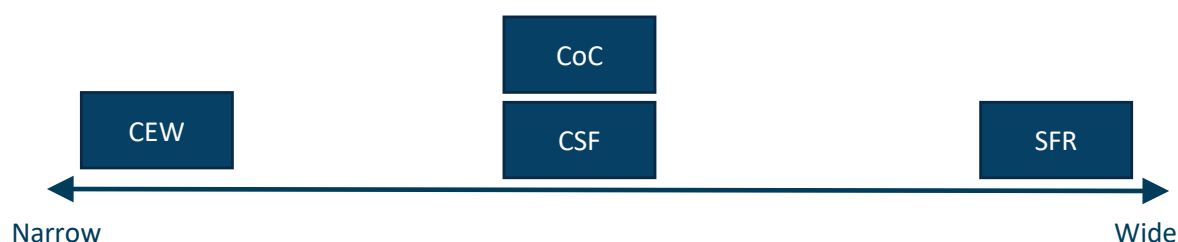
As some types of development cooperation are locally confined, a national measure of fragility might be of only limited use. However, limited public data are available on the subnational and regional level, especially in areas with limited statehood (Stollenwerk, 2017, p.9). Although there is a broad trend of data disaggregation in the social sciences, there are as yet no global data on the subnational variance of fragility. The most encompassing attempt to date is by Lee and Zhang (2016). Moreover, there are some approaches for similar phenomena such as conflict (Raleigh et al., 2010; Sundberg and Melander, 2013; Trinn and Wencker, 2018) or risk (De Groeve et al., 2015).

Extension: narrow vs. wide concepts

The concepts under analysis have different empirical extensions. On the one hand, concepts should be broad enough to cover all empirical phenomena that are instances of fragility. On the other, concepts that are too broad are of only limited use in causal analysis due to possible endogeneity. The causal relationship between state fragility and intrastate conflict, for instance, cannot be investigated if intrastate conflict is a constitutive attribute of fragility (see Vreeland, 2008). The same might apply to analyses on the link between fragility and development in cases where both concepts draw on the provision of basic services. Consequently, excessively broad concepts make it difficult to study relationships between phenomena.

SFR is the broadest concept in the analysis because it does not exclusively cover attributes pertaining to statehood. For instance, it includes environmental risks such as exposure to natural hazards. The CEW is the narrowest concept due to its exclusive focus on violence.

Some uses of SFR suffer from endogeneity. For instance, the OECD's fragility measure has been linked to the achievement of SDGs (OECD, 2018; UN Foundation, 2018). However, both concepts share indicators. Hence, it is no surprise that fragile contexts fare poorly with regard to SDG achievement. Moreover, the endogeneity of both concepts precludes an analysis of causal influence of fragility, as understood by SFR, and SDG achievement. This affects dimensions of sustainable development such as poverty, (gender) inequality, governance, education, health, violence, and the effects of disaster. Hence, if it is argued that more must be done on fragility to achieve the SDGs, this is not only because fragility negatively affects SDGs but also because both concepts partly refer to the same things. Consequently, causal analyses investigating the effect of fragility on SDG achievement can benefit from narrower concepts of fragility such as those based on a Weberian understanding.



One- vs. multidimensional concepts

We have seen that concepts should be broad enough to cover all empirical phenomena that are instances of the definiendum. At the same time, however, they should be sufficiently differentiated not to blur important differences between various types of fragility. Multidimensional concepts offer a solution as they can combine a sufficiently large extension while mapping differences through internal differentiation.¹⁶

All concepts rely on more than one defining attribute. SFR uses five, while all other concepts draw on three.¹⁷ SFR is thus the broadest and most multidimensional concept. It is also the only one that uses the same indicators in multiple dimensions. This is not inherently problematic, especially since the resultant double weighting of indicators is statistically controlled for (OECD, 2018, p.267). However, when comparing the risks and coping capacities in different dimensions of fragility, it must be kept in mind that some are related by design.

Aggregation

Multidimensionality can mitigate the fuzziness of broad concepts. All multidimensional concepts need to be aggregated to derive the overall value of state fragility. While taking the arithmetical mean of indicators is a popular and common choice of aggregation, other methods such as the weakest link, the best shot, or logical combinations are possible (Goertz, 2006). Table 5 illustrates three methods of aggregation for hypothetical values.

More recently, statistical approaches to aggregation have grown in popularity. Here, the empirical distribution of values determines the weighting of the individual indicators and/or the final classification.

Table 5 Methods of aggregation

Mean			Weakest link			Best shot		
30			5			60		
60	5	25	60	5	25	60	5	25

Source: own table.

We find that all concepts use different approaches to aggregation. CSF combines the weakest link with a data-driven approach. In each of the three dimensions, the indicator with the lowest value represents the aggregated score. Consequently, indicators within a dimension cannot act as substitutes. Countries scoring low in one indicator also score low in this dimension, even though they might possibly have high ratings for other indicators in the same dimension. In a second step of aggregation, CSF applies finite mixture modelling to identify six empirical constellations of fragility.

SFR applies a purely data-driven approach to aggregation. It applies two principal component analyses to aggregate from indicator to attribute level and from attribute level to the final fragility index. The first stage

¹⁶ Multidimensionality and extension are not the same. CEW, for instance, is narrow as it focuses on conflict escalation but uses three attributes for measurement, namely structural conflict factors, conflict-enhancing processes, and strategies of conflict resolution and use of force.

¹⁷ The CEW has two more dimensions that do not, however, contribute to the measurement of the violence tendency.

results in three to five types of fragility per dimension. At its highest level of aggregation, SFR puts forward a one-dimensional scale of “levels of fragility” using an arbitrary threshold to differentiate extreme fragility from fragile states and the rest of the world. Even though SFR depicts a one-dimensional scale of fragility on its highest level, the OECD advocates against its use (OECD, 2018, p.82).

CEW and the CoC do not use data-driven aggregation. CEW uses weighted sums to arrive at an overall value for violence tendency. The CoC takes the mean of all indicators for each dimension to result in one value per dimension.

The choice of a method of aggregation should be derived from the understanding of the underlying phenomenon. Since all concepts in our comparison refer to similar phenomena, it is surprising that they make use of very different approaches of aggregation. The weakest link approach, for instance, implies that each indicator is a necessary condition for the respective dimension. More specifically, it implies that, for example, access to drinking water and protection from disease cannot compensate for a lack of primary school enrolment. As an example, in 2015 the United States was classified as semi-functional due to low scoring of single indicators in all three dimensions.

Table 6 provides a further illustration. The weakest link approach causes Brazil, Mexico, and South Africa to score low on the authority level due to a large number of homicides. All three are classified as low-authority countries. This rating does not take into account the fact that there were no deaths due to armed conflict and that the monopoly of violence rating of Brazil and South Africa is in the mid-range compared to the other countries included in the Bertelsmann Transformation Index (BTI).

Table 6 Illustration of weakest link approach to aggregation

	Authority	Battle-related deaths	Homicides	Monopoly of violence (BTI)
Brazil	0.26	1.00	0.26	0.70
Mexico	0.36	1.00	0.36	0.50
South Africa	0.21	1.00	0.21	0.70

Source: own table based on 2015 data from Grävingholt et al. (2018).

Using the mean or sum, on the other hand, allows for substitution across indicators. The CoC, for instance, uses the very same definitional attributes of authority, capacity, and legitimacy as the CSF, but aggregates indicators based on their arithmetical mean. Thus, although both concepts are semantically equivalent on the level of attributes, they reflect a fundamentally different understanding of fragility.

The use of statistical methods of aggregation has the practical consequence that a change in or addition of data can affect the whole classification because the weights and/or classes are a function of the distribution of all indicators. As a practical consequence, countries might be classified differently although the underlying indicator values do not change. When new data are added, this might also happen retrospectively. One possibility for arriving at a stable classification, however, is to fix parameter values at a value determined for one specific sample.

Data-driven classifications thus adjust to the empirical reality. This could be seen as an advantage, as data-driven approaches reduce the empirical complexity of multidimensional data to a manageable level while preserving as much information from the data as possible. However, data-driven aggregation procedures are often less transparent than the use of thresholds. Perhaps even more importantly, the weights change with the data and are not derived from theoretical reasoning. It can be difficult for practitioners to retrace the steps in aggregation and the relative importance of the underlying indicators.

Where no statistical methods are used, classification often draws on thresholds. This is the case in the two concepts used by the BMZ. The categories are predefined and thresholds are chosen quite arbitrarily.

Typologies

While all concepts are multidimensional in drawing on multiple attributes, only CSF and SFR explicitly derive types from their concept. SFR stands out in that it combines a data-driven approach with qualitative reasoning by experts. The results of clustering within the first stage are qualitatively described in respect of their most distinctive features and ranked on a six-level scale (OECD, 2018, p.272–278). With regard to the overall level of fragility, the OECD applies two arbitrary thresholds to distinguish three levels of fragility on a one-dimensional scale.

Typologies using thresholds usually divide the empirical space into mutually exclusive and exhaustive categories. This means that any one country can only belong to one category at a single point in time and that all countries belong to one (and only one) category. CSF uses non-exclusive categories by providing a probability of membership for different categories. SFR offers a middle way in providing the position of cases in a multidimensional space. These, however, are then coded nominally (by their predominant qualities) and ordinally (by level of fragility) by experts into mutually exclusive and exhaustive categories.

The types of concept are as follows:

Table 7 Types of fragility

DIE – Constellations of State Fragility	OECD – States of Fragility	BMZ – Fragility index in Catalogue of Criteria	BMZ – Crisis Early Warning
Dysfunctional Low-authority Low-capacity Low-legitimacy Semi-functional Well-functioning	Extreme fragility Fragile 3–5 qualitative types based on clustering in each dimension	Most fragile states	Low violence tendency (green) Heightened violence tendency (yellow) Acute violence tendency (red)

Source: own table.

Predictive concepts vs. concepts focusing on the status quo

A concept can either focus on the status quo or be predictive. The former assesses a country's state of fragility in a given year. Predictive indices, on the other hand, try to assess the likelihood of future events.

CSF and the CoC are status-quo oriented. They aim to measure the actual level of fragility understood as an actual state, rather than the likelihood of a future state. By contrast, CEW is explicitly directed at identifying the risk of escalation of political conflict and is thus clearly predictive.

The SFR approach is ambiguous. The notions of high risk and low coping capacity refer to an unrealized reaction to external or domestic shocks. In line with this, the authors argue that each indicator represents “an increased likelihood or impact of a negative event” (OECD, 2016, p.150). The framework explicitly selects indicators that “alter [...] likelihood or impact *ex ante*” (ibid., p.151). On the other hand, the authors state that “the fragility framework should not be interpreted as predictive” (ibid., p.74). This leads to an understanding of fragility as a vulnerability or a possibility rather than something that has occurred. It is in this way that the definition of fragility by SFR departs significantly from the Weberian notion of statehood.

Data sources

The concepts under analysis draw on different types of data. CSF and SFR rely on publicly available sources such as indices. CEW relies on multiple-choice questions answered by one expert per country. These answers are assigned values according to their contribution to violence tendencies. The CoC similarly relies on questions answered by country experts. Each question is rated on a scale of 1 to 5. Comments on each rating substantiate the assessments. Thus, while CSF and SFR rely on quantitative data from publicly available data, the BMZ approaches rely on quantitative and qualitative data based on expert codings.

In addition, if expert coding is commissioned by the data user, the timeliness of the data is usually better. The timeliness of publicly available data, by contrast, is subject to those who provide them. Although extrapolation of data is possible, the validity of the generated data is highly questionable.

2.5 Results and recommendations

Our conceptual comparison has shown that no two concepts are alike. Even semantically similar definitions can have diverse concept designs. This has concrete empirical implications in terms of which countries are considered fragile. This is crucial since country classifications influence allocation and programming decisions (see Chapter 3). The examination of concepts of fragility in this evaluation is hence not an exercise for some abstract academic research interest, but linked to one of the most important questions in development cooperation and policy-making in general: on what basis are scarce resources distributed?

We summarize the conceptual comparison as follows. First, we sum up the conceptual comparison. Second, we discuss how the current use of concepts can be improved in the short term. Third, we provide some ideas on possible improvements in the longer term. Finally, the main points are summarized in a single recommendation.

Summary of conceptual comparison

Among the concepts in our comparison, there is no single concept that we would endorse unequivocally. Based on our comparison criteria, however, we consider that Constellations of State Fragility (CSF) is of greatest use when combined with more timely data.

One advantage that CSF shares with the OECD's States of Fragility over the BMZ's Catalogue of Criteria and Crisis Early Warning is that it relies on publicly available data. This makes CSF (and SFR) more transparent and less dependent on the assessment of individual experts. Public availability puts data under public scrutiny, ultimately contributing to greater measurement equivalence across countries and years. This is not to say that concepts should exclusively be founded on quantitative indicators rather than expert codings. On the contrary, it is problematic to measure some phenomena by numbers alone. For instance, state legitimacy – the intersubjective belief in the rightfulness of a given order – is a highly complex and subjective feature of statehood that is difficult to put into numbers. Moreover, concepts such as lack of state authority might be empirically instantiated in myriad manifestations such as riots, terrorist events, intrastate conflict, crime or a *coup d'état*. The changing nature of social reality might not be captured by a predefined set of quantitative indicators. In addition, if expert coding is commissioned by the data user, the timeliness of the data can be much better than of publicly available data. Although extrapolation of data is possible, the validity of the generated data is highly questionable. Consequently, the solution is not to entirely remove judgemental elements from the measurement of state fragility but, rather, to ensure their reliability and validity. A mix of quantitative and qualitative indicators appears best for concepts on political phenomena such as fragility. While both CSF and SFR draw on a mix of qualitative and quantitative indicators, CEW and the CoC rely on expert coding.¹⁸

A second advantage of CSF and SFR over the CoC and CEW is that, while the latter two mostly focus on partner countries, the former include worldwide data. This might appear to be less important for decision-making on German development cooperation. However, a view beyond partner countries allows results to be put in a broader context, facilitates comparison with other (e.g. potential future or past partner) countries, and increases the number of cases for analytical use. For instance, worldwide data coverage allows global trends to be discerned from trends that only affect partner countries. This, in turn, allows the effects of development cooperation to be more precisely estimated by comparing countries with German development cooperation to those without it. Consequently, even though collection of data on non-partner

¹⁸ A more experimental approach to retrieving timely data is to quickly process data from public sources through, e.g., text-mining (Brandt et al., 2011; Chadeaux, 2014; Chiba et al., 2017; Schrodtt and Gerner, 2000).

countries might not be of core interest to political decision-makers in the BMZ, particularly in the light of limited resources, such data could improve our understanding of fragility.

Our conclusion regarding the primacy of a certain method of aggregation is not so clear cut. The data-driven methods of aggregation and classification by CSF and SFR provide a means to optimally reduce empirical complexity to a manageable typology while preserving as much information as possible. On the other hand, statistical properties might be less transparent to practitioners than clear-cut criteria such as thresholds. Moreover, aggregation based on constant and predefined criteria ensures the stability of resulting types.

When it comes to deciding between CSF and SFR, our greatest reservation over the latter is its very wide empirical scope. Broad concepts of fragility face the danger of becoming “catch-all” terms (see Lindemann, 2014). If we conceive state fragility to be an overarching term describing a broad range of diverse deficiencies in the political, economic, societal, environmental, and possibly other dimensions, the term loses analytical value. Clearly, the question of what constitutes “too broad” is an empirical question up for debate. Few would question the pivotal role of personal security in defining fragility. The role of service delivery, however, is already disputed, with proponents of a minimal state advocating a very restricted set of state functions and contemporary liberalism endorsing redistributive state intervention. Greatly extending the scope beyond statehood, however, leads to catch-all terms. Obviously, state fragility is not the only phenomenon of interest for programming and project design. And SFR looks at dimensions that are certainly important for the success of development cooperation. However, it remains doubtful whether all need to be lumped together in a single concept of fragility that finally aggregates to a ternary scale. CSF, with its narrower extension than SFR, allows the role of statehood to be taken specifically into account. It allows for a greater number of phenomena to be investigated vis-à-vis fragility. This includes, for instance, many indicators underlying the SDGs. Attributes or indicators that are part of a concept cannot be studied in a causal analysis including the concept due to endogeneity. Consequently, broader concepts restrict the analytical potential.

In summary, no single concept fulfils all criteria of our conceptual comparison. Concepts with a wide scope and suitable extension, and founded on qualitative and quantitative public data and explicit concept design, are readily available but usually do not provide timely data. Timely data are, however, available through the existing reporting system currently used in the BMZ. As it would be inefficient to develop yet another approach to defining and measuring fragility, the challenge lies in combining existing approaches. This requires a standardized process of combining and weighting information through mixed- and multi-method approaches (Gibler, 2017). For instance, quantitative and qualitative assessments of similar phenomena could be systematically compared across approaches to identify whether or not they agree. As a step further, data from different sources could be combined, for example to increase the precision of imputed quantitative data by incorporating evidence based on more timely expert assessments. In summary, rather than developing a new approach from scratch, a combination of approaches seems to be expedient in order to exploit the potential of existing approaches.

How to improve the use of concepts for policy-making

Every country is unique. Policy-making in development cooperation requires heuristics to cope with this empirical complexity. In particular, political decision-makers need to assess state fragility without delving into the complexities of individual cases. In the end, uniqueness should not preclude meaningful comparisons between countries.

Indices of state fragility provide a means to reduce empirical complexity. They allow comparisons between states on a single, possibly multidimensional scale. Without well-crafted heuristics, a lack of guidance on what information should be considered relevant combined with time constraints might lead decision-makers to over-rely on volatile subjective reasoning. The alternative would be to rely on ad hoc criteria that are prone to psychological biases (Kahneman, 2012; Koehler, 2007; Tversky and Kahneman, 1983). Consequently, concepts such as those under review in this chapter provide useful heuristics to be systematically integrated in strategic management. The mandatory use of indices as currently practised in the development of country strategies thus promises to improve decision-making. At the same time, policy decisions should not be exclusively based on the results of indices. Portfolio management, programming,

and implementing development cooperation requires expert knowledge on countries and sectors. An idiographic, in-depth understanding complements the more generalized comparative approach provided by indices and typologies.

With reduced complexity comes the danger of oversimplification. Most of the concepts presented above provide an easily manageable number of resultant types and scales to assess and compare states. The OECD's States of Fragility, for instance, aggregates 49 indicators into a one-dimensional ternary scale. In a similar vein, the BMZ's CEW makes use of 35 indicators in order to come up with a ternary measure of escalation potential. Such highly aggregated measures are easy to apply. Moreover, clear-cut typologies that assign cases to a single category greatly simplify our view of differences between states. However, it is doubtful that highly aggregated scales and clear-cut typologies adequately reflect the underlying empirical complexity in fragile states (see also Wagner and Sattelberger, 2017).

One approach to reduce the danger of oversimplification is to present results of indices more comprehensively. This would put to use the wealth of information provided by multidimensional concepts rather than drawing on the highest level of aggregation for the sake of practical value. The use of arbitrary thresholds, for instance, conceals the continuous nature of (most) underlying criteria. A presentation that takes into account the multilevel nature of concepts, i.e. one that includes indicators and overall results, would allow for a differentiated yet focused view. One possible way to present such results concisely is by a visual representation. This could include the individual indicators and the applied weighting mechanism. It would be even better to present data interactively, allowing users to specify component indicators, weights, and methods of aggregation. This would allow the impact of concept design on the classification of cases to be assessed.

Moreover, in many cases, assigning cases to types is not as clear-cut as typologies convey. For instance, a country classified as non-fragile might change to being classified as fragile one year later because it barely passed a certain threshold. Since the underlying measurement is uncertain, this change in category might be entirely due to chance. The arbitrariness that arises from selecting clear-cut thresholds for continuous phenomena can be mitigated by explicitly modelling and presenting the uncertainty of assigning certain cases. Instead of using crisp sets in which cases are either in or outside a certain category, fuzzy sets assign cases a probability of membership. Such data are readily available in existing indices such as CSF and SFR and can be useful in understanding state fragility.

Moreover, the uncertainty of measurement should be explicitly modelled. Quantitative indicators convey a certainty that seldom exists. In fact, measurement is never perfect. King et al. (2004) provide methods for estimating the uncertainty of indicators. Coppedge et al. (2016) explicitly model and report uncertainty for a concept of democracy. One way to take account of the uncertainty of measurement is to rely on distributions rather than pseudo-exact point estimates. Reporting of confidence intervals or posterior densities allows policymakers to visually assess the certainty of an assessment and could be used to decide on further data collection on a certain case.

How to improve existing concepts

As shown above, recent improvements in concept designs allow for adaptations in practice that can be realised in the short term. The following proposes three improvements that could be implemented in the longer term.

First, taking into account how fragility spreads across borders could improve the understanding of fragility. Conflict, and thus instability, spreads among countries (Buhaug and Gleditsch, 2008; Gleditsch, 2002; Grechyna, 2018). Moreover, there is evidence that mechanisms of diffusion are conditional on state capacity (Braithwaite, 2010). The concepts under review do not – except for a single indicator in CEW – take account of the role of the fragility of neighbouring countries. DFID (2016), for instance, explicitly mentions the instability of neighbouring states in its fragile states methodology. Based on existing evidence on the spatial effects of political instability, measures of fragility would benefit from taking into account processes of spatial diffusion. Consequently, country assessments could widen the view beyond the nation-state to specifically take account of how fragility spreads across borders.

Second, spatial disaggregation of measures of fragility should improve our understanding because the level of fragility often varies within states, specifically in developing states. The recent developments in measuring fragility have either broadened the view to consider more and more dimensions (as in SFR) or disaggregated the concept along its constitutive dimensions (as in CSF). In Chapter 5, we present one approach to making use of geographically disaggregated data to study state authority on a subnational basis.

A third approach is to apply the notion of fragility and its constitutive dimensions to the regime, administration, and government of a state. This “third generation of state theory” (Mazzuca, 2017) could further enhance the added value of fragility indices in context-sensitive programming.

In summary, we make the following recommendation:

Recommendation 1

Based on a systematic conceptual comparison of four approaches to defining and measuring fragility, this evaluation recommends that the BMZ continue to use fragility indices for strategic portfolio management. Furthermore, the BMZ should continue to constantly evaluate and possibly improve its definitions and measurement of fragility on the basis of clearly defined criteria. This should include an assessment of the extent to which existing approaches could be combined.

As part of the current revision of approaches to measuring fragility, the BMZ should ensure that the results of indices are presented to political decision-makers more comprehensively in order to make underlying concept designs transparent and to do justice to the multidimensional nature of fragility. To achieve this, the BMZ should adopt a visual representation of concepts that includes indicators and the applied weighting mechanism. Second, the BMZ should include the uncertainty of classification and measurement in presentations of fragility scores.

Moreover, with a view to possible future revisions, the BMZ should further disaggregate and contextualize its measurement of state fragility in order to further increase the value of indices for the context-sensitive management of development cooperation. Further disaggregation could be achieved by reporting information for lower-level geographical units or more regularly and by differentiating fragility between different actors within partner countries. Contextualization could be achieved by taking account of possible spillover effects of fragility across borders.

3. STRATEGIES

3.1 Introduction

Development cooperation in fragile states is considered a key tool of German civil crisis prevention (Bundesregierung, 2004, 2012, 2017). Consequently, it is no surprise that the increasing attention given to fragile contexts has found expression in a growing number of guidelines on context-sensitive implementation of development cooperation. This chapter examines whether strategy papers released by the BMZ since 2000 have repercussions on German ODA commitments.

For development cooperation to be successful, it needs to be aligned to fragile contexts (Manning and Trzeciak-Duval, 2010; Zoellick, 2008). Strategy papers contribute to this alignment by providing an operational framework for policymakers to adapt development cooperation in fragile contexts. They provide guiding norms and principles for operations, define policy goals, set action areas, propose appropriate instruments, and explicate assumptions underlying development cooperation in fragile contexts. Beyond influencing policies and politics, strategy papers signal to the public and partner countries the German government's preferences concerning fragility, state-building, and crisis prevention.

Strategies can also help to reduce fragmentation between donors and instruments by supporting policy coordination, complementarity, and coherence. Fragmentation can be particularly problematic in fragile contexts because of the limited coordinating capacity of partner countries.

However, some practitioners and researchers argue that official policies as embodied in strategies have no effect on practice. Furness (2014) argues that, for the EU, development cooperation in fragile states reveals a “policy-operations” gap, i.e. differences between written policies and development cooperation on the ground. He argues that rapidly changing conditions require greater flexibility to adjust to local contexts, whereas strategies were long term and vague. Practitioners follow a similar reasoning when they claim that strategies are especially hard to implement in fragile contexts. The volatility and complexity of fragile contexts clearly impedes the detailed planning of interventions.

We use three steps to investigate whether allocation patterns reflect the content of strategies. First, we describe changes and continuities in the BMZ's strategy towards fragile states. Second, we extract testable implications from strategy documents that evolved after critical junctures. And third, we test empirically to what extent ODA commitments¹⁹ match these testable implications.

The analysis contributes to the transparency of German development cooperation. It responds to the criticism voiced by the non-governmental Advisory Board for Civilian Crisis Prevention (*Beirat für zivile Krisenprävention*) that the Federal Government's crisis-prevention policy lacks transparency and strategic communication (Beirat Zivile Krisenprävention, 2008, 2014; Nachtwei, 2010). In a similar vein, an analysis of spending patterns contributes to implementing the International Aid Transparency Initiative (IATI), of which Germany was a founding member in 2008. Last but not least, since the BMZ devotes considerable resources to the formulation of strategies, it has a legitimate interest in understanding to what extent changes in strategy have tangible consequences. This is particularly important in the complex institutional set-up of German development cooperation, with a specialized ministry and implementing agencies, while foreign policy and humanitarian assistance sit in a different ministry. To the best of our knowledge, there is no other study that investigates the empirical implications of strategies. Hence, this study provides relevant information for all BMZ staff working in the field of fragility as well as on relevant country desks.

¹⁹ Commitments up to 2017 data are defined by the OECD DAC as “those flows to countries and territories on the DAC List of ODA Recipients and to multilateral institutions which are: i. provided by official agencies, including state and local governments, or by their executive agencies; and ii. each transaction of which: a) is administered with the promotion of the economic development and welfare of developing countries as its main objective; and b) is concessional in character and conveys a grant element of at least 25 per cent (calculated at a rate of discount of 10 per cent)” (OECD, n.d.).

3.2 Method

The description of the evolution of Germany's strategy towards fragile states is organized using critical junctures that provided the opportunity for development policy change. To trace strategic and organizational developments in the fragile states agenda, we draw on eight key strategy papers published between 2004 and 2017 by the German Federal Government and the BMZ (see Table 8). In our empirical analysis, we focus on the 2005, 2007, 2009, and 2013 strategies, which operationalize the more abstract guidelines published by the Federal Government. While strategies released by the Federal Government provide an abstract and interministerial regulatory framework, BMZ strategies more specifically detail concrete policies for development cooperation. To assess whether allocations match strategies, we derive empirical implications from the papers, operationalize them based on commitments as recorded in data from the OECD DAC's Creditor Reporting System (CRS), and finally test them empirically.²⁰ Hence, we use allocations as a proxy for changing priorities in German development cooperation. We classify countries as fragile based on CEW as this is the classification used by the BMZ.²¹

Table 8 Key strategy papers relating to state fragility published by the German Federal Government and the BMZ, 2004–2017

Year	Fed. Government	BMZ Peace and Security division	BMZ Governance division
2004	2004 “Civilian crisis prevention, conflict transformation and peace consolidation”	2005 “Cross-Sectoral Strategy for Crisis Prevention, Conflict Transformation and Peace Building in German Development Cooperation”	
2005			
2006			
2007			2007 “Development-Oriented Transformation in Conditions of Fragile Statehood and Poor Government Performance” ²²
2008			
2009			
2010			
2011			
2012	2012 “Interministerial Guidelines for a Coherent Policy of the German Federal Government in Fragile States”	2013 “Development for Peace and Security. Development Policy in the Context of Conflict, Fragility and Violence”	
2013			
2014			
2015			
2016			2013 “Strategy on Transitional Development Assistance: Strengthening Resilience – Shaping Transition”
2017	2017 Guidelines “Preventing Crisis, Resolving Conflicts, Building Peace”		

Source: own table.

²⁰ For the full list of sources see the online appendix.

²¹ Since the data are only available for the years after 2006, we impute values for the years 2004 to 2006 using UCDP data on armed conflict as well as CSF data (see online appendix for more details).

²² The 2009 document “Promotion of Good Governance in German Development Policy” is another comprehensive strategy paper by the BMZ, which is valid for all countries. Since it refers to the 2007 paper regarding fragility issues mentioned here, it is not listed in Table 8.

Critical junctures

Germany's development policy adapts to international and domestic developments and is subject to constant change. The opportunity for substantial policy changes, however, is greatest during critical junctures, i.e. "relatively short periods of time during which there is a substantially heightened probability that agents' choices will affect the outcome of interest" (Capoccia and Kelemen, 2007, p.348). Critical junctures allow for the establishment of new institutional arrangements in moments of uncertainty (Collier and Munck, 2017). Uncertainty stems from external events or shocks that increase the demand for or reveal the incapacity of the existing institutional framework to deal with changed conditions. In line with this reasoning, Faust et al. (2013, p.14) state that "the important steps forward towards an agreed-upon international agenda for addressing state fragility have become possible in the immediate aftermath of severe crises".

In the following, critical junctures function as a heuristic device. Based on the above reasoning, we can assume that it is most likely to be during critical junctures that policymakers revise strategies, alter policy-making processes, and adapt aid allocation patterns towards fragile states. By contrast, path dependency, i.e. the influence of past policy decisions on current and future policy, governs the comparatively longer periods between critical junctures.

Although critical junctures following international crises provide leeway for change, policy-making does not happen in a vacuum. Preceding conditions, most importantly existing domestic institutions and the international fragile states agenda, and policymakers' individual preferences, influence policy-making and implementation processes (Capoccia, 2016; Collier and Munck, 2017). Consequently, the following discussion sheds light on the domestic and international environment preceding critical junctures.

Empirical analysis

To evaluate the coherence between strategies and country portfolios, we first derive empirical implications from analysed documents. Subsequently, we test these empirical implications based on German bilateral aid commitments. Commitments reflect decisions better than disbursements (Berthelemy, 2006; Boussalis and Peiffer, 2011). To operationalize strategy implementation, we identify the type of development cooperation mentioned in the strategy by identifying the closest possible match in the CRS (e.g. a certain sector and/or recipient). We restrict our data to committed grants and loans. The online appendix provides a detailed summary of all operationalizations.

To test strategy implementation, we combine a descriptive and an inferential approach. The descriptive analysis compares, for each empirical implication separately, whether total or relative allocations to fragile countries changed after the implementation of the respective strategy. The analytical approach applies a fixed-effect regression analysis on the panel data to rule out alternative explanations for changes in allocations (see online appendix for more information). We assume an implementation gap of two years for TC and four years for FC.²³

We then combine the results qualitatively, taking into account the descriptive and inferential evidence. A team of researchers at DEval rated the empirical results using the Likert scale (see Table 9). We then aggregated the evidence for each critical juncture to come to an overall assessment.

²³ Due to data constraints, we only apply an implementation gap of one year for the 2013 strategy.

Table 9 Evidence aggregation

Strong evidence of allocation patterns opposite to implication	Weak evidence of allocation patterns opposite to implication	No clear evidence	Weak evidence of allocation patterns in line with implication	Strong evidence of allocation patterns in line with implication
--	-	O	+	++

Source: own table.

Structure of the German development cooperation system

To delineate the structure and processes of the German bilateral development cooperation system, the evaluation team reviewed internal and public documents describing the decision-making procedures. The review placed particular emphasis on those steps in the process that resort to strategies on development cooperation in fragile contexts or fragility assessments. In order to tailor recommendations to the needs of the BMZ, GIZ, and KfW, our suggestions refer to procedures as described in the BMZ's Joint Procedural Reform (GVR).

3.3 Germany's strategy towards fragile states

Germany strives to align its development strategy in fragile contexts to the international development agenda as laid out in the Fragile State Principles (OECD and OECD DAC, 2007) and the New Deal for International Engagement in Fragile States (Bundesregierung, 2012, p.3; International Dialogue on Peacebuilding and Statebuilding, 2011). Although harmonization of donor activities in fragile contexts has not yet been fully achieved (OECD DAC, 2011), it is high on the international development agenda (High Level Forum, 2005). In order to do justice to the fact that Germany develops its strategy as part of the international community and based on the insight that the agenda on state fragility probably changed most significantly following serious international crises, we structure the following discussion based on two critical junctures of international importance: the aftermath of the Al-Qaeda attacks in 2001 and the "Arab Spring" of 2012.

3.3.1 The post-9/11 juncture: 2004 Action Plan

The 2004 Interministerial Action Plan "Civilian crisis prevention, conflict transformation and peace consolidation" can be considered Germany's first joined-up strategy targeting a distinct group of "fragile states", even though the notion of fragility is absent from the report. Germany formulated this strategy relatively late compared to other Western donor countries, e.g. the US and the UK (Cammack et al., 2006).

The 2004 Interministerial Action Plan is one expression of fundamental institutional and policy change after 2001. The Al-Qaeda attacks on 11 September were a "critical turning point" (Klotzle, 2006), linking state fragility and international security. The attacks raised domestic security demands to address the perceived terrorist threat emanating from failed states in the "Global South" (François and Sud, 2006; Furness, 2014). Considered a national security issue, failed or fragile states became a geopolitical priority for Western states after 9/11 (Rotberg, 2002).²⁴ After this, it was discussed by almost all major development agencies (Bueger and Bethke, 2014). In a similar vein, Putzel and Di John (2012, p.5) state that "[s]ince the attacks on the World Trade Centre on 11 September 2001, the problems of weak states [...] have been a growing concern among international development agencies".

²⁴ In February 2002, German chancellor Gerhard Schröder stated: "Security promotes development, but development also promotes security. The beginning of the 21st century has been overshadowed by a horrible tragedy, the destruction of the Twin Towers. But I think that a new international awareness and, as such, new international cooperation has arisen out of the ruins of ground zero" (Schröder, 2002).

Responding to the institutional vacuum in Germany after 9/11, the Action Plan established a distinct architecture for implementing preventive policies in fragile states. In particular, the strategy attempts to anchor crisis prevention as a cross-cutting task, coordinating different ministries in this policy area, strengthening collaboration with non-state actors, and developing specific structures for crisis prevention (Bundesregierung, 2004). Two newly created bodies monitor and assist the implementation of the Action Plan: the interministerial steering group for crisis prevention (Ressortkreis zivile Krisenprävention) and a non-governmental Advisory Board (Beirat für zivile Krisenprävention).

Contextual factors

Two developments at the national and international level influenced institutional and policy changes after 2001, among them the 2004 Action Plan: a renewed focus on intrastate conflict after the end of the Cold War and the transformation of the Bundeswehr (Federal Armed Forces) into an operational army.²⁵

With German reunification in 1990, foreign and development policies became less and less influenced by the ideologies of the Cold War era. With this came an increased focus on intrastate conflicts and discussions about the causes of political violence in the post-ideological era (Fukuyama, 1989).²⁶ One outcome of these discussions was the recognition that societal transformation and national development carry the potential to produce conflict.

This led state actors to question their role in and methods of conflict engagement. The challenges posed by intrastate conflicts eroding the core functions of statehood required new policy responses beyond traditional peace-making (Kaldor, 2013). In particular, development cooperation was tasked with the design of new instruments for preventing the violent escalation of conflicts and for post-conflict reconstruction, e.g. through civilian crisis prevention and conflict transformation measures. Furthermore, new principles for conflict-sensitive cooperation in fragile states such as “do no harm” emerged (Kirschner, 2007).²⁷

The academic discourse on “new wars” influenced policymakers’ perceptions of fragile states. Unlike their Anglo-Saxon colleagues, German scholars emphasized state decay in the age of globalization as the main cause of “new wars” (Brzoska, 2004; Münkler, 2003). Accordingly, the 2004 Action Plan states that “the breeding ground for most domestic conflicts is the merely formal or rudimentary existence of state structures or the failure of the state as such” (Bundesregierung, 2004, p.7). Thus, the German government considered fragile states as a major concern in the prevention of violent conflict and regional instability.

As a second contextual factor, following a Constitutional Court judgment in 1994, the Bundeswehr progressively transformed from an army of deterrence (“*Abschreckungsarmee*”) to a parliament-controlled operational army (“*Einsatzarmee*”). If authorized by a parliamentary majority, Germany’s participation in “out-of-area” NATO- and UN-led military missions was now no longer a breach of the German Constitution.

This sparked a debate about the nature of Germany’s engagement in armed conflicts (Dames, 2015). The mandate authorizing German troops to participate in the NATO-led mission in Afghanistan (International Security Assistance Force, ISAF) and the US-led Operation *Enduring Freedom* in the Horn of Africa finally required the development of a civilian strategy in response to crisis situations alongside military means.

At the same time, policymakers developed an interest in promoting the civilian component of interventions abroad. The mandates for the Bundeswehr missions were adopted by a left-wing parliamentary majority and under the Schröder coalition government between the Social Democrats (SPD) and the Greens. While the

²⁵ Most notably, in the left-wing government’s decision to intervene militarily in Kosovo.

²⁶ A prominent discussion revolved around the concept of so-called “new wars”. The term predominantly refers to intrastate conflicts. “New wars” are characterized by the involvement of regular and irregular conflict actors. They are not financed by the state through taxes but through war economies and external sources. Motives are less ideological but revolve around identity politics. “New wars” are not fought for particular policies. Rather, the goal is political mobilization, i.e. to gain support and achieve political control through war and terror against the “outgroup”, potentially involving displacement of and violence against civilians (Kaldor, 1999; Münkler, 2003).

²⁷ Projects “do no harm” if they avoid producing unintended consequences such as increasing incentives for state capture by rebel groups (Grossman, 1992).

wider German public moved from post-World War II anti-militarism to indifference (Fleckenstein, 2005), left-wing voters in particular remained sceptical of military intervention in the early 1990s (Baumann and Hellmann, 2001).²⁸ Policymakers emphasized the civilian dimension of crisis prevention as a legitimizing factor for foreign interventions, hoping to avoid further alienation of left-wing voters.

Two approaches: peacebuilding and institution-building

Following the publication of the 2004 Action Plan, the BMZ released two major strategies to operationalize its 161 recommendations: the Peace and Security division primarily focused on the peacebuilding component while the Governance division was concerned with long-term institution-building. This resulted in two approaches to development cooperation in fragile states throughout the first decade of the 21st century: peacebuilding and institution-building.

Although the foci of the two approaches are complementary, they have not always been seen as well integrated in the organizational framework of development cooperation in Germany. For example, in 2006 Cammack et al. considered that “[a]t the federal government level, interaction is not particularly coordinated because of tensions between the various ministries, and within the BMZ, because of the fragmentation of development actors. At field level, there is also evidence of lack of coherence and coordination” (2006, p.57).

Peacebuilding approaches

The 2005 “Cross-Sectoral Concept in Crisis Prevention, Conflict Transformation and Peace Consolidation in German Development Cooperation” established the peacebuilding approach in the BMZ. It defines “binding standards and recommendations for the design, implementation, and steering of German official development cooperation” (BMZ, 2005, p.8) in states at risk as defined in the BMZ’s annual CEW (see Chapter 2).

Overall, the 2005 strategy aims to improve the BMZ’s contribution to civilian crisis management. The BMZ’s country portfolios and country strategies now required the systematic integration of crisis prevention, conflict transformation, and peace consolidation in fragile and conflict-affected states. Germany’s changing engagement in armed conflicts lent particular importance to the strategy as the BMZ sought to take over primary responsibility for civilian reconstruction in Afghanistan, but faced resistance from the German foreign ministry (Auswärtiges Amt, AA) (Hofmann, 2007).

Empirical implications of the 2005 strategy

The 2005 BMZ cross-sectoral concept suggested integrating four major types of intervention into the country portfolios of fragile states, based on a Peace and Conflict Needs Assessment.

- Technical cooperation at the regional and national levels, e.g. in the field of early warning, to strengthen institutional capacities and staff competences
- Financial development cooperation for rebuilding infrastructure and implementing employment programmes, especially in post-conflict countries
- Development-oriented emergency and transitional aid to close the gap between humanitarian assistance and long-term approaches, e.g. by reintegrating ex-combatants
- Non-governmental support to states where official assistance is difficult due to political sensitivities or limited capability of ODA in terms of funding, staff, and methods of engagement, e.g. in security sector reform. The Civil Peace Service is a BMZ-funded instrument to link non-governmental and governmental organisations which can be deployed at a local level (Ziviler Friedensdienst, 2017).

²⁸ In 1990 and 1991 the majority of Germans disagreed with Helmut Kohl’s proposition to amend the Basic Law to legalize Germany’s military participation in UN missions (1990: 53% in former East and 66% in former West Germany; 1991: 49% in former East and 66% in former West Germany). Paradoxically, at the same time Germany contributed DM 18 billion to fund military interventions by its allies in the Gulf War. It also provided personnel and logistic support in other crises (Namibia, Bosnia, Macedonia) (Baumann and Hellmann, 2001).

Civil Peace Service staff implement non-violent conflict-resolution activities in the field of media assistance, conflict and human rights monitoring, victim support and reconciliation, peace education, and training of local partners. The strategy makes clear that development cooperation is always of a civilian character and can only support peace processes, not resolve conflicts.

The scale of aid allocation is expected to change if a country is identified as being at risk of or in a crisis situation. This implies that country portfolios of crisis-affected and at-risk countries had to be re-evaluated after 2005. The 2005 strategy outlines this process as follows. After an in-depth conflict analysis, the relevant country department (*Länderreferate*) re-evaluates country portfolios, reformulates development objectives, translates these objectives into a country strategy, and realigns the country portfolio. In conflict-affected countries this process takes place more regularly, and independently from negotiations with the partner government. BMZ country teams coordinate with other ministries, local partners, and non-governmental organizations in this process.

In addition, the strategy prescribes specific conflict-sensitive methods for engagement in countries identified by CEW. For example, Post Conflict Needs Assessments and participatory processes for poverty-reduction strategies, which include various conflict parties, are considered crucial in conflict-affected contexts. Further, donors should take into account partners' different absorption capacities depending on the stage of a conflict and link fast-impact projects with long-term measures.

Beyond BMZ-specific instruments, the ministry also participates in the interministerial crisis-prevention steering group. This group promotes the development of interdepartmental country strategies in crisis-affected situations by country working groups (*Ländergesprächskreise*). It is also responsible for ensuring a common interministerial understanding of practices, methods and indicators related to crisis prevention, especially Peace and Conflict Impact Assessments and "do no harm". For this purpose, the former GTZ was tasked with the development of these methods, including the awareness of staff members of the 'do no harm' principle.²⁹

To sum up, during the first decade of the 21st century German development policy introduced its first systematic concept to deal with fragile states. If the integration of civilian crisis prevention and fragility expressed in the 2005 strategy had implications for development policy on the ground, we expect to observe the following empirical implications.³⁰

	Implication
1	More aid is committed to conflict transformation and peacebuilding
2	Countries with a German military presence receive more aid on average
3	Financial development cooperation for reconstruction increases in post-conflict countries

Institution-building approaches

In 2007 the Governance, Democracy and Rule of Law division ("Governance division") of the BMZ established the process of institution-building approaches as embodied in the "Development-Oriented Transformation in Conditions of Fragile Statehood and Poor Government Performance" strategy. The division aspired to contribute to "sustainable stabilization" through "the creation of reliable state structures in the field of the rule of law, democracy, human rights security and strengthening civil society" (BMZ, 2007,

²⁹ "Sektorvorhaben Krisenprävention und Konfliktbearbeitung". The GTZ (Deutsche Gesellschaft für Technische Zusammenarbeit) became part of the GIZ (Deutsche Gesellschaft für Internationale Zusammenarbeit) in 2011.

³⁰ The provisions of the Action Plan should have two more observable implications. First, information-sharing and joint conflict assessments between different ministries should improve. Second, conflict-sensitive methods should be systematically integrated into development cooperation. Although we cannot test these two implications based on allocation patterns, we mention them here for possible future qualitative analysis.

p.12). Constituting “a strategic entry point for tackling crisis prevention” (ibid.), these activities were aimed at complementing the BMZ’s 2005 strategy, which was limited to conflict-affected and post-conflict countries where development and military actors worked together (ibid., p.6–7).

The main innovation of the 2007 strategy was to assess the performance of partner countries in governance in terms of not only effectiveness but also legitimacy. This echoes the OECD DAC’s recognition in the 2005 Paris Declaration that ownership is crucial for aid effectiveness and reflects the idea that donors can influence domestic political processes through strategically and well-designed aid. In particular, aid conditionality is considered a means for rewarding or sanctioning partner governments’ willingness to implement governance reforms (BMZ, 2007).

The 2007 strategy proposed six fields of action to strengthen the legitimacy of partner governments and to enhance the capacity for non-violent change:

- Promoting social inclusion through democratic participation and support of legislative authorities³¹
- Strengthening the rule of law and social justice³²
- Reforming the security sector to increase public safety and civilian control³³
- Enhancing the public administration’s responsiveness to citizens³⁴
- Providing basic social services to poor and marginalized groups³⁵
- Improving macroeconomic performance for pro-poor growth³⁶

The strategy recommends designing interventions in each of these action fields depending on the the governance level (low, medium, high) and the trend of development orientation (deteriorating, constant, improving).

Influenced by the 2008 Accra Agenda for Action, the BMZ’s Governance division later extended the concept of recipient “ownership” to include the wider population. While the 2007 strategy is centred on the state, the 2009 Special “Promoting Resilient States and Constructive State-Society Relations: Legitimacy, Transparency and Accountability”³⁷ referred to the OECD concept of state-building as an endogenous and political process (BMZ, 2009), and extended governance support beyond technical cooperation for state institutions to “build or strengthen resilient democratic state structures” (BMZ, 2009, p. 15). As these

³¹ Activities include strengthening national parliaments, civil society, pluralistic media, and political participation of marginalized groups as well as supporting political-administrative decentralization and good governance initiatives at regional level.

³² Activities include adapting partner countries’ legal practice to international human rights standards, promoting and supporting the prosecution of war crimes through technical assistance and capacity-building of the judiciary, and strengthening independent, non-discriminatory, and efficient justice and legal systems to facilitate non-violent conflict resolution.

³³ The aim of security sector reform is to increase constitutional, parliamentary, and civilian oversight. Activities include strengthening authorities to investigate and sanction human rights abuses by security forces, promoting access to marginalized groups, improving policing, community-level prevention work, and combating cross-border crime.

³⁴ Activities include strengthening transparency and accountability of the executive in public resource spending and compliance with state action, technical assistance for capacity and organizational development in public administrations, strengthening the fiscal base, transparent public finance systems and public participation in budget planning, promoting institutions to combat corruption (e.g. African Peer Review Mechanism), and the accountable and responsible use of natural resource revenues, as well as cooperating with local actors and international initiatives to monitor extractive industries.

³⁵ Interventions aim to promote basic health services, strengthen cooperation between governmental and non-governmental actors in the health sector, promote community-based schooling and safe places for learning and education that are culturally sensitive through curriculum reform, and improve decentralized infrastructure.

³⁶ Activities in this area aim to build institutional capacity to combat criminal economies, improve access to markets and employment for poor households, promote sustainable resource management and land distribution/rights for both sexes as well as training and job opportunities for vulnerable young people, and strengthen regional economic integration and border controls.

³⁷ In German, the paper was published as: BMZ Strategiepapier 01-2010, „Die Förderung konstruktiver Staat-Gesellschaft-Beziehungen – Legitimität, Transparenz, Rechenschaft.“
http://www.bmz.de/de/mediathek/publikationen/archiv/reihen/strategiepapiere/Strategiepapier298_01_2010_de.pdf.

processes are considered political, development cooperation cannot limit its support to technical assistance for state institutions.

Empirical implications of the 2007 and 2009 strategies

The 2007 strategy focused on long-term institution-building and emerged in parallel with the ministry's 2005 peacebuilding strategy. The 2007 strategy suggested instruments to achieve results in the six above-mentioned action areas and depending on the governance level and development orientation of a fragile state.³⁸

The 2009 Special extended governance support to non-state actors in fragile states. German development cooperation intended to improve governance through better state-society interactions. While remaining vague on which specific interventions are supported, the 2009 document clearly committed Germany to channelling aid to or via non-state actors in fragile states.

We derive the following empirical implications from the strategic developments in 2007 and 2009.

	Implication
4	More aid is channelled to non-state actors in countries with low governance and deteriorating development orientation
5	More aid is committed to the direct delivery of basic services in countries with low governance and deteriorating development orientation
6	Aid supports public institutions, e.g. ministries, national parliaments or local governments, in countries with low governance and improving development orientation
7	Financial development cooperation supports the government in its national poverty and sectoral strategies in countries with low governance and improving development orientation
8	More aid is allocated in basket funds in countries with low governance and constant or improving development orientation
9	More aid is used to support local governments in countries with low governance and constant development orientation
10	More assistance for crisis prevention is given to relevant regional bodies (regions) in countries with low governance and constant development orientation
11	The relative share of governance aid to non-state actors, e.g. the media, increases

Summary

The work on peacebuilding and on institution-building is complementary. While the former establishes the methods of engagement and means to address conflict in fragile states, the latter focuses on questions of long-term governance. In the BMZ the strategies and resulting instruments exist in parallel, with the peace and security division focusing primarily on the former while the governance division is concerned with the latter. While at one point the ministry was criticized for not being adequately coordinated at the level of operations (Cammack et al., 2006), the BMZ now emphasizes the degree to which the relevant divisions work together (see also Federal Government of Germany, 2017). At the same time, the BMZ is of course only one of several German actors on fragile states, including the AA, the Ministry of Defence, and even the Ministry of Interior (e.g. for police training).

³⁸ For the detailed recommendations, see online appendix.

3.3.2 Responding to the Arab Spring: a new approach to stabilization

In 2012, the Federal Government released the “Interdepartmental Guidelines for a Coherent Policy of the German Federal Government in Fragile States”. These provide the basis for a reform of the BMZ strategy towards fragile states, laid out in the 2013 strategy “Development for Peace and Security”.

The uprisings in MENA countries – the so-called “Arab Spring” – illustrated that the existing approach of development cooperation had not led to stabilization, but might even have fuelled instability (Grävingholt, 2016). The Arab Spring led to a new focus on the countries of the MENA region and brought new issues onto the agenda of foreign and development policy-making. This is illustrated, for instance, in the fourth report of the Bundestag on the implementation of the 2004 Action Plan (Bundesregierung, 2014). This explicitly mentions the political and social upheavals in the Arab world as a trigger for changes in the direction of development cooperation. We therefore focus on developments around 2012 as a second critical juncture.

Contextual factors

Developments at the national and international levels preceded the second critical juncture in Germany’s fragile states policy. These conditioned and influenced the development of the 2012 and 2013 documents.

The domestic environment

The creation of a German strategy on fragile states as embodied in the 2004 Action Plan was an important step. However, academics and civil society criticized its substance and operationalization. In general, observers agree that a lack of political will and organizational resources obstructed the plan’s operationalization.

At the political level, the development of two approaches within the BMZ and the absence of coordination with other German actors over fragile states was blamed on the lack of strategic prioritization in the 2004 Action Plan, which put forward 161 action points without setting any priorities (Beirat Zivile Krisenprävention, 2014; Kirschner, 2007). A lack of prioritization is particularly problematic in fragile and conflict-affected states because of multiple dilemmas that occur in these contexts and require actors to make a trade-off (Faust et al., 2013; Furness, 2014; Schneckener, 2007).

Moreover, the political dilemma between public demands at home and in recipient countries (Faust et al., 2013) was not adequately addressed. One proposed means to deal with this shortcoming was investment in a dedicated communications strategy about the Action Plan to the German public (Beirat Zivile Krisenprävention, 2008; Nachtwei, 2010).

Furthermore, the absence of political leadership by one ministry (Nachtwei, 2010) and fragmentation between coexisting approaches on fragile states created bureaucratic problems. Actors’ and ministries’ uncoordinated trade-offs led to incoherencies, a lack of coordination, and even competition between actors, e.g. during reconstruction in Kunduz, Afghanistan (Hofmann, 2007).

At the organizational level, the main criticism related to budget allocations, civil society inclusion, and shared early warning of crises. While the budget for civilian crisis prevention increased over time, it was insufficiently coordinated between ministries. The interdepartmental steering group’s mandate was reduced from the allocation of funds to different ministerial activities to non-binding information exchange (Nachtwei, 2010; Stengel and Weller, 2008).

In addition, in its response to the government’s implementation report, the Advisory Board regretted that its expertise was not sufficiently taken into account. This was blamed on the operational capacity and influence of the steering group (Beirat Zivile Krisenprävention, 2008). By 2010 no shared early warning mechanism had been established as announced in the Action Plan and demanded by the Advisory Board (Faust, 2011; Faust et al., 2013).

The international environment

At the international level, the disappointing results and academic criticism of liberal state-building efforts in Afghanistan, Iraq, and the DRC (Burgess, 2017; Paris, 2010) increasingly required donors to justify their

engagement in fragile contexts (Mac Ginty and Richmond, 2013; Pospisil and Kühn, 2016). The 2011 “New Deal For Engagement in Fragile States” between traditional donor countries and the self-declared g7+ group of fragile states served as a legitimizing factor for Western state-building (Hingorani, 2015; Mac Ginty and Richmond, 2013; Nussbaum et al., 2012). Its focus on recipient–donor relations and the emphasis on legitimacy through local involvement shaped international state-building strategy. The five peace- and state-building goals (PSG5) concluded in the New Deal subsequently served as a reference for German policies towards fragile states.³⁹

However, the criticism of Germany’s fragile states agenda was only heard at the decision-making level during the Arab Spring. German foreign minister Guido Westerwelle concluded that “for too long regime or government stability was confused with state stability”⁴⁰ (Boldt, 2012).

The implications were twofold. First, autocratic partner regimes in North Africa and the Arabian Peninsula were explicitly considered fragile after popular upheaval and erupting conflicts. The popular uprisings had been predicted neither by political analysts nor by fragility indices (Buterbaugh et al., 2017; Gause, 2011). Second, popular upheavals in autocratically governed partner countries raised questions about how Germany could strategically and coherently engage with partner regimes that suffered from serious legitimacy deficits (Asseburg, 2011).

Towards coherent policies: the comprehensive approach and the local turn

In 2012, the “grand coalition” government released “Interdepartmental Guidelines for a Coherent Policy of the German Federal Government in Fragile States”. The BMZ Peace and Security division operationalized these guidelines in the 2013 strategy “Development for Peace and Security”. The latter replaced the division’s 2005 strategy and is its current reference document for development cooperation in fragile states. By explicitly integrating elements from the 2007 BMZ governance strategy, the 2013 strategy seeks to bring the two tracks together inside the ministry (BMZ, 2013a, p.4).

Comprehensive approach

The 2012 guidelines address the problem of multiple coexisting strategies by introducing the comprehensive approach (“*Vernetzte Sicherheit*”) to German cooperation with fragile partner countries. Introduced in the 2006 *Weißbuch* published by the Federal Ministry of Defence, *Vernetzte Sicherheit* treats security as a global and multidimensional concept. It is assumed that the military can no longer ensure security on its own. Rather, a networked response to crises by various German and international actors is required.⁴¹ The comprehensive approach builds on the assumption that German development, military, and other actors (can) have common goals and methods of engagement and complement each other according to their comparative advantage in fragile states (Wittkowsky and Meierjohann, 2011). Put succinctly, civilian and military crisis responses should no longer follow different logics and objectives.

This has implications for development cooperation: while it remains civilian in nature, the comprehensive approach extends its mandate beyond crisis prevention. The BMZ strategy requires “the overall portfolio of instruments [to be geared] systematically and strategically to the respective development policy objectives” (BMZ, 2013b, p.19) and the sequencing of these instruments to be improved.

The local turn

The 2012 guidelines also emphasize the importance of supporting endogenous processes and local ownership, reflecting the “local turn” in state-building and the importance of “ownership” in development

³⁹ The five goals are: legitimate politics: foster inclusive political settlements and conflict resolution; security: establish and strengthen people’s security; justice: address injustices and increase people’s access to justice; economic foundations: generate employment and improve livelihoods; revenues & services: manage revenue and build capacity for accountable and fair service delivery.

⁴⁰ Exact quote in German: “*Viel zu lange wurde die Stabilität von Staaten mit der Stabilität von Regierungen verwechselt*”.

⁴¹ “*Sicherheit kann daher weder rein national noch allein durch Streitkräfte gewährleistet werden*” (Bundesministerium der Verteidigung, 2006, p.25).

cooperation more generally. Accelerated by the g7+ activism, disappointing results of state-building in the 1990s, and declining enthusiasm for the “War on Terror”, the “local turn” seeks to strengthen inclusive bottom-up processes in fragile states (Mac Ginty and Richmond, 2013; Pospisil and Kühn, 2016).

The BMZ’s 2013 strategy moves beyond a purely institutionalist understanding of governance. This has led some to claim that the 2012 and 2013 publications eclipse the central state (Pospisil and Kühn, 2016). We, however, would argue that the strategy rather reflects a changed understanding of the concept of state: it remains central but is increasingly understood in relation to wider society and local indigenous (less formalized) institutional settings (see Grävingholt et al., 2015). Consequently, development cooperation aims to “strengthen states and their societies in terms of their adaptability and resilience to external shocks and crises” (BMZ, 2013b, p.10) through political settlement and improved state-society relations. Ideally, activities link the central state to the local level, or at least support both.

Finally, to display a change in donor-recipient relations towards more local ownership in state-building, the core areas of action in the 2013 strategy converge with the New Deal’s Peacebuilding and Statebuilding Goals (PSG5).

Empirical implications of the 2013 strategy

The BMZ’s embracing of the comprehensive approach in 2013 aims to improve civilian-military links, the sequencing of instruments, and coordination with other foreign actors in fragile states. Furthermore, the 2013 strategy proposes a wide array of instruments in fragile states: technical and financial cooperation, private-sector development cooperation, and the Civil Peace Service (BMZ, 2013b).

Structure-supporting transitional aid (ESÜH) links humanitarian and development efforts. This requires coordination with the AA, which is responsible for humanitarian aid. The BMZ 2013 “Strategy on Transitional Development Assistance” reforms the instrument (BMZ, 2013c). It is different from development assistance and previous transitional aid because support is not limited to BMZ partner countries, but explicitly prioritizes fragile, post-conflict, and protracted crisis contexts where cooperation with the recipient government is impossible (BMZ, 2013c, p.10). The BMZ’s control of ESÜH extends its competences for participating in interventions in crisis situations where development cooperation is not possible. This strengthens its role in fragile states beyond governance support and in relation to the AA, though it should be noted that previously the BMZ was already in charge of developmental emergency and transitional aid (ENÜH).

Unlike the EU, Germany regards general budgetary support as an inadequate form of aid for fragile states. This may somewhat contradict the emphasis on ownership in the New Deal, a reference document for Germany’s 2013 strategy, because exiting from budget support can reduce the ownership of partner countries (Orth et al., 2018).

The introduction of the national cross-sectoral FS marker reflects the increased links between civil and military crisis management. It labels projects contributing to peace and security objectives, thereby giving the FS marker a broader application than the KR marker it replaced.

In summary, we derive the following empirical implications from the discussion. As above, the empirical analysis is limited to those implications that can be tested with data on allocations.

	Implication
12	Non-partner countries that are affected by armed conflict or in the transition phase receive transitional aid
13	More aid was allocated to conflict prevention programmes
14	No aid was granted in the form of budgetary support

3.3.3 At the crossroads: the 2017 interministerial guidelines “Preventing crises, managing conflicts, building peace”

In 2017, the Federal Government released the interministerial guidelines “Preventing Crises, Managing Conflicts, Building Peace”. Unlike the 2012 guidelines, these replaced the 2004 Action Plan. The release of the guidelines did not follow a critical juncture. With a view to phasing out German engagement in Afghanistan, the Federal Government aimed to preserve its knowledge on engagement in fragile states.

The 2017 version fulfils the task formulated in the 2012 guidelines to define Germany’s priorities for bilateral engagement in fragile states. These priorities should be determined by the immediate threat that the situation in a partner country poses to Germany’s and Europe’s domestic security, the international community’s expectations of Germany, and Germany’s capacity to contribute to improvements in partner countries (Federal Government of Germany, 2017).

Unlike the EU’s Global Strategy and the 2013 BMZ strategy, the 2017 guidelines do not juxtapose resilience and fragility (Wagner and Anholt, 2016). Rather, resilience is portrayed as an objective to strengthen the absorption capacities of all countries, irrespective of their fragility. This reflects the SDGs’ principles of universality for all UN member states. Resilience can thus be regarded as an umbrella term for sustainable long-term stabilization in fragile states beyond military means. It includes and connects conflict prevention and post-conflict rehabilitation, which extend over a longer period. Furthermore, unlike previous buzzwords in international development – such as “democracy” or “human security” – the malleability of resilience may provide an opportunity for networked security. Consequently, resilience may provide common ground for actors from different backgrounds with diverging interests, e.g. military and civilian staff, traditional and non-traditional donors, humanitarian and development workers, and state and non-state actors (ibid.).⁴² The influence of the resilience discourse in Germany is apparent in the creation in 2015 of a division in the AA responsible for “conflict prevention, stabilization and conflict rehabilitation” (Wittkowsky, 2017).

However, the guidelines are careful not to use resilience as a leitmotif. They seldom mention resilience (nine times) compared to stabilization (31 times) and prevention (85 times). This may be related to the fact that resilience remains an ambiguous term that leaves room for debate about its aims and implementation (Chandler, 2013).

Efforts to operationalize the 2017 guidelines are currently ongoing across various ministries. In the BMZ, different cross-sectoral divisions (*Querschnittsreferate*) operationalize the guidelines with particular attention paid to the 2030 Agenda for Sustainable Development. In addition, the BMZ participates in theme-specific interministerial task forces. For example, the BMZ’s Governance division is a member of a newly created task force on justice and the rule of law. At present, however, the 2013 strategy “Development for Peace and Security” remains the framework of reference for official development cooperation in fragile states.

Summary

German development cooperation policy towards fragile states has constantly evolved, as evidenced by continuous debate and strategic reorientation.

The understanding of fragility and risks when operating in fragile contexts has improved and taken up academic debates on how to define fragility (see Chapter 2). Moreover, policymakers have adjusted strategies to changing requirements in the global environment. On the one hand, this refers to international developments in states affected by fragility. On the other, strategies have taken into account policy-making at the international level. For example, the German Federal Government released the 2004 Action Plan for Crisis Prevention and the 2017 guidelines one year after the 2003 EU Security Strategy (ESS) and the 2016

⁴² The BRICS and other emerging donor countries refused to sign the 2011 “New Deal on Peacebuilding and Statebuilding” in Busan. In particular, non-traditional donors accuse the good governance paradigm of interventionism, the promotion of Western concepts of legitimate statehood, and stigmatizing aid recipients through the “fragile states” label.

EU Global Strategy (EUGS) respectively. Developments in the OECD DAC's meetings on aid effectiveness and the findings from the States of Fragility Report published annually since 2012 also inform German policy. Similarly, the World Bank's 2011 World Development Report "Conflict, Security and Development" influenced the BMZ's 2013 strategy. The former reveals a rethinking of legitimacy and the involvement of grass-roots actors by emphasizing the importance of citizens' "institutional trust" and "layered approaches" in conflict-affected states.

Strategies on development cooperation in fragile contexts increasingly incorporate insights from academic research. This is particularly true with regard to the conceptualization of state fragility (Paris, 2011). Nonetheless, there remain gaps in relation to the operationalization and implementation of academic insights (Baranyi and Desrosiers, 2012). In the following section, we put our empirical implications to the test.

3.4 Empirical results

Table 10 presents the aggregated results of the empirical analysis. All descriptive and inferential results are included in the online appendix.⁴³

Table 10 Results of empirical analysis

	Implication	Year	Result
1	More aid is committed to conflict transformation and peacebuilding	2005	++
2	Countries with a German military presence receive more aid on average	2005	++
3	Financial development cooperation for reconstruction increases in post-conflict countries	2005	++
4	More aid is channelled to non-state actors in countries with low governance and deteriorating development orientation	2007	-
5	More aid is committed to the direct delivery of basic services in countries with low governance and deteriorating development orientation	2007	+
6	Aid supports public institutions, e.g. ministries, national parliaments or local governments, in countries with low governance and improving development orientation	2007	++
7	Financial development cooperation supports the government in its national poverty and sectoral strategies in countries with low governance and improving development orientation	2007	--
8	More aid is allocated in basket funds in countries with low governance and constant or improving development orientation	2007	(o)
9	More aid is used to support local governments in countries with low governance and constant development orientation	2007	+
10	More assistance for crisis prevention is given to relevant regional bodies (regions) in countries with low governance and constant development orientation	2007	NA
11	The relative share of governance aid to non-state actors, e.g. the media, increases	2009	-

⁴³ We cannot test implication 10 since we do not have data on the regional level because it cannot be linked to national data on fragility. We nevertheless include the implications in the table for reasons of completeness.

	Implication	Year	Result
12	Non-partner countries that are affected by armed conflict or in the transition phase receive transitional aid	2013	++
13	More aid was allocated to conflict prevention programmes	2013	+
14	No aid was granted in the form of budgetary support	2013	+

2005 strategy

With regard to implication 1, we find strong evidence that allocations to conflict transformation and peacebuilding increased. The analytical results show a positive and significant coefficient and thus point in the same direction.

Concerning implication 2, we find steadily increasing total commitments to countries with Bundeswehr missions. The relative share of commitments does not show a steadily increasing trend, but rather an overall increase. The analytical results show a positive and significant coefficient. Overall, we find strong evidence that commitments follow strategic implication 2.

Tests of the last implication of the 2005 strategy show an increase in financial development cooperation for reconstruction in post-conflict countries, particularly after 2009, when taking into account total and relative commitments. The econometric approach shows a positive and significant coefficient. Overall, we find strong evidence that commitments follow strategic implication 3.

In summary, we find that allocations follow the 2005 strategy.

2007 and 2009 strategies

On implication 4, we find a small increase in total commitments to non-state actors in countries with low governance and deteriorating development orientation. The relative share of commitments fluctuates and shows no clear trend. The coefficient on the interaction in our inferential approach is negative and significant. Overall, we find weak evidence against the implementation of the strategy with regard to implication 4.

Looking at implication 5, we find that total aid committed to the direct delivery of basic services in countries with low governance and deteriorating governance orientation increased, whereas the development of relative share does not show a clear pattern. The inferential results return a positive but non-significant coefficient. We regard this as weak evidence that commitments developed according to implication 5.

Commitments supporting public institutions in countries with low governance and improving development orientation vary greatly from year to year but show an overall increase in respect of total commitments and the share of commitments. This result is supported by the inferential analysis. The coefficient is positive but not significant. Since all approaches point in the same direction, we rate this as strong evidence that aid commitments were allocated according to implication 6.

With regard to implication 7, we find almost no allocations in the sectors concerned. The statistical analysis returns a non-significant positive coefficient, which is probably due to a single project in 2017.⁴⁴ We regard this as strong evidence against implementation of this implication. However, implication 7 was superseded by implication 14, which might explain the lack of implementation.

Implication 8 considers aid commitments in the form of basket funds to countries in the BMZ portfolio with low governance and constant or improving development orientation. Since, due to missing data, the first observation dates from 2010,⁴⁵ we do not test the implication statistically. We find no clear trend in respect

⁴⁴ The sharp increase in 2017 is due to a single KfW project on behalf of the Federal Ministry of Finance.

⁴⁵ The use of the CRS Code "Basket Funds/pooled funding" started in 2010.

of total commitments or the share of commitments. As the available data show no clear pattern, there is uncertainty about implementation of implication 8.

Implication 9 states that more aid supports local governments in countries with low governance and constant development orientation. Neither the total nor the relative commitments show an increase in the first years after publication of the strategy. By contrast, the inferential statistic shows a significant positive coefficient, although the results are driven by an increase in commitments after 2013, which is seven years after publication of the strategy. In summary, we find weak evidence that commitments were allocated in accordance with implication 9.

Implication 10 focuses on aid commitments to relevant regional bodies. Due to missing data on the regional level, we cannot test this hypothesis.

The 2009 strategy states that commitments in the form of governance aid to non-state actors should increase. We test this in implication 11. The total commitments show a steady decline after 2006 followed by a temporary increase in 2013, 2014, and 2017. The relative share of commitments declines continuously after 2007 except for 2013 and 2017. The inferential approach yields a significant negative coefficient. We thus find weak evidence against implementation of the 2009 strategy with regard to implication 11.

2013 strategy

Total and relative transitional aid commitments to non-partner countries in conflict or transition rose after publication of the 2013 strategy. The coefficient of the inferential model is positive and significant. We thus conclude that there is strong evidence that allocations are in line with implication 12.

Conflict prevention programmes in fragile states see an increase in total commitments after 2013, as indicated in implication 13. The result for the relative share of aid commitments is less clear. The treatment effect from the inferential analysis is negative, but very small and not significant. Overall, we interpret this as weak evidence for the implementation of the strategy.

The descriptive analysis of implication 14 shows a large decline in aid commitments in the form of general budgetary support, which started in 2011. We find an increase in 2017, which is due to a single project. Treating this project as an exception to an otherwise consistent trend, we interpret our findings as weak evidence that aid commitments were allocated according to implication 14.

Discussion

Our investigation of patterns of commitment shows that aid allocations broadly follow the strategies. Before presenting the results in more detail, we highlight three limitations of the study. First, evidence on changes in commitments does not provide direct insights into aid effectiveness. However, in respect of the latter, the results in Chapter 5 broadly suggest that the success of projects in fragile contexts is not significantly lower than in other contexts. Second, while the statistical study of policy changes after critical junctures across cases delivers insights on *whether* written strategies affect policy-making, it does not sufficiently illuminate the causal mechanisms explaining *how* the written word might or might not shape policies in individual cases. We have taken steps in this direction by systematically looking at decision-making processes. However, this evaluation has not systematically tested the impact of organizational features on strategy implementation. Third, we base our findings exclusively on an analysis of commitments. Consequently, we do not test all implications derived from the strategies. For instance, we cannot adequately test whether the planning and implementation of interventions follows awareness of the “do no harm” principle. Future research/evaluations should invest in systematically evaluating the quality of interventions in fragile states in terms of principles such as conflict analysis and do no harm. With these limitations in mind, the following summarizes our findings.

With some exceptions, we find significant changes in the distribution of obligations for sectors mentioned in the strategies. These results show that allocation patterns are mostly in line with policy recommendations as laid out in strategy documents.

However, two policy recommendations were not as well implemented. First, although we should be cautious due to data limitations, we do not find evidence of an increase in commitments to basket funds (implication 8). Basket funds, i.e. the joint financing of a specific purpose by multiple donors through a single account, have the potential to improve development cooperation in fragile countries. Although evidence on basket funds remains scarce (D'Aquino et al., 2019), they promise greater donor coordination and harmonization as well as increased ownership (Manuel et al., 2012). Hence, the use of basket funds is in line with the aid effectiveness agenda (3rd High Level Forum on Aid Effectiveness, 2008; High Level Forum, 2005). Moreover, basket funds have the potential to reduce the administrative burden faced by partner countries by reducing the need for bilateral coordination with multiple donors and uniform procedures (Commings et al., 2019). This can be particularly important in states with low capacity.

Second, there is weak evidence of decreasing support to non-state actors against implications 4 and 11. These findings support the view that German development cooperation prioritizes an active role of the state specifically in service delivery, in contrast to donors that favour market-oriented delivery of public services such as Canada, the United Kingdom or Scandinavian countries, which often bypass state institutions (Dietrich, 2016). This touches upon the question of whether state or non-state actors are suitable partners for development cooperation in fragile contexts. Arguably, the use of country systems in fragile states entails risks that might be mitigated using a parallel system such as channelling support through non-state actors (Manuel et al., 2012). This probably explains the finding that, in general, donors channel development cooperation through non-state actors where governance is low (Acht et al., 2015; Dietrich, 2013). Here, service delivery through non-state actors might be more efficient. Our result in Chapter 5 that project success in the health sector is lower in low-capacity countries also points in this direction. However, in authoritarian countries – where the commitment to delivering public goods can be expected to be lower than in democracies (Bueno de Mesquita et al., 2003; Lieberman, 2018) – bypassing partner countries' governments can reduce incentives to demand policy reforms and accountability, thereby strengthening authoritarian regimes (DiLorenzo, 2018). This is important since German development cooperation specifically emphasizes the importance of non-state actors to “support a bottom-up approach towards democratisation and active political involvement” (BMZ, 2009a, p.29), especially in countries affected by crisis where efficient government-to-government aid is not possible. Future research could further differentiate the effects of bypass aid for different types of development cooperation, partner countries, and non-state partners. In any case, our findings support the observation that “engaging with non-state actors and legitimate local organisations to strengthen state-society relations remains a challenge for development partners” (OECD, 2011a, p.15).

3.5 Results and recommendations

We summarize the analysis of strategies and allocations as follows. First, we discuss the results of the analysis of strategic shifts and continuities and link them to the academic debate. Second, we summarize the allocation analysis. We conclude each of these with a recommendation.

Based on an analysis of eight key strategy papers published between 2004 and 2017 by the German Federal Government and the BMZ, our analysis has identified important strategic shifts during two critical junctures: 9/11 and the Arab Spring. With each of these, new discourses have emerged on how development cooperation can be adapted to the challenges arising from fragility. At the same time, Germany has constantly revised strategies and concepts based on international and domestic developments. Consequently, the approach to fragility has evolved over the years and has seen fundamental changes.

The Al-Qaeda attacks on 11 September 2001 brought fragile states onto the international agenda. In Germany, the 2004 Interministerial Action Plan “Civilian crisis prevention, conflict transformation and peace consolidation” can be considered the first joined-up strategy targeting a distinct group of fragile states.

Since then, two approaches have emerged in German development cooperation towards fragile states: peacebuilding and institution-building. Although we have seen major innovations within each of these approaches and, more recently, attempts to integrate them more closely, they have both maintained a

considerable degree of autonomy. This is evidenced by the fact that they are anchored in two different divisions – today even in different directorates-general – in the BMZ, draw on different concepts of fragility (see Chapter 2), and are represented by different strands of strategy papers.

The institutional specialization must be considered against the background that the foci of peace- and institution-building are distinct but complementary. They are distinct because, on the one hand, peacebuilding is about more than building strong and inclusive institutions. For instance, it also includes disarmament, demobilisation, and reintegration, or the reconstruction of infrastructure. On the other hand, institutions affect environmental, economic, and social development and thus go beyond peacebuilding. At the same time, peace- and institution-building are complementary (Federal Government of Germany, 2017). Political institutions and development orientation are probably among the key determinants of the risk of violent conflict escalation. For instance, institutions include or exclude groups from power, wealth, and decision-making processes, thereby aggravating or mitigating inequalities in society which in turn can lead to conflict (Cederman et al., 2013; Østby, 2008; Stewart, 2008). Moreover, there is evidence that institutional quality such as rule of law, low corruption, and bureaucratic quality can reduce the likelihood of the onset of civil war (Fjelde, 2009; Taydas et al., 2010), whereas good governance can reduce the likelihood of conflict recurring (Hegre and Nygård, 2015). Put more generally, institutions form the structural conditions that shape the incentives of societal, economic, and political actors and thereby influence the decision to resort to violence or play by the rules as embodied in institutions. For similar reasons, peacebuilding processes often centre on the question of shaping institutions. But peace and conflict also affect institutions: institutional changes due to violent conflict can render obsolete an assessment of governance and development orientation.

In the light of the distinct nature of peace- and institution-building and the fact that strategies have been implemented well, the institutional specialization within the BMZ seems fit for purpose. Moreover, it seems advisable to maintain separate concepts for crisis early warning – such as the CEW, overseen by the Peace and Conflict division – and governance, such as the Catalogue of Criteria for Assessing Development Orientation, overseen by the Governance division (see Chapter 2).

At the same time, successful development cooperation in fragile states must always take into account peace- and institution-building together. A close collaboration among the responsible organizational units would ensure complementarity. The joint reflection of development orientation, the level of governance, and CEW are a good step in doing justice to the complementary yet distinct nature of peace- and institution-building.

The current revision of the CEW and the measurement of the level of governance provides an opportunity to further deepen cooperation, develop a uniform understanding of fragility, and eliminate potential redundancies by possibly integrating existing approaches (see Recommendation 1). On the one hand, this may include more clearly defining and demarcating the objectives of both approaches. On the other, joint data collection on shared indicators could increase efficiency and thereby improve assessments.

We thus make the following recommendation:

Recommendation 2

Based on a systematic conceptual comparison of approaches to defining and measuring fragility, an analysis of strategy documents, a review of the structure and processes of the German bilateral development cooperation system, and interviews, this evaluation recommends that the BMZ maintain its practice of dealing separately with questions of peace and security, and of governance. At the same time, the BMZ should ensure coordination and complementarity between the responsible organizational units wherever possible. The BMZ should systematically compare concept designs across both divisions with the aim of clearly demarcating approaches in order to avoid duplication and redundancy. Where concept designs complement one another, a formal mechanism for sharing indicators and data collection efforts should be established.

The investigation of patterns of commitment shows that aid allocations broadly follow the strategies. After 2005, more aid was committed to conflict transformation and peacebuilding, to fragile countries with a German military presence, and to reconstruction in post-conflict countries. After 2007, more aid was committed to the direct delivery of basic services in countries with low governance and deteriorating development orientation; there were more commitments supporting public institutions in countries with low governance and improving development orientation; and more was committed to support local governments in countries with low governance and constant development orientation. After 2013, aid commitments to non-partner countries in conflict or transition rose, and conflict-prevention programmes in fragile states saw an increase in total and relative commitments. We also find that the more general strategy of exiting from budgetary support also applies to fragile states.

Although these results show that allocation patterns are mostly in line with policy recommendations as laid down in strategy documents, we also find some strategies to be less well implemented. First, we do not find evidence of an increase in commitments to basket funds, an instrument promising greater donor coordination and harmonization as well as increased ownership and reduced administrative burden faced by partner countries. Second, we find weak evidence of decreasing support to non-state actors, contrary to strategic recommendations. Both findings touch upon the question of whether state or non-state actors are suitable partners for development cooperation in fragile contexts.

In summary, one of the main challenges lies in the strategic orientation of German development cooperation with regard to the question of what channels of delivery – among them the partner government and non-state actors – offer the greatest benefits of development cooperation in fragile contexts. The strategies under analysis in this chapter already go a long way in tailoring instruments to the local context by making recommendations that are conditional on the fragility, governance, and development orientation of partner countries. However, in some cases implementation is falling behind.

We thus make the following recommendation:

Recommendation 3

Based on an analysis of strategy documents and commitments, and a screening of the literature, this evaluation recommends that the BMZ ensure the implementation of its strategies specifically with regard to the support of non-state actors in fragile contexts. The BMZ should also develop a concrete action plan to assess and possibly adapt guiding principles on cooperation with non-state actors in fragile contexts based on available evidence.

Moreover, the BMZ should continue the good practice of constantly adapting strategies to the ever-improving understanding of successful development cooperation in fragile contexts. The existing practice of tailoring strategic recommendations to contextual conditions is likely to do justice to the increasing analytical precision of scientific research and the validity of scientific explanations that depend on particular background conditions.

4. EVALUATION

4.1 Introduction

This chapter takes a meta-evaluation perspective investigating the relationship between fragility and evaluation quality. Church and Shouldice (2002) stress the importance of improving the practice and testing of theories of change for project success and point out the challenge of doing so in evidence-based ways. However, more than half (55%) of the agencies surveyed by Steets et al. (2016) stated that they were “not so satisfied” or “not satisfied at all” with monitoring and evaluation of their implementing partners.

In a recent meta-study, Noltze et al. (2018c) found that the quality of evaluation reports varies across a range of factors. The quality of reports is crucial for understanding the success of development cooperation in two respects. On the one hand, report quality may be an influencing factor when it comes to assessing project success in a context of fragility. On the other hand, evaluation reports are crucial for understanding what works in fragile contexts (Winckler Andersen et al., 2014). A negative impact of fragility on evaluation quality would impede better knowledge about aid effectiveness, particularly in those contexts where effective development cooperation is most needed.

This chapter examines how fragile contexts might influence the quality of evaluation. We explore the total effect of fragility on evaluation quality and examine the explanatory power of individual mechanisms. Furthermore, we formulate recommendations concerning strategically and operationally effective evaluation in fragile contexts and suggest ways to address potential challenges.

4.2 Theory: quality determinants

Fragility may affect evaluation quality through the local context, the characteristics of the project, and the nature of the evaluation. This section discusses mechanisms in each of these spheres as derived from evaluation guidelines, scientific research papers, and monitoring reports on lessons learnt published by international organizations. Figure 4 summarizes the discussion.

First, context characteristics may impede evaluators’ mobility and/or their access to the population. Potential causes are:

- Concerns about the personal safety of evaluators and other actors such as interviewees
- Security regulations of international or local agencies limiting movement
- Restricted access to specific areas due to a poor security environment
- Poor infrastructure, e.g. communications and transport (Church and Shouldice, 2002; Steets et al., 2016; Stoddard and Jillani, 2016).

The volatile environment in a fragile country may also influence the planning and implementation of evaluations. When under time pressure, compromises are made in terms of baseline data, control groups, or other evaluation quality factors. Moreover, a sensitive political climate, security concerns, and lack of mutual trust within a society may influence the response behaviour or the willingness of local actors (especially the population) to share information (Chapman et al., 2009; Church and Shouldice, 2002; Krueger and Sagmeister, 2014). People who are unsure about their personal safety or that of their family, or respondents in a societal atmosphere of mistrust, may think twice about whether a particular piece of information they might give could have negative consequences.

Second, fragility can affect development projects. While Steets et al. (2016) mainly see capacity challenges, Chapman (2009) and Davenport and Moore (2015) emphasize interests, ownership or “willingness” issues. Regarding capacities, M&E systems or indicators might be weak or lacking. This may affect local partners and aid agencies alike. Possible causes are the lack of resources in terms of knowledge, time, and funding, or high staff turnover due to work pressures and rotation requirements in fragile countries (Chapman, 2009; Development Researchers Network, 2011). In terms of “willingness”, reduced motivation to share information openly may arise from an “affirmative culture” or perceived pressure to report overly positive results (e.g. in order to secure future funding). In addition, concerns about public image, uncooperative work atmosphere within the project, and relations with the project donor may affect openness to evaluation

(Church and Shouldice, 2002; Davenport and Moore, 2015; HAP - Humanitarian Accountability Partnership, 2013; Heathershaw, 2012; Ruppert et al., 2016; Stoddard and Jillani, 2016).

Third, evaluation itself can be challenging in fragile contexts. Fragility may affect:

- The recruitment of evaluators (it may be difficult to enlist qualified personnel for missions in insecure conditions)
- The conceptual framing (e.g. learning or accountability)
- The choice of evaluation questions, projects to evaluate, interviewees, and data sources (if specific areas or groups of people are not accessible)
- Resources for the evaluation
- The methods applied
- The sources disclosed (if the security of informants cannot be guaranteed, their identity should not be revealed by the evaluation)

Whether these mechanisms are at work in fragile contexts may depend on resources in terms of time and money available to evaluators, as well as on the power or dependency relationship with donors, local entities and other stakeholders. Moreover, Winckler Andersen et al. (Winckler Andersen et al., 2014) point out that evaluation in fragile contexts puts not only the evaluator but also local partners at risk. In addition, the evaluator might be confronted with situations requiring particular skills, e.g. building trust and using informal spaces, sensitivity and risk of misinterpretation, or reaction to the vulnerability of respondents and integrity issues (Chapman et al., 2009; Davenport and Moore, 2015; Davison, 2004; Fassin, 2012; OECD, 2012).

In summary, we derive the following hypothetical causes, hereinafter referred to as “mechanisms”, of low evaluation quality in fragile contexts from the theoretical review:

1. The mobility of evaluators reduces their access to the population
2. There is a low level of trust within the population
3. An affirmative culture among project staff causes respondents to report project results too positively
4. Due to a lack of resources, the monitoring and evaluation capacity within the project evaluated is low
5. Evaluators – depending as they do on a cooperative environment – work in a tense atmosphere of power relations and interest constellations, and may be influenced by stakeholders
6. Due to a lack of resources, the monitoring and evaluation capacity of the evaluator or the evaluation team is low

4.3 Method

The assessment of evaluation quality is based on Noltze et al. (2018b). They studied a range of evaluation criteria derived from the literature on evaluation (Patton, 2008; Scriven, 2009; Stufflebeam, 2001; Widmer, 2006) and development cooperation (Carlsson and Wohlgemuth, 1996; Hageboeck et al., 2013; Leeuw and Cooksy, 2005). Moreover, evaluation guidelines from the GIZ and KfW and other meta-evaluations are taken into account (Caspari, 2014; Huber et al., 2014; Raetzell and Krämer, 2013). Noltze et al. (2018b) used 16 different criteria for the quality of evaluation reports that take on the value 1 if the criterion is fulfilled, and 0 otherwise. Examples include research design (e.g. whether a control group existed) and robustness (e.g. whether data was triangulated) (see Noltze et al. 2018b, Table 4, pp. 74–75).

To operationalize our main explanatory variables on fragility, we draw on Constellations of State Fragility (see Chapter 2) and measure fragility through its three dimensions of authority, capacity and legitimacy

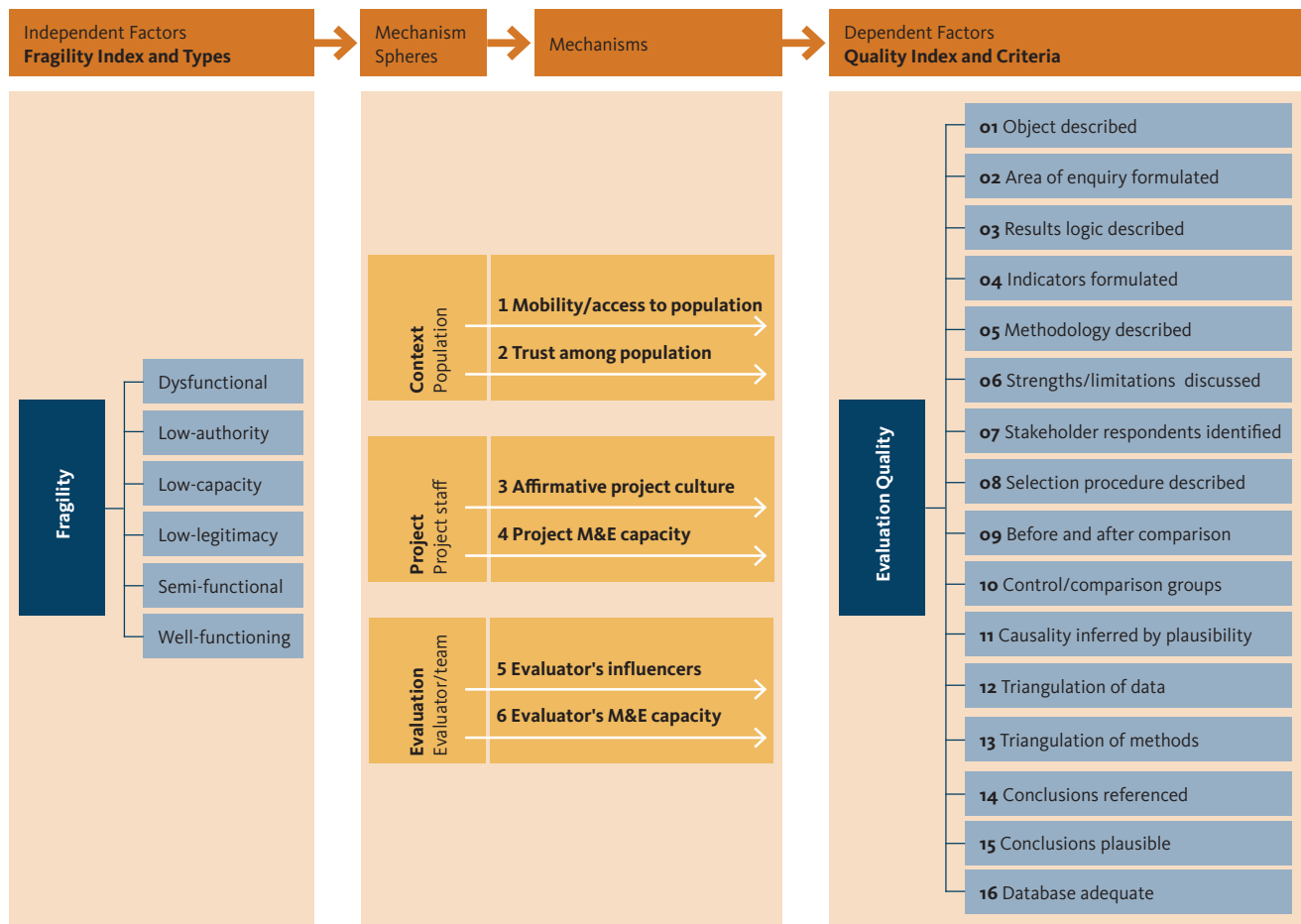
separately.⁴⁶ In order to test the overall effect of fragility, we construct an index of state fragility in which we aggregate the three dimensions by calculating their mean. To facilitate interpretation in the descriptive graphs (Figure 5) and for the mediation analyses, we also construct a binary measure of fragility which equals 1 for values of a fragility value larger than 0.5, and 0 otherwise. As such cut-off points always risk being arbitrary, we use different cut-off points in the statistical analyses to assess the robustness of our findings.⁴⁷ Figure 4 depicts the hypothesized relationship between fragility and evaluation quality via the six proposed mechanisms. We also consider different types of fragility (see Chapter 2) in the descriptive quantitative analyses.⁴⁸

Suitable operationalizations for a quantitative test of potential causal mechanisms exist for three variables: mobility, trust among the population, and project M&E. The degree of mobility at country level is expressed by the variable “timeliness”, which forms part of the World Bank Logistics Performance Index (Arvis et al., 2018). This reflects the share of ordered shipments that reach their final destination within the scheduled period. To operationalize trust among the population, we use the variable “social capital” consisting of the concepts “level of trust” and “civil society activity” from the Bertelsmann Transformation Index (Hartmann et al., 2018) as a proxy. To measure M&E capacity, we use the binary indicator on whether an evaluation included project monitoring data (Noltze et al., 2018c).

⁴⁶ Based on feedback from the reference group, we disaggregate fragility in its different dimensions, in addition to using the overall index, as we did in earlier versions of this report. For more details on those dimensions, see Chapter 2.

⁴⁷ The main models in the mediation analyses are run using a 50th percentile cut-off point. Values above the cut-off are regarded as high and below as low levels of fragility. Robustness checks are run for the 25th and 75th percentile.

⁴⁸ While the multivariate statistical analyses use the disaggregated three dimensions of fragility, they are not run with the different types of fragility as we have too few observations for some values of certain types.

Figure 4 State fragility and evaluation quality, and possible mechanisms linking them

Source: own figure.

In order to test our hypotheses, we employ a mixed-method design combining quantitative and qualitative analysis (Creswell, 2014). A quantitative analysis explores if fragility is relevant for the quality of evaluation reports. Subsequent expert interviews generate insights on hypotheses that cannot be tested quantitatively, and identify challenges and lessons learnt for future evaluation. Triangulation of data and methods – comparing the findings across different data sources and methods applied – allows us to estimate the validity of statements and findings.

The sample used in the quantitative analysis consists of 388 evaluation reports by the GIZ and KfW submitted to the BMZ between 2006 and 2016.⁴⁹ It includes project data such as information on monitoring, project duration, or success rating according to the five DAC criteria. Moreover, it contains details of the evaluation such as the type of evaluation and an assessment of the evaluation quality of each report (Noltze et al., 2018b). To this, we add the data on our explanatory variable (fragility) and the variables operationalizing the mechanisms (mobility, trust, M&E capacity), as described above. To estimate the influence of fragility on evaluation quality, we rely on descriptive and inferential analysis. For the inferential analysis we apply linear regression, penalized logistic regression and mediation analysis.

⁴⁹ Starting from a dataset of 513 observations, several observations covering more than one country are excluded for this country-level analysis. Moreover, some observations are dropped due to the merging processes of different databases and variables. This dataset varies slightly from that used in this chapter on project success, since for this analysis we rely on a 50% sample, because only those reports for which information on evaluation quality was available could be included. Nevertheless, the source of both datasets is the same (GIZ and KfW evaluations sent to the BMZ between 2006 and 2016, see above). For details on population, sampling strategy, and dataset preparation, see Noltze et al., 2018b.

The qualitative analysis is based on five semi-structured interviews involving explorative questions. We chose interview partners based on two criteria. The first was the role of the institution in the field of evaluation and fragility. Our aim was to choose experts covering the perspectives of implementing organizations concerned (GIZ and KfW), research (Global Public Policy Institute), and practitioners (consultancies: GOPA – Gesellschaft für Organisation, Planung und Ausbildung; and GFA – Gesellschaft für Agrarprojekte in Übersee). The second selection criterion was at the personal level: all experts were required to have at least five years of experience in the field of evaluation in fragile contexts and/or work with related questions on a daily basis. Based on the coding of the interview protocols, we then conducted a content analysis (Flick et al., 1995; Kuckartz, 2014).

The empirical analyses face a number of methodological challenges. First, the quality criteria may not fully reflect all facets of evaluation quality in fragile contexts. However, the quality criteria were carefully chosen to obtain a valid measure of evaluation quality. Second, the analysis draws on secondary data. We did not conduct case studies but included practical experiences through expert interviews. Third, different quality criteria might apply for different types of evaluation. To address this issue, only those quality criteria were selected that are commonly applied in the respective guidelines for the different evaluation types. Fourth, multiple evaluators are used to code evaluation quality. This involves the risk of coders disagreeing on the rating. The intercoder reliability analysis revealed strong, though not perfect, consistency (see Landis and Koch, 1977) with a Kappa coefficient of 0.63.⁵⁰ Finally, the sampling strategy originates from a different research question, raising the question of representativeness. However, the sample was randomly drawn from the population stratified by evaluation type (about half of reports of each evaluation type were sampled).

⁵⁰ For this analysis, we applied Cohen's Kappa coefficient. A Kappa value of 0 depicts maximal divergence between evaluators or coders, while a value of 1 represents perfect concordance. In addition, we included an external perspective. For this, we compared the quality assessment from the meta-evaluations by DEval and GIZ. Results showed that the maximum scores reached were similar.

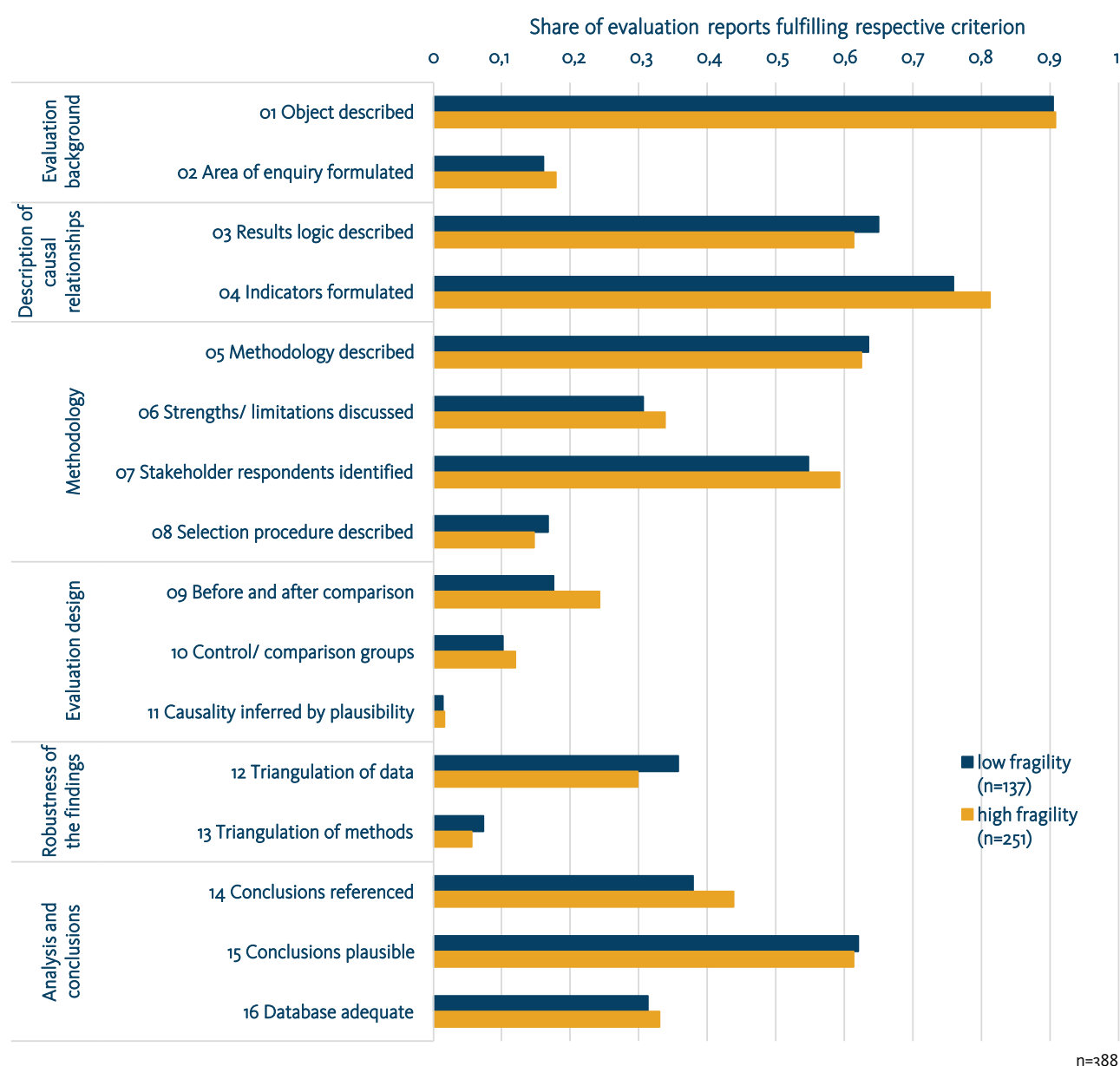
4.4 Empirical results

This section presents the findings of the quantitative and qualitative analyses.

Quantitative analysis

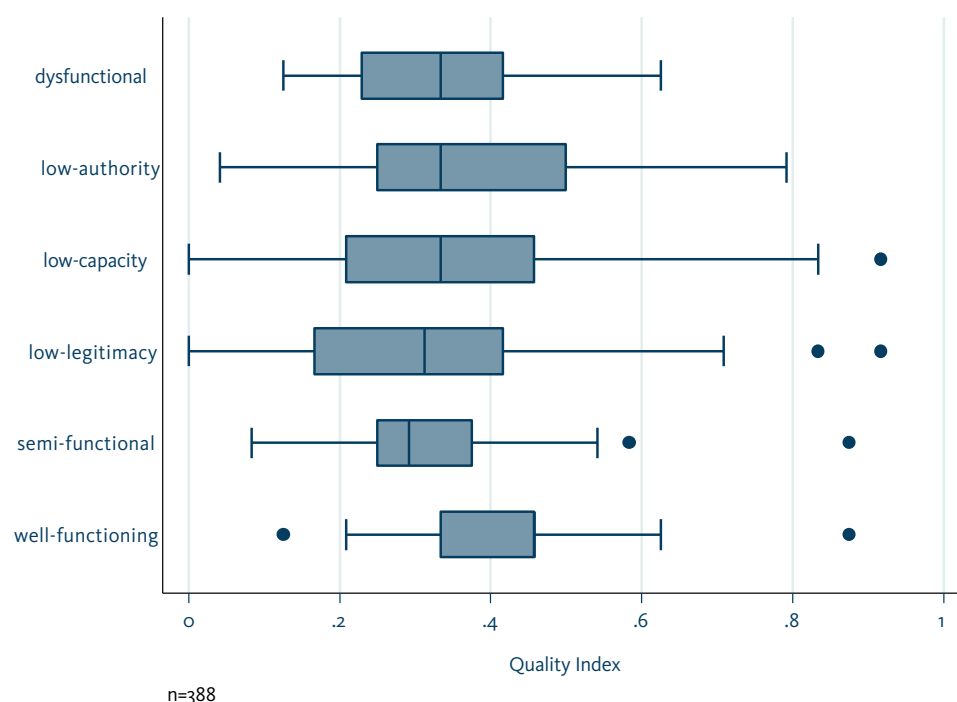
We begin with an inspection of evaluation quality depending on a binary variable fragility status (“high” or “low”) and by type of fragility. Findings suggest that the quality of evaluations does not systematically vary conditional on the fragility of the country in which the evaluation was conducted.

Figure 5 Fulfilled quality criteria by high or low fragility context



Source: own figure.

Figure 5 and Figure 6 show the descriptive and univariate results of the analysis. Figure 5 shows no big differences in evaluation quality between countries with low and high levels of fragility. Similarly, Figure 6 does not show large variations in the aggregated quality index between countries of different types of fragility. We apply the Kruskal–Wallis test to examine the null hypothesis that the means of all fragility types are equal, and find that it could not be rejected at the level of 5% or below.

Figure 6 Boxplots depicting the evaluation quality index differentiated by type of fragility

Source: own figure.

We now turn to the results from the multivariate regressions. The dependent variables are the different quality criteria. Since the dependent variable is a dummy variable of one quality criterion, we run penalized logistic regressions (Heinze and Schemper, 2002). We also run linear regressions with the aggregated quality index as dependent variable. The key independent variables are different dimensions of state fragility (i.e. authority, capacity and legitimacy). We control for sectors where aid could plausibly affect fragility levels and the quality of the evaluation simultaneously. We also control for the number of days that evaluators spent in the field, as this might influence evaluation quality.

The results on the aggregated quality index show no clear evidence that fragility is an important determinant of evaluation quality. We only find statistically significant associations between the fragility dimension of capacity and two evaluation quality criteria when we look at them separately. We find that a higher state capacity is negatively associated with the likelihood that indicators were formulated (Q-o4) and with the likelihood that strengths and limitations were discussed (Q-o6). Table 11 summarizes the results.⁵¹ The regression tables are included in the online appendix. However, we refrain from interpreting these results due to multiple hypothesis testing.⁵² The results lend further robustness to our overall finding of no significant relation between fragility and evaluation quality.

⁵¹ We assess the results of all regressions at the 5% significance level ($p < 0.05$).

⁵² Given a p-value of 5% or less, the chance of falsely rejecting a true null hypothesis of no effect is 5%.

Table 11 Summary results from multivariate regression analysis of the effect of fragility on evaluation quality

Outcome	Independent Variables			
	Fragility Index	Authority	Capacity	Legitimacy
Q-01 Object described
Q-02 Area of inquiry formulated
Q-03 Results logic described
Q-04 Indicators formulated	.	.	–	.
Q-05 Methodology described
Q-06 Strengths/limitations	.	.	–	.
Q-07 Stakeholder respondents identified
Q-08 Selection procedure described
Q-09 Before and after comparison
Q-10 Control/comparison groups
Q-11 Causality inferred by plausibility
Q-12 Triangulation of data
Q-13 Triangulation of methods
Q-14 Conclusions referenced
Q-15 Conclusions plausible
Q-16 Database adequate

Source: own figure.

We also conduct mediation analyses to investigate whether fragility is related to the quality of evaluation reports through the three selected potential causal mechanisms. A mediation analysis allows us to assess whether fragility affects the quality of evaluation reports through intervening variables, as there might be mediation effects even if the direct effect is not significant. Thus, mediation analysis allows indirect effects that cancel out each other to be revealed.⁵³ We use proxies to investigate the potential causal mechanisms of mobility, trust, and project M&E, as outlined above. Interpretation of mediation analyses is facilitated by conceptualizing the treatment variable as binary. Thus, we created binary independent variables for authority, capacity, and legitimacy. To assess robustness, we also employ cut-off points at the 25th and 75th percentiles. Significant results are summarized in Table 12.⁵⁴

Our analyses suggest that the most important mediator of the effect of fragility on evaluation quality is mobility. The left-hand column in Table 12 shows the three dimensions of fragility. The second column depicts the potential causal mechanisms (mediators), and the third column shows the quality criteria. The three columns on the right-hand side differentiate three different cut-off points between the “high” and

⁵³ See Hayes (2009) and Imai (2011) for more details on mediation analyses.

⁵⁴ For simplicity, null findings are excluded from the table and are available upon request.

“low” value of the respective fragility dimension of the first column. The symbols show statistically significant effects. Those criteria to which the respective fragility dimension is robustly related are shaded in yellow. Regression tables are provided in the online appendix.

Table 12 Summary results of the mediation analysis

Combination			Cut-Off in Percentage		
Independent variable	Mediator	Dependent variable	25%	50%	75%
Authority	Mobility	Q-16 Database adequate	.	+	.
Capacity	Mobility	Q-14 Conclusions referenced	.	+	+
Capacity	Mobility	Q-16 Database adequate	+	+	+
Capacity	Mobility	Q-15 Conclusions plausible	.	+	+
Legitimacy	Trust / Social Capital	Q-05 Methodology described	+	+	+
Legitimacy	Mobility	Q-08 Selection procedure described	.	-	.
Legitimacy	Mobility	Q-16 Database adequate	-	-	-
Legitimacy	Trust / Social Capital	Q-09 Before and after comparison	-	-	-

Legend: + = positive mediation effect; - = negative mediation effect; . = no mediation effect

Source: own figure.

In respect of authority, we find evidence that higher levels of authority increase the likelihood that the data and methodology of evaluations is qualitatively and quantitatively sufficient (“database adequate”) through increased mobility.⁵⁵

On capacity, we find that higher levels are positively related to the likelihood that conclusions in the evaluation report are plausible (Q-15) through mobility. Some 42% of the total effect is mediated through mobility.⁵⁶ Capacity is also positively correlated to Q-14, i.e. whether conclusions were referenced. This effect is again mediated through mobility; in fact, 37% of the total effect is mediated through mobility.⁵⁷ There is also a conditional effect of capacity on whether the database of an evaluation report is adequate (Q-16) through mobility, with some 47% of the total effect being mediated.

The direction of the relationship between the fragility dimension of legitimacy and different quality criteria is not as consistent as is the case for capacity. On the one hand, the effect of legitimacy on evaluation quality is positively mediated through our measure of trust among the population. While we find no evidence of a direct effect of legitimacy on whether the evaluation methodology was described (Q-05), 67% of the total effect is mediated through trust. A higher level of legitimacy is associated with a higher likelihood of evaluation criterion Q-05 being fulfilled through higher mobility. There is no evidence of a direct effect of legitimacy on whether a before-and-after comparison was undertaken (Q-09); but 58% of the total effect is mediated through trust. Interestingly, higher legitimacy seems to reduce the likelihood that the evaluation was based on before-and-after comparisons. This might be due to the fact that other rigorous types of

⁵⁵ However, this effect is not robust to using the 25th percentile cut-offs for the binary authority variable.

⁵⁶ While this effect is robust to a number of different model specifications, the mediation effect is not robust to using the cut-off of the 25th percentile for the binary fragility variable of capacity.

⁵⁷ While this effect is robust to a number of different model specifications, the mediation effect is not robust to using the cut-off of the 25th percentile for the binary fragility variable of capacity when controls are included in the model.

research design are used in such contexts of high trust among the population.⁵⁸ Legitimacy has a positive direct effect on whether the database in an evaluation was considered as adequate (Q-16). In addition, 40% of this effect is mediated through mobility. Again, higher values of legitimacy are associated with lower quality.

Qualitative analysis

The results of the quantitative analysis are complemented by a qualitative analysis of five expert interviews, covering all hypotheses. Furthermore, it provided the opportunity to obtain first impressions of best practices. However, since we only conducted five interviews, the results must be interpreted with care.

The hypothetical mechanism on reduced mobility and poor project M&E capacities received the strongest support from experts. The main mobility challenges were travel restrictions and poor infrastructure. Project M&E is related to various factors such as resource questions – in terms of budget as well as availability and competence of staff – leading to lacking or poor-quality indicators and other project data available for external evaluators.

As listed in Figure 4, mechanisms 2 (trust among the population), 3 (affirmative culture) and 5 (evaluator's influencers) reflect power relations, interests, and the motivational situation of the various actors in the spheres (the population, the project staff and the evaluator or evaluation team). On the other hand, mechanisms 1 (mobility), 4 (project M&E) and 6 (evaluator's M&E capacity) explore the capacity side of the actors, which is closely linked to resource questions.

All experts confirmed that they expect fragile contexts to have a negative impact on evaluation quality. For instance, as external actors, evaluators have specific room for manoeuvre, but they also rely on the information provided by the project. This is even more true in the case of travel restrictions, which may even prevent evaluators from travelling to the respective country. Mechanism 4, suggesting that evaluation quality may be reduced due to a project's poor M&E systems, was supported by all five experts.⁵⁹ Mechanism 1 was also confirmed by all experts. Mechanisms 2, 3 and 6 were predominantly supported. Mechanism 5, on the other hand, was supported only by two of the five experts. In contrast to the quantitative analysis, the qualitative analysis supported all mechanisms except mechanism 5 as being relevant to evaluation quality in fragile contexts.

The most examples and specific challenges were expressed concerning the realm of project M&E capacities and mobility. The other four mechanisms were much less strongly supported by examples. This is in line with the findings of the quantitative analysis, suggesting that mobility and trust are relevant determinants for evaluation quality.

As for mechanism 1, most interviewees mentioned travel restrictions imposed by the sending organization in Germany. Poor infrastructure in the partner country (such as transportation or communications) was mentioned less often. A denial of travel by the headquarters of the German sending institution may result in a visit by project staff to a safer neighbouring country or a desktop study in Germany. In the event that the evaluator or evaluation team travels to the partner country despite security and infrastructure concerns, pragmatic decisions might be taken (e.g. "You visit those schools that are accessible"), or local consultants may visit those regions which are not available to international evaluators due to security restrictions.⁶⁰ On the other hand, the cost of implementing security provisions may increase, which may lead to an evaluation team deciding not to travel if the costs of making the journey safe exceed the funds available.

⁵⁸ It would be plausible for this effect to be driven by evaluation reports that do not fulfil the criterion of before-and-after comparisons having been based on control/comparison groups. However, this does not seem to be the case.

⁵⁹ For the purpose of this analysis, a hypothesis or mechanism was regarded as confirmed if the interviewee mentioned at least one explaining factor or supporting example illustrating the mechanism.

⁶⁰ The strategy of sending local staff into a region that is not accessible to the evaluators as planned was mentioned by almost all experts in the interviews, but was also controversial, weighing both normative and pragmatic factors.

Mechanism 2 focuses on trust issues among the population. If the level of trust in a society is low, the population's response behaviour may be negatively influenced. Experts explained that this could limit information-sharing by interviewees, but may also cause a systematic bias, e.g. by pressure to respond according to the narrative of the specific group the respondent belongs to, or to follow the conflict logic.

For mechanism 3, an affirmative project culture was confirmed by the experts, but backed up with fewer examples. Experts pointed out that an affirmative project culture may influence what topics are raised and what evaluation questions are asked. Moreover, if the evaluation team relies on support from the project when organizing the evaluation mission, it might also affect case selection. One explanation mentioned was that project staff may see evaluators as controlling inspectors rather than facilitators of a learning process. Nevertheless, internal reasons such as incentives/"boni" were also mentioned, exacerbating the situation. In addition, project staff may find themselves in insecure situations, e.g. running the risk of being expelled from the country.

The experts put much emphasis on mechanism 4 (project M&E capacity). Among various aspects, they mentioned data availability as a core challenge, along with the poor quality of indicators (e.g. they may not refer to the project proposal or may not be adapted to the context) and hence low data quality. As possible causes, experts mentioned a lack of qualified staff and other resources for M&E tasks in the project, which may be caused by funding rivalry between the implementation and evaluation of projects, usually financed from the same budget. Moreover, especially in fragile contexts, projects may operate under "high pressure", often coping with a situation of high demands and urgency (e.g. the management of a hospital) and limited resources, which impedes the dedication of resources to M&E tasks. In addition, this situation may increase turnover among project staff, negatively affecting its level of competence and ownership. Other aspects mentioned to a lesser extent were gender-blind staffing, poor leadership support for the evaluation within the project, and low M&E orientation among project staff and no follow-up by them.

Examining mechanism 5 (evaluator's influencers), internal issues in the institution employing the evaluator (e.g. pressure not to be too critical) outweighed local context influences such as corruption.

On mechanism 6 (evaluator's M&E capacity), the factor mentioned most often was inadequately qualified personnel, followed by lack of knowledge about the context. These observations were usually accompanied by remarks about tight resources combined with a demanding environment. This led to another issue mentioned: that guidance by the project is crucial for successful evaluation, especially in fragile contexts.

4.5 Results and recommendations

This chapter has investigated how fragile contexts may influence evaluation and its quality, how potential challenges can be addressed, and how evaluation practices in a fragile context can be improved. The following summarizes our findings and concludes with a recommendation.

In a first examination of the direct correlation between fragility and overall evaluation quality, results show that, at a general level, evaluation quality does not vary greatly depending on fragility. The same is true for the individual evaluation quality criteria. We can thus conclude from the quantitative analysis that, overall, even in dire circumstances evaluators manage to maintain evaluation quality at normal levels. This finding is surprising as qualitative findings suggest that fragility may affect evaluation quality in various ways.

In a second step, we explore the role of different mechanisms as mediators between fragility and evaluation (quality). This reveals that low state capacity is associated with reduced mobility and thus indirectly with a lower evaluation quality.

According to the qualitative analysis, the core challenges to fully fledged quality evaluations in fragile contexts were: poor data availability and project M&E; limited mobility due to travel restrictions; the image of evaluators among the population; adverse internal incentives for project staff and, to a lesser extent, evaluators; and issues around the qualifications and contextual knowledge of evaluators. Experts furthermore emphasized in the interviews that, apart from these challenges, more pressure at the level of both project and evaluation is added by the increasing speed of project implementation and shortened

project cycles, as well as an orientation towards outputs rather than outcomes. Consequently, evaluators employed different coping strategies, e.g. they adopted expectations and benchmarks.

The findings of both the quantitative and qualitative analyses show that the mobility of evaluators and their (logistical) access to interview partners is a core issue in fragile contexts. We further conclude that, across quantitative and qualitative findings, the evidence is mixed concerning the role of societal trust and project M&E capacities. While in the quantitative analysis, respective variables are not significantly associated, experts emphasized the special importance of M&E systems at the project level.

Mobility restrictions, and thus access to field data, in fragile states are a challenge for evaluations (see OECD, 2006). Often, security challenges mean that mobility and thus access to stakeholders becomes more expensive. It is therefore important that monitoring and evaluations are sufficiently, separately (i.e. from project implementation), and securely funded.

A high M&E capacity of projects ensures availability and quality of project data. A critical review of M&E systems could be useful to identify potential to improve data quality and availability. This could include reserving sufficient resources for M&E tasks at project/programme level, incentives such as M&E-capacity-related annual objectives for leadership personnel, or sensitization training specifically for leadership personnel. Moreover, the evaluation capacity of local partners could be improved by creating a sustainable pool of evaluation experts in partner countries.

Possible approaches to increase M&E capacity are budgeting and project planning. With regard to budgeting, separate budgets for project implementation on the one hand and M&E on the other should limit competition for resources. This has already been recently implemented in respect of evaluations. For instance, the GIZ has a separate budget for financing evaluations. According to the recent GIZ evaluation reform, project evaluations will now be more centralized, being commissioned by the GIZ headquarters evaluation department. However, there is room for further improvement. With regard to project monitoring, module proposals (*Modulvorschläge*) for technical development cooperation require reporting of expected costs differentiated by output. However, monitoring costs are not separately reported, but either subsumed under the respective output or labelled as “overall costs”. Separate reporting of monitoring costs could improve the BMZ’s steering capacity in respect of the quality of monitoring and evaluation. For example, the share of costs allocated to monitoring and evaluation could be adapted depending on expected challenges of M&E in the operating environment. Moreover, implementing organizations would be required to inform the BMZ in the event of major budget adjustments.

Another possible remedy that specifically addresses mobility challenges would be to further strengthen the capacity of conducting remote monitoring and evaluation (see Glass and Schönhofen, 2018). Suitable approaches to data collection include remote sensing, digital data collection by handheld devices, and remote surveys (BMZ and GIZ project Internet and Sustainable Development, 2016; Lech et al., 2017, 2019; Steets et al., 2016). To make use of these new types of digital data in decision-making processes, it is important to strengthen the capacity to apply and process data (Pawelke et al., 2017).

Finally, in order to improve the overall quality of evaluations, the BMZ, GIZ and KfW should continue to implement the recommendations in DEval’s meta-evaluation on sustainability (Noltze et al., 2018c).

We thus make the following recommendation:

Recommendation 4

Based on qualitative interviews, observational evidence from an evaluation synthesis, and a screening of the literature, this evaluation recommends that the BMZ, GIZ, and KfW address mobility restrictions on evaluators in fragile contexts and improve M&E at the project level.

First, we recommend that the BMZ require the GIZ and the KfW to further separate out budgets for project implementation on the one hand, and project monitoring and evaluation on the other, to further strengthen the prevention of fungibility and competition for resources.

Second, this evaluation recommends that the GIZ and the KfW address mobility challenges in fragile contexts through the increased use of innovative data-collection methods. This includes strengthening capacity in conducting remote evaluation and monitoring using, e.g., geospatial methods, remote surveys, and digital data collection.

5. DEVELOPMENT INTERVENTIONS IN FRAGILE CONTEXTS

5.1 Introduction

Fragile contexts are challenging environments for development cooperation. This chapter synthesizes the results of 471 digitized evaluation reports on GIZ and KfW projects in 85 countries between 2006 and 2015 to compare project ratings across different levels of fragility.

To assess the success of development cooperation in fragile contexts, we draw on two measures of fragility. First, we use the multidimensional yet state-based concept of state fragility by Grävingholt et al. (2018). This appears most effective in measuring state fragility, as discussed in Chapter 2. Second, we develop a subnational measure of fragility based on violent events in the vicinity of georeferenced locations. Linking georeferenced aid projects to violent events provides a clearer picture of subnational variations of fragility.

We find fragility to be mostly uncorrelated to project success as measured by evaluation ratings. State capacity, however, is an exception in that more successful development cooperation projects reside in countries with higher levels of capacity.

5.2 What we know about the success of development cooperation

There is an extensive literature on the success of development cooperation projects (for an overview, see Arndt et al., 2010; Doucouliagos and Paldam, 2009; Faust and Leiderer, 2008). As a rather broad phenomenon, state fragility is linked to many factors that have been discussed in terms of the success of development cooperation. The following reviews this literature using four groups of factors: 1) relations between extending agencies, implementing agencies, and partner countries; 2) donor characteristics; 3) characteristics of the partner country; and 4) characteristics of the intervention.

Relations between extending agencies, implementing agencies, and partner countries

Development cooperation involves a long chain of delegation. Taxpayers provide funds to governments, which engage implementing aid agencies (e.g. GIZ, KfW) that provide instruments of development cooperation to partner countries, possibly indirectly through contractors.⁶¹ These can be thought of as principal–agent relationships (Killick, 1998; Martens et al., 2002).⁶²

Concerning development cooperation in fragile contexts, relationships between governmental and non-governmental aid agencies (principals) on the one hand and partner-country governments as well as field agents (agents) on the other are of primary interest. Lack of information for taxpayers, the legislative and executive branches of government, and the implementing aid agencies allows field agents to pursue their own interests. In this scenario, tighter control through monitoring and close management might appear an appropriate means to mitigate potential harm from information asymmetry.

However, greater autonomy of field agents might also have beneficial effects, particularly in fragile contexts. Honig (2018) finds that in unpredictable environments tight top-down management of field agents is less effective than reduced principal control. He argues that, although fragility has an overall negative effect on project success, this can be mitigated by giving field agents greater autonomy of decision-making to react more flexibly in volatile environments.

A second relational characteristic is donor interest (Alesina and Dollar, 2000; Dreher et al., 2015; Minoiu and Reddy, 2010). Whether allocation decisions are made based on the donors self-interest or on the needs of

⁶¹ In Germany, financial development cooperation usually works through contractors, while technical development cooperation works directly with recipients.

⁶² Principal–agent theory models hierarchical relations between a delegating entity, i.e. the principal, and a subordinate entity selected to execute the delegated tasks, i.e. the agent. Principals and agents are modelled as utility-maximizers under incomplete information. The theory argues that, in these circumstances, conflicts of interest and information asymmetry might allow agents to act in their own rather than the principals interest, leading to suboptimal results. The respective role of the involved actors depends on the relationship under study. For instance, beneficiaries and partner countries might act as principals, e.g. towards aid agencies and contractors, or agents, e.g. in cases of conditionality contracts.

the partner countries influences the quality of development cooperation interventions (Knack and Rahman, 2007).

Donor characteristics

A second group of characteristics that might influence project success involves donor countries. Minasyan et al. (2017) draw on the index of donor performance from the Center for Global Development to show that higher-quality foreign aid – aid maximizing efficiency, fostering institutions, reducing the burden on partner countries, and enhancing transparency and learning – leads to income gains in recipient countries (Roodman, 2012). In 2014, Germany ranked 14th in the index of donor performance. This index identifies a comparatively large share of aid going to well-governed countries as a strength of German ODA. More recently, Mitchell and McKee (2018) rank Germany lowest out of 27 state donors and 13 multilateral agencies. However, they find development cooperation to well-governed countries to be successful.

Faust et al. (2016) find that, from a partner perspective, large multilateral donors and small specialized bilateral donors tend to be more influential and helpful than large bilateral donors. Findley et al. (2017) argue that multilateral and non-state donors are driven less by self-interest and focus more on the needs of partner countries. Looking at state donors, Faust (2008) and Dietrich (2016) investigate the effect of domestic political and economic institutions in donor countries on the effectiveness of aid.

Characteristics of the partner country

A broad strand of the literature argues that good policies in partner countries are decisive for the success of development cooperation (Burnside and Dollar, 2000, 2004; Collier, 2008). However, the robustness of the results has been questioned (Easterly et al., 2004; Rajan and Subramanian, 2008).

A similar argument concerns the institutional quality of partner countries (Collier and Dollar, 2002), political stability (Chauvet and Guillaumont, 2003), and the type of political regime (Svensson, 1999). With regard to the latter, democracy might lead to greater aid effectiveness as democratic leaders have greater incentives to create public goods (Bueno de Mesquita et al., 2003). More recent arguments have advanced the debate by using more refined regime typologies (Lührmann et al., 2017).

Characteristics of the intervention

Beyond donor and partner-country characteristics, the type of instrument may affect effectiveness. The literature differentiates between budget and project support as well as tied and untied aid (Dreher et al., 2008; Milner, 2006).⁶³ It has moreover investigated the role of the volume (Clemens et al., 2012) and fragmentation (Gehring et al., 2017; Knack and Rahman, 2007) of development cooperation.

When considering success in fragile countries, channels of funding have been identified as crucial. Generally, there are two: country systems and non-state systems.⁶⁴ The use of country systems for ODA supports the principles of aid effectiveness and is seen as good practice. However, the use of country systems might be risky, particularly in low-capacity states (Manuel et al., 2012). More generally, Manuel et al. (2012) and Honig (2018) argue that flexibility of aid instruments allows them to be constantly adapted to potentially changing environments, which seems to be crucial in fragile contexts.

Orth et al. (2017) find that a mix of four different components in budget support programmes (financial transfer, policy dialogue, conditionalities, and technical assistance) makes budgetary support more successful. Moreover, they argue that the consequent use of country systems improves public financial management in partner countries.

⁶³ In Germany, development cooperation is organizationally divided into technical and financial development cooperation.

⁶⁴ Manuel et al. (2012) discuss the funding channels of the six categories of ODA instruments according to the OECD's International Network on Conflict and Fragility (INCAF): (1) general budget support (country system); (2) sectoral budget support (country system); (3) government-managed pooled funds (country system); (4) jointly managed trust funds (parallel systems); (5) project support (parallel systems); and (6) support through and to non-state actors (parallel systems).

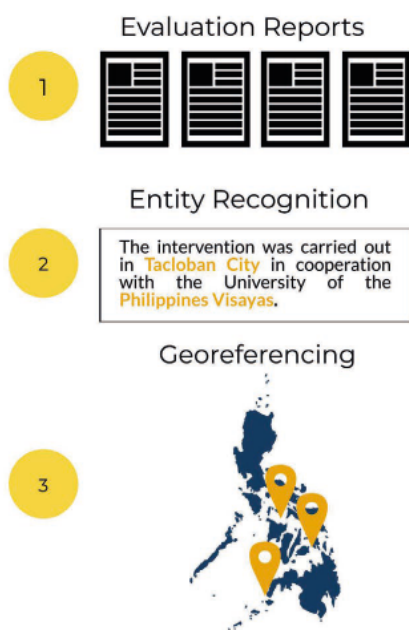
Besley and Persson (2011) use a game-theory model looking at the influence of types of development cooperation on the fragility of partner countries and project success. Their model allows empirical expectations to be derived on the role of cash- or non-cash aid as technical, military and post-conflict assistance. Cornell (2013) develops a measure of democracy aid to investigate the impact of development cooperation on democratization in autocratic regimes.

In summary, the academic debate finds the success of development projects to be conditional on 1) relations between extending agencies, implementing agencies, and partner countries; 2) donor characteristics; 3) characteristics of the partner country; and 4) characteristics of the intervention.

5.3 Data and method

Data

To assess the impact of fragility on project success, we rely on evaluation reports from the GIZ and KfW starting in 2006. Since then, evaluation draws on the standardized DAC evaluation criteria. To address the lack of independence between consecutive reports, we only keep the most recent evaluation for each project. Moreover, we exclude projects that span more than one country and projects in countries where no data on fragility is available. This leaves us with 471 projects conducted in 85 countries between 2006 and 2015.



Our outcome of interest is project success. We operationalize project success as the arithmetical mean of the ratings for the five DAC criteria: relevance, effectiveness, efficiency, impact, and sustainability. Where adequate, we also disaggregate by criterion. Thus, we do not investigate whether projects had an impact on fragility but look at success in fragile contexts.

As our main explanatory variable, we employ both a national and a subnational measure of fragility. To measure fragility at the national level, we draw on the Constellations of State Fragility (CSF) by Grävingholt et al. (2018). We include both the aggregated type of fragility and the three dimensions of fragility (authority, capacity, and legitimacy) as predictors (see Chapter 2).

A new measure of subnational fragility

The level of fragility often varies within states, e.g. between the centre and the periphery. The effectiveness of interventions may depend to a significant degree on these local conditions.

The use of macroeconomic indicators or national fragility scores is thus insufficient. Where precise geographical information on projects is available through evaluations, indicators assessing contextual factors should reflect a comparable level of precision. Most empirical analyses, however, use data that are aggregated to the country-year level, thereby losing project-specific information. Spatial disaggregation is thus a promising strategy to improve our knowledge of development cooperation in fragile contexts.

While current approaches to geocoding largely rely on close reading, we rely on text mining. We retrieve subnational location information from the evaluation reports through natural language processing. We then assign coordinates to each of these locations through Google Maps. This provides us with a clearer picture of distributional patterns of aid within partner countries.

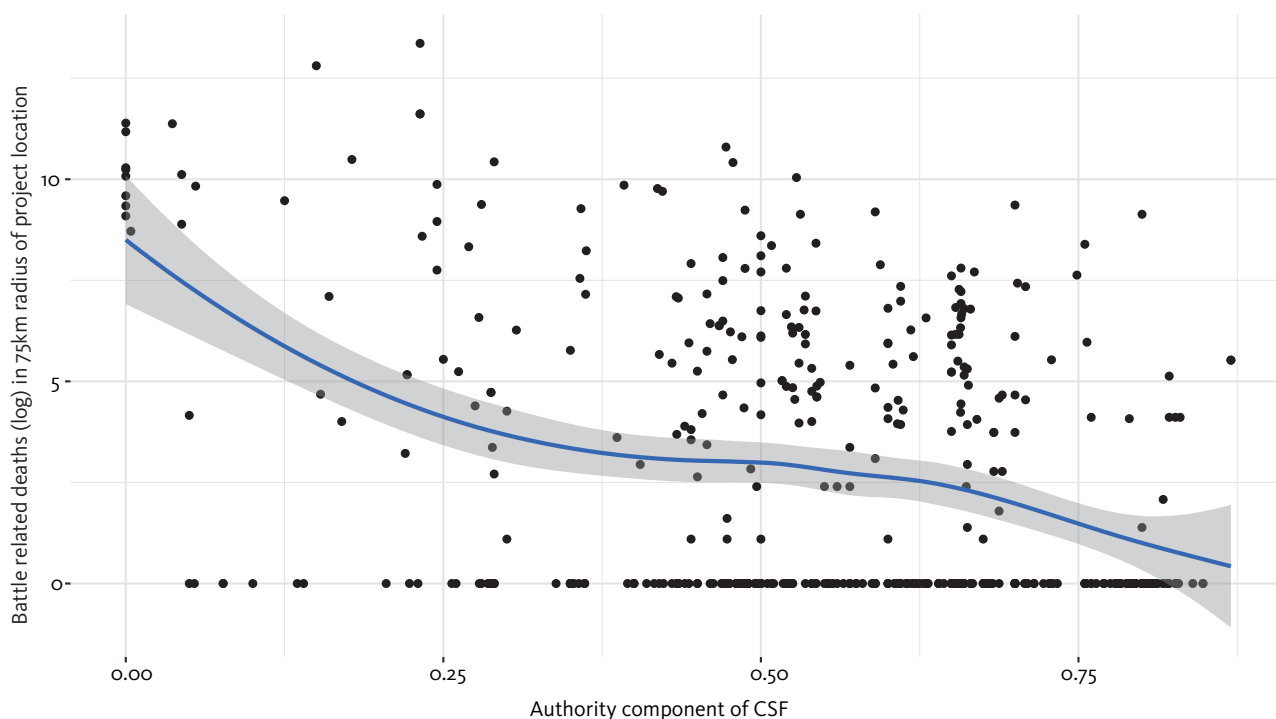
In order to improve the automated process of geocoding, we apply several algorithms to render our results more precise. These algorithms draw on (1) the status of a country as a partner for bilateral development cooperation, (2) the type of location, and (3) a cluster algorithm based on the distance of each location from the centroid of all locations. Moreover, we test for different specifications of our measure. For instance, we

include a measure that weighs results by the number of times a certain location is mentioned in the document. An illustration of the process is included in the online appendix.

To measure state fragility subnationally, we combine our data on project locations with data on conflict events from the Uppsala Conflict Data Program Georeferenced Event Dataset (UCDP GED) (Sundberg and Melander, 2013). We operationalize state authority as the number of battle-related deaths occurring in intrastate conflict within a 75-km radius around the location of projects. Due to the skewness of the data, we draw on the natural logarithm.

Figure 7 compares the distribution of the newly developed subnational measure of fragility with the authority dimension of the constellations of fragility dataset. The graph shows that, although there is a correlation between the two measures, there is great variability of subnational fragility. This indicates that our subnational measure can improve our understanding of state fragility. On the one hand, we observe many interventions conducted in peaceful local contexts (indicated by low values on the Y-axis) in states with low authority (indicated by low values on the X-axis). On the other, we observe interventions in regions with high levels of violence in states with comparatively high levels of authority.

Figure 7 Subnational fragility by CSF authority component



Source: own figure based on GIZ and KfW evaluation reports and UCDP GED.

Method

We combine descriptive and econometric analyses to investigate project success in fragile contexts. The descriptive analysis quantitatively synthesizes the ratings from the available evaluation report conditioned on various measures of fragility. In addition, we draw on regression analysis to account for the influence of possible confounding variables and to address possible imbalances in our sample. For more information, see the online appendix.

Most importantly, we should be cautious in interpreting the following results causally. Selection into the sample is not random, and unobserved confounding variables might influence the results. For instance, due to the assessment of risk and institutional capacity in the preparation stage preceding project implementation, it is likely that projects are tailored to the specific context. Consequently, projects in fragile contexts may be comparatively less complex, leading to better ratings in evaluation reports. Moreover, aid is not distributed randomly (Briggs, 2018). As we do not yet know enough about the mechanisms of aid

allocation, these issues of self-selection might bias the findings of the present analysis. However, the results in Chapter 4 suggest that evaluation quality does not vary greatly by type of fragility. Hence, evaluation quality does not seem to be a source of systematic bias in our outcome indicators.

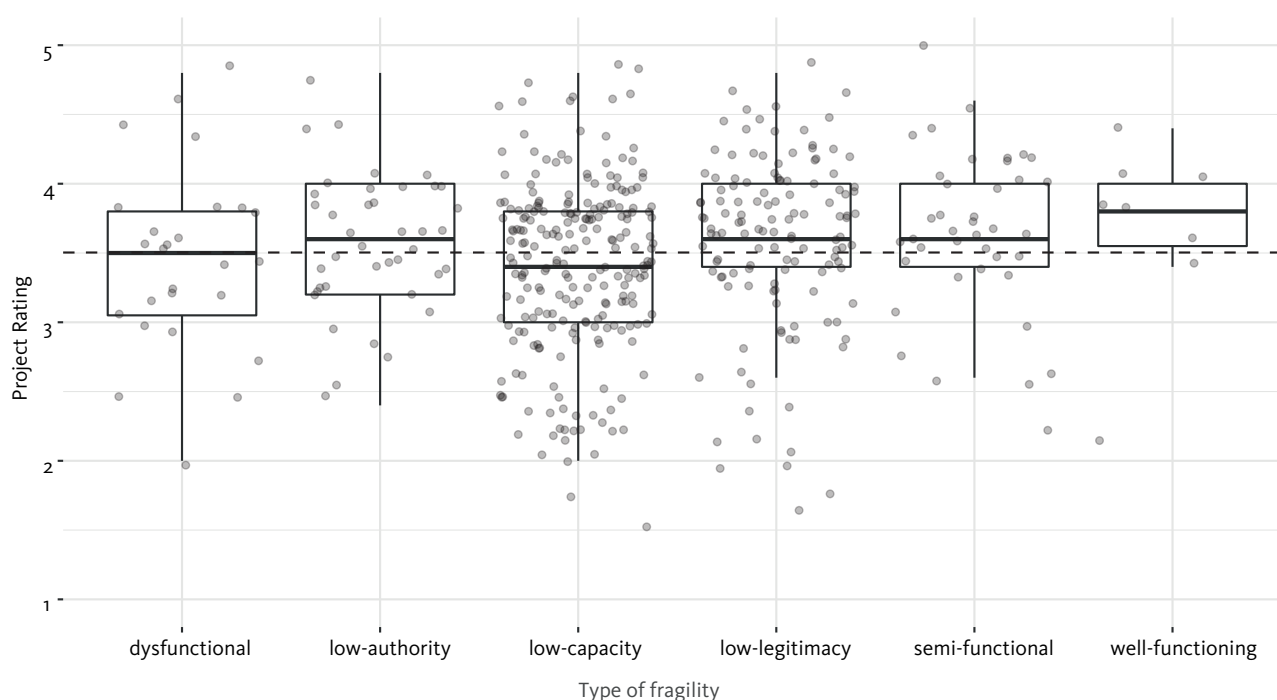
5.4 Empirical results

Results

Figure 8 plots project success rating by type of fragility. The horizontal dashed bar shows the arithmetical mean of project ratings across all evaluations. The graph indicates that the majority of projects take place in countries with low capacity and low legitimacy. Not surprisingly, projects in well-functioning states are rare as these states are typically not partner countries. We find no robust relationship between the type of fragility and project success. Although the figure shows that project ratings are somewhat lower in low-capacity states, we do not find robust evidence for lower ratings in our econometric analysis (regression results are included in the online appendix).

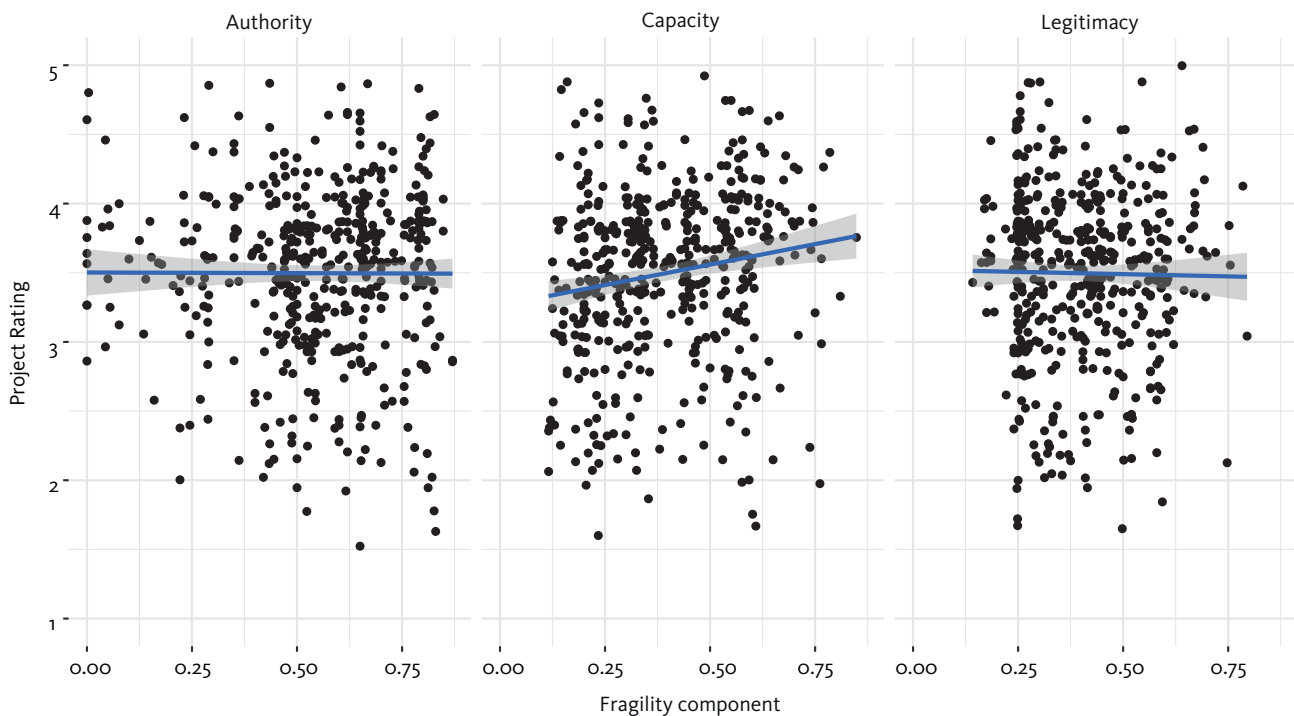
We further investigate how much ratings vary. It may be that volatile contexts lead to a greater variance of project outcomes. However, a comparison of the standard deviation of project ratings conditional on fragility does not show substantial differences. We thus find no evidence that project ratings differ according to the fragility of contexts as measured by types of fragility.

Figure 8 Project ratings by type of fragility



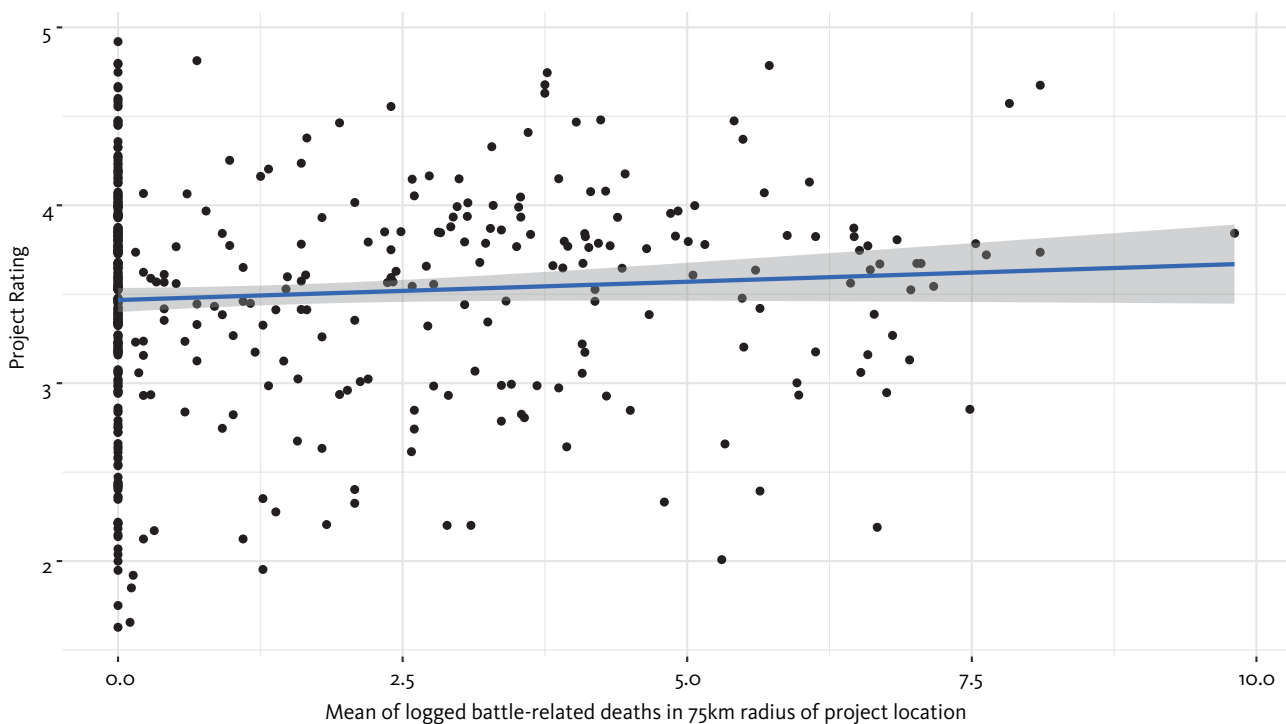
Source: own figure based on GIZ and KfW evaluation reports and CSF.

We can further investigate the relationship between fragility and project ratings by looking at the three dimensions of fragility. Figure 9 plots project ratings by authority, capacity, and legitimacy. We find no indication of a relationship between authority and legitimacy on the one hand and project ratings on the other.

Figure 9 Project ratings by fragility component

Source: own figure based on GIZ and KfW evaluation reports and CSF.

We can take a closer look at the relationship between authority and project ratings by drawing on our subnational measure of fragility. Figure 10 plots project ratings against our geographically disaggregated measure of authority. Again, we find no systematic correlation between state authority in the vicinity of project locations and the ratings of success.

Figure 10 Project rating by subnational fragility

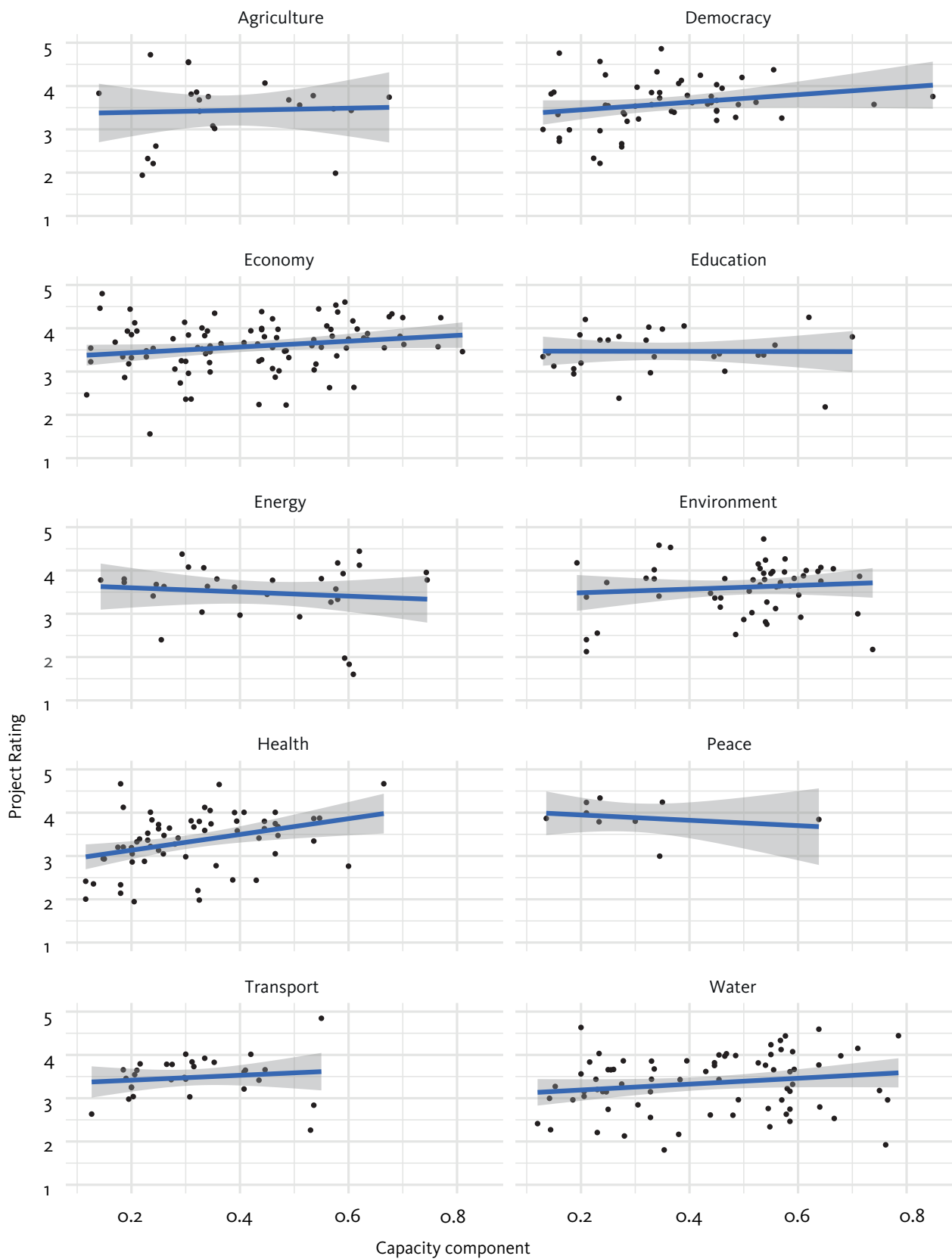
Source: own figure based on GIZ and KfW evaluation reports.

In contrast to authority and legitimacy, we find project ratings positively correlated with state capacity. This result appears robust in our econometric analysis and for different specifications of capacity such as taking state capacity in the last year of the intervention or the mean capacity rating over project time (see online appendix for regression results). Moreover, the effect is similar for technical and financial development cooperation.

A closer look at the descriptive data shows that the positive correlation between project ratings and state capacity is primarily driven by the ratings for the DAC criteria of effectiveness, impact, and sustainability and much less by ratings for efficiency and relevance.

We can further investigate the variation of project ratings by looking at different sectors. Figure 11 shows project ratings by sector according to the capacity dimension of state fragility. Based on the descriptive data, we find that the positive relationship between capacity and project rating is strongest for projects in the health sector.⁶⁵ These results should be treated with caution, however, due to a possible problem of endogeneity: the authority dimension of CSF includes child mortality as an indicator (see Chapter 2).

⁶⁵ Due to the low number of observations in each sector, we only test our results in a reduced regression model controlling for the implementing organization, type of evaluation, and other components of fragility. We find positive coefficients for the health, democracy, economy, and water sectors, although results are only significant for the health sector. But these results should be treated with caution. For instance, in the health sector, we only have 54 observations and, in light of the earlier results, we do not expect a significant effect. Consequently, a regression approach would probably be underpowered.

Figure 11 Project rating by state capacity in different sectors

Source: own figure based on GIZ and KfW evaluation reports and CSF.

Discussion

The empirical results show that state capacity correlates with higher project success in terms of effectiveness, impact, and sustainability. The correlation between state capacity and project success ratings is particularly strong in the health sector. Below we discuss possible implications in respect of state capacity development.

Capacity development is usually understood in rather broad terms encompassing all three dimensions of fragility, namely authority, capacity, and legitimacy (GIZ, 2015). The disaggregated approach to fragility in this evaluation suggests that capacity development aimed at improving bureaucratic quality and the ability of the state to provide public services to its population is key. In the following, capacity development refers to this narrower understanding.

Capacity development takes time (Henning, 2017). Nonetheless, the empirical results in this chapter suggest that it does pay off in terms of higher project success ratings. This seems to be particularly true for interventions in sectors that rely on high bureaucratic quality and infrastructural strength in the partner countries. The ability of states to implement policies based on an efficient administration that works throughout the territory thus appears to be decisive for the success of development cooperation. This underlines the importance of long-term institution-building in contexts of fragility, where, at the same time, the need to achieve results quickly might be particularly important.

The finding that greater state capacity correlates with greater project success reveals a possible conflict of objectives. On the one hand, low-capacity states are probably among those in greatest need of external support. On the other, development interventions in countries where state capacity is already high hold out the prospect of being more successful. In light of our findings, this conflict of objectives could be resolved by prioritizing capacity-building measures in low-capacity partner states as this should increase the success of development interventions. One possible approach might be via reform partnerships (BMZ, 2017b). Most countries that entered into reform partnerships were classified in 2015 as low-capacity states (Grävingholt et al., 2018).⁶⁶

The results in this chapter must also be seen in the context of decreasing commitments to non-state actors (see Chapter 3). Where states lack capacity, donors are inclined to rely on channels of delivery that bypass the state. This in turn might undermine efforts to increase state capacity, possibly threatening the success of development cooperation in the long run. At the same time, building state capacity is an endogenous process that needs to be conceived as a social relation between the state and society (OECD, 2011b). This implies that improving state capacity necessarily links the central state to the local level, or at least supports both (see also the discussion on the local turn in Chapter 3.)

Moreover, we find the correlation of state capacity and success of development cooperation to be sector-specific. Although we should, as discussed above, be cautious in interpreting these results, they resonate well with quantitative and qualitative evidence. Holmberg and Rothstein (2011) find that it is the quality of government rather than spending on healthcare that leads to higher life expectancy, higher subjective health, and lower child mortality. Based on two case studies, Briebe (2018) argues that better health outcomes can be explained by public health systems that are well organized and uniformly distributed across a country's territory. More generally, he suggests that it is specifically bureaucratic quality and infrastructural strength that link state capacity to better development outcomes. Our results point in a similar direction. Interventions in the health sector in particular seem to achieve better ratings with increasing state capacity. For other development outcomes, we find some evidence of a positive relationship in the democracy, economy, and water sectors. Interestingly, there is no evidence of a similar positive

⁶⁶ The predicted probability of being classified as a low-capacity state in 2015 for the countries that entered reform partnerships was as follows: Côte d'Ivoire 99%, Ethiopia 98%, Ghana 80%, Senegal 77%. Morocco and Tunisia, by contrast, are classified as low-legitimacy states with a probability of 95% and 56%, respectively.

relationship in education, which has traditionally been a sector heavily dependent on state capacity (Rajkumar and Swaroop, 2008).

5.5 Results and recommendations

This chapter sheds new light on the nexus between fragility and project success by introducing two innovative features that should improve our understanding of the role of state fragility. First, the analysis draws on a multidimensional measure of state fragility by Grävingholt et al. (2018). Rather than classifying states as fragile or non-fragile, this allows six types and three dimensions of fragility to be differentiated (see Chapter 2). Second, this chapter employs an innovative geographically disaggregated measure of fragility by retrieving information on subnational locations from the evaluation reports through natural language processing and linking them to violent conflict events. This approach acknowledges the fact that state fragility typically does not affect the whole country to the same extent. In developing countries in particular, state fragility varies spatially. Nonetheless, state fragility is often exclusively conceived as a measure at the country level. Georeferencing aid projects and assessing the local situation provides a more realistic picture of the context in which development cooperation takes place.

The following summarizes the results, discusses mechanisms and methodological shortcomings that might explain these results, and discusses possible implications for the strategic management of development cooperation. A single recommendation concludes the section.

Project success ratings do not vary by type of state fragility

The empirical analyses based on types of fragility and the subnational measure of fragility do not reveal a significant overall effect of fragility on project success. Although we are unable to provide definitive answers as our data do not permit the estimation of a credible counterfactual, our findings are in line with earlier research showing that ratings for development interventions are no lower in fragile contexts than in situations of consolidated statehood (Carvalho, 2006).

The mechanism behind this perhaps surprising result is not immediately clear. A first possible explanation is that projects are well adapted to fragility in terms of programming and design. Our finding in Chapter 3 – that strategies on fragility are quite well implemented in German development cooperation – seems to point in this direction. Related to this, the goals of interventions in fragile contexts might be set lower, which would make them easier to achieve. However, interventions in highly fragile contexts usually have a dual objective in that they also aim at stabilization. This might lead to other, but probably not necessarily less ambitious, goals. A second possible explanation is that a negative effect of fragility would be offset by a positive effect of higher marginal returns. Future analyses should investigate this more closely, based on detailed data on the costs and outcomes of interventions.

Methodological shortcomings might also be in play. One possible explanation is that evaluations come too early to witness the adverse effects of fragility on project success. However, the fact that our findings are robust across different types of evaluation and that our analysis includes predominantly *ex post* evaluations suggests otherwise. A second methodological explanation for the finding could be that evaluators take account of the fragility of the context in their ratings, which would lead to spurious results. There is some evidence that conflict and regime type affect project ratings differently than project outcomes pointing in this direction (Metzger and Guenther, 2015). In addition, an evaluation by Noltze et al. (2018b) shows that only about 60% of evaluation reports contain plausible conclusions that follow from the analysis of the data.

Project success ratings vary by state capacity

When looking more closely at the dimensions rather than types of fragility, the evaluation reports in our synthesis indicate that project success is conditional on state capacity. We find evidence that the extent to which development interventions achieve their objectives inducing sustainable positive change depends on the capacity of partner countries. By contrast, whether interventions do so in cost- and time-efficient ways and whether they are well adapted to the target group seems to depend less on capacity.

This finding supports the results of the evaluation of sustainability by Noltze et al. (2018b). They find the role of partner countries to be particularly important in rating sustainability. Thus, the achievement of sustainable results seems to be a particular challenge of development interventions in low-capacity states.

Disaggregating the analyses between different sectors reveals that the correlation between state capacity and project ratings is particularly prominent in the health sector. This resonates with existing findings that interventions aimed at improving health require efficient and independent bureaucracies as well as an efficient infrastructure of service provision across the whole country (Briebe, 2018). However, we should be cautious with these results as health outcomes are also included in the measurement of state capacity.

More generally, the relation between fragility and project success turns out to be context-, sector-, and outcome-specific. Tailoring development cooperation to fragile contexts requires a deeper understanding of the complex mechanisms that determine project success in different sectors across different contexts. Referring back to the discussion in Chapter 2, a multidimensional understanding of fragility is certainly helpful in this regard. The findings in this chapter add insights on sector- and outcome-specific effects. Although further research and better data are needed, strengthening state capacity seems to be crucial in order to create a context in which development cooperation is successful.

We conclude from our results that it is advisable to continue, and possibly extend, the good practice of assessing state capacity in the management of bilateral development cooperation. Differentiating fragility profiles (Federal Government of Germany, 2017) and a detailed Peace and Conflict Assessment (BMZ and GIZ Sektorvorhaben Frieden und Sicherheit, 2014) is an important step in this direction. The risk assessment could also draw on (the successor to) the Governance Criteria Catalogue as well as monitoring and evaluations of similar interventions in similar contexts. Moreover, measures of state capacity could be further improved by disaggregating capacity between different actors and by taking into account subnational variations in state capacity (also see Recommendation 1).

Furthermore, a higher standardization of evaluations should improve the comparability of evaluation reports. From an evaluator's point of view, validity of indicators with regard to project goals and comparability of indicators across different projects are of primary importance. Current efforts by the BMZ, GIZ, and KfW in the joint working group "AG Evaluierung" are aimed in this direction and hence should improve the quality of observational data analysis across different contexts.

We thus make the following recommendation:

Recommendation 5

Based on weak observational evidence from an evaluation synthesis, a screening of research and evaluations, and a review of the structure and processes of the German bilateral development cooperation system, this evaluation recommends that the BMZ systematically assess the risks emanating from a lack of capacity of partner countries for the success of development projects even more than has hitherto been the case. This should include a detailed assessment in module proposals of implementation risks in low-capacity states.

Moreover, the analysis points to the added value of improving the quality of, and standardizing, monitoring and evaluation. Consequently, the joint BMZ-GIZ-KfW working group "AG Evaluierung" should strive to increase comparability across evaluations with a view to enhancing the validity of evaluation syntheses and thereby facilitate joint learning.

6. REFERENCES

- 3rd High Level Forum on Aid Effectiveness (2008)**, “Accra Agenda for Action”, 3rd High Level Forum on Aid Effectiveness, Accra.
- Acht, M. et al. (2015)**, “Corrupt Governments Do Not Receive More State-to-State Aid: Governance and the Delivery of Foreign Aid through Non-State Actors”, *Journal of Development Economics*, Vol. 114, pp. 20–33.
- Alesina, A. and D. Dollar (2000)**, “Who Gives Foreign Aid to Whom and Why?”, *Journal of Economic Growth*, Vol. 5, pp. 33–63.
- Arndt, C. et al. (2010)**, “Aid, Growth, and Development: Have We Come Full Circle?”, *Journal of Globalization and Development*, Vol. 1, No. 2.
- Arvis, J.-F. et al. (2018)**, “Connecting to Compete 2018. Trade Logistics in the Global Economy. The Logistics Performance Index and Its Indicators”, International Bank for Reconstruction and Development; World Bank, Washington D. C.
- Asseburg, M. (2011)**, “Der Arabische Frühling. Herausforderung und Chance für die deutsche und europäische Politik.”, SWP-Studie, Stiftung für Wissenschaft und Politik (SWP), Berlin.
- Baranyi, S. and M.-E. Desrosiers (2012)**, “Development Cooperation in Fragile States: Filling or Perpetuating Gaps?”, *Conflict, Security & Development*, Vol. 12, No. 5, pp. 443–459.
- Baumann, R. and G. Hellmann (2001)**, “Germany and the Use of Military Force: “Total War”, the “Culture of Restraint” and the Quest for Normality”, *German Politics*, Vol. 10, No. 1, pp. 61–82.
- Beirat Zivile Krisenprävention (2008)**, “Zivilgesellschaftliche Perspektiven zum Aktionsplan. Bericht und Stellungnahme des Beirats „Zivile Krisenprävention“”, Berlin.
- Beirat Zivile Krisenprävention (2014)**, “Zivile Krisenprävention in Globaler Verantwortung – Deutsche Friedensförderung vor neuen Aufgaben. Impulspapier des Beirats Zivile Krisenprävention zur Weiterentwicklung der Strukturen ziviler Krisenprävention, Konfliktbearbeitung und Friedenskonsolidierung in Deutschland zu Beginn der 18. Legislaturperiode”, Berlin.
- Berthelemy, J.-C. (2006)**, “Bilateral Donors’ Interest vs. Recipients’ Development Motives in Aid Allocation: Do All Donors Behave the Same?”, *Review of Development Economics*, Vol. 10, No. 2, pp. 179–194.
- Besley, T. and T. Persson (2011)**, “Fragile States and Development Policy”, *Journal of the European Economic Association*, Vol. 9, No. 3, pp. 371–398.
- Blattman, C. et al. (2014)**, “How to Promote Order and Property Rights under Weak Rule of Law? An Experiment in Changing Dispute Resolution Behavior through Community Education”, *American Political Science Review*, Vol. 108, No. 01, pp. 100–120.
- BMZ (2005)**, “Übersektorales Konzept zur Krisenprävention, Konfliktbearbeitung und Friedensförderung in der deutschen Entwicklungszusammenarbeit”, Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung, Bonn, Berlin.
- BMZ (2007)**, “Development-Oriented Transformation in Conditions of Fragile Statehood and Poor Government Performance”, No. 153, Strategies, Federal Ministry for Economic Cooperation and Development, Bonn, Berlin.
- BMZ (2009a)**, “Promoting Resilient States and Constructive State-Society Relations – Legitimacy, Transparency and Accountability”, No. 168, Federal Ministry for Economic Cooperation and Development, Bonn, Berlin.
- BMZ (2009b)**, “Promotion of Good Governance in German Development Policy”, No. 178, Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung, Bonn, Berlin.
- BMZ (2013a)**, “Entwicklung für Frieden und Sicherheit: Entwicklungspolitisches Engagement im Kontext von Konflikt, Fragilität und Gewalt”, Strategiepapier, Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ), Bonn/Berlin.

- BMZ (2013b)**, “Development for Peace and Security. Development Policy in the Context of Conflict, Fragility and Violence”, BMZ Strategy Paper 4/2013e, German Federal Ministry for Economic Cooperation and Development, Bonn / Berlin.
- BMZ (2013c)**, “Strategy on Transitional Development Assistance. Strengthening Resilience- Shaping Transition.”, 6, Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung, Bonn, Berlin.
- BMZ (2017a)**, “Entwicklungspolitik als Zukunfts- und Friedenspolitik - 15. Entwicklungspolitische Bericht der Bundesregierung”, Bundesministerium für Wirtschaftliche Zusammenarbeit und Entwicklung (BMZ).
- BMZ (2017b)**, “Africa and Europe – A New Partnership for Development, Peace and a Better Future. Cornerstones of a Marshall Plan with Africa”, Federal Ministry for Economic Cooperation and Development, Bonn/Berlin.
- BMZ (no date)**, “State Fragility – a Challenge for Development Policy”, *Federal Ministry for Economic Cooperation and Development*, http://www.bmz.de/en/issues/state_fragility/index.html (accessed 03.12.2018).
- BMZ and GIZ project Internet and Sustainable Development (2016)**, “Toolkit – Digitalisation in Development Cooperation and International Cooperation in Education, Culture and Media”, Federal Ministry for Economic Cooperation and Development, Bonn, Berlin.
- BMZ and GIZ Sektorvorhaben Frieden und Sicherheit (2014)**, “Peace and Conflict Assessment. Factsheet Zum Methodischen Rahmen „Peace and Conflict Assessment (PCA)“”, https://www.bmz.de/en/zentrales_downloadarchiv/themen_und_schwerpunkte/frieden/Peace_and_Conflict_Assessment_Factsheet.pdf.
- Boldt, K. (2012)**, “Leitlinien für den Umgang mit Fragilen Staaten”, *Entwicklungspolitik Online (EPO)*, Berlin.
- Börzel, T. A. and T. Risse (2018)**, “From the Euro to the Schengen Crises: European Integration Theories, Politicization, and Identity Politics”, *Journal of European Public Policy*, Vol. 25, No. 1, pp. 83–108.
- Boussalis, C. and C. Peiffer (2011)**, “Health, Need and Politics: The Determinants of Bilateral HIV/AIDS Assistance”, *The Journal of Development Studies*, Vol. 47, No. 12, pp. 1798–1825.
- Braithwaite, A. (2010)**, “Resisting Infection: How State Capacity Conditions Conflict Contagion”, *Journal of Peace Research*, Vol. 47, No. 3, pp. 311–319.
- Brandt, P. T. et al. (2011)**, “Real Time, Time Series Forecasting of Inter- and Intra-State Political Conflict”, *Conflict Management and Peace Science*, Vol. 28, No. 1, pp. 41–64.
- Briebe, D. (2018)**, “State Capacity and Health Outcomes: Comparing Argentina’s and Chile’s Reduction of Infant and Maternal Mortality, 1960–2013”, *World Development*, Vol. 101, pp. 37–53.
- Briggs, R. C. (2018)**, “Poor Targeting: A Gridded Spatial Analysis of the Degree to Which Aid Reaches the Poor in Africa”, *World Development*, Vol. 103, pp. 133–148.
- Brinkerhoff, D. W. et al. (2018)**, “Distance, Services, and Citizen Perceptions of the State in Rural Africa”, *Governance*, Vol. 31, No. 1, pp. 103–124.
- Brzoska, M. (2004)**, ““New Wars” Discourse in Germany”, *Journal of Peace Research*, Vol. 41, No. 1, pp. 107–117.
- Bueger, C. and F. Bethke (2014)**, “Actor-Networking the “Failed State” — an Enquiry into the Life of Concepts”, *Journal of International Relations and Development*, Vol. 17, No. 1, pp. 30–60.
- Bueno de Mesquita, B. et al. (2003)**, *The Logic of Political Survival*, MIT Press, Cambridge.
- Buhaug, H. and K. S. Gleditsch (2008)**, “Contagion or Confusion? Why Conflicts Cluster in Space”, *International Studies Quarterly*, Vol. 52, No. 2, pp. 215–233.

- Bundesregierung (2004)**, “Civilian Crisis Prevention, Conflict Resolution and Post-Conflict Peacebuilding”, Berlin.
- Bundesregierung (2012)**, “Ressortübergreifende Leitlinien - Für eine kohärente Politik der Bundesregierung gegenüber Fragilen Staaten”, Berlin.
- Bundesregierung (2014)**, “Deutsches Engagement bei Krisenprävention - Drucksache 18/3213”, Deutscher Bundestag, Berlin.
- Bundesregierung (2017)**, “Krisen verhindern, Konflikte bewältigen, Frieden fördern. Leitlinien der Bundesregierung”, Berlin.
- Burgess, S. F. (2017)**, “International Support for State Building in War-Torn Africa: Are There Alternative Strategies?”, *Comparative Strategy*, Vol. 36, No. 2, pp. 129–142.
- Burnside, A. C. and D. Dollar (2004)**, “Aid, Policies, and Growth: Revisiting the Evidence”, *World Bank Policy Research Working Paper*, No. 3251.
- Burnside, C. and D. Dollar (2000)**, “Aid, Policies, and Growth”, *The American Economic Review*, Vol. 90, No. 4, pp. 847–868.
- Buterbaugh, K. N. et al. (2017)**, “Predicting Revolt: Fragility Indexes and the Level of Violence and Instability in the Arab Spring”, *Terrorism and Political Violence*, Vol. 29, No. 3, pp. 483–508.
- Cammack, D. et al. (2006)**, “Donors and the “Fragile States” Agenda: A Survey of Current Thinking and Practice”, Overseas Development Institute, Japan International Cooperation Agency, London.
- Capoccia, G. (2016)**, “Critical Junctures”, in Fioretos, K.-O., T. G. Falleti and A. D. Sheingate (eds.), *The Oxford Handbook of Historical Institutionalism*, Oxford University Press, Oxford New York NY, First edition., pp. 118–140.
- Capoccia, G. and D. R. Kelemen (2007)**, “The Study of Critical Junctures: Theory, Narrative, and Counterfactuals in Historical Institutionalism”, *World Politics*, Vol. 59, No. 3, pp. 341–369.
- Carlsson, J. and L. Wohlgemuth (1996)**, “Capacity Building and Networking - A Meta Evaluation”, Sida Evaluation, Department for Evaluation and Internal Audit.
- Carvalho, S. (2006)**, *Engaging with Fragile States: An IEG Review of World Bank Support to Low-Income Countries under Stress*, Weltbank, Independent Evaluation Group study series, World Bank, Washington, DC.
- Caspari, A. (2014)**, “Sektorbezogene Querschnittsauswertung: Meta-Evaluierung Ländliche Entwicklung”, Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH, Eschborn.
- Cederman, L.-E. et al. (2013)**, *Inequality, Grievances, and Civil War*, Cambridge University Press, Cambridge.
- Chadefaux, T. (2014)**, “Early Warning Signals for War in the News”, *Journal of Peace Research*, Vol. 51, No. 1, pp. 5–18.
- Chandler, D. (2013)**, “International Statebuilding and the Ideology of Resilience”, *Politics*, Vol. 33, No. 4, pp. 276–286.
- Chapman, N. et al. (2009)**, “Evaluation of Donor-Supported Activities in Conflict-Sensitive Development and Conflict Prevention and Peacebuilding in Sri Lanka”.
- Chauvet, L. and P. Guillaumont (2003)**, “Aid and Growth Revisited: Policy, Economic Vulnerability and Political Instability”, *ABCDE*, pp. 95–109.
- Chiba, D. et al. (2017)**, “The Shape of Things to Come? Expanding the Inequality and Grievance Model for Civil War Forecasts with Event Data”, Nygård, H. M. and Wucherpfennig, J., Eds. *Journal of Peace Research*, Vol. 54, No. 2, pp. 275–297.

- Church, C. and J. Shouldice (2002), “The Evaluation of Conflict Resolution Interventions: Framing the State of Play”, International Conflict Research, Derry, Londonderry.
- Clemens, M. A. et al. (2012), “Counting Chickens When They Hatch: Timing and the Effects of Aid on Growth”, *The Economic Journal*, Vol. 122, No. 561, pp. 590–617.
- Collier, D. and G. L. Munck (2017), “Building Blocks and Methodological Challenges: A Framework for Studying Critical Junctures”, *Qualitative & Multi-Method Research*, presented at the Symposium on Critical Junctures and Historical Legacies, American Political Science Association, Durham.
- Collier, P. (2008), *The Bottom Billion: Why the Poorest Countries Are Failing and What Can Be Done about It*, Univ. Press, Oxford, 1. OUP paperback.
- Collier, P. and D. Dollar (2002), “Aid Allocation and Poverty Reduction”, *European Economic Review*, Vol. 46, No. 8, pp. 1475–1500.
- Commins, T. et al. (2019), “Pooled Funding to Support Service Delivery. Lessons of Experience from Fragile and Conflict-Affected States”, DFID.
- Coppedge, M. et al. (2016), “V-Dem Methodology V6”, SSRN Scholarly Paper, No. ID 2951040, Social Science Research Network, Rochester, NY.
- Cornell, A. (2013), “Does Regime Type Matter for the Impact of Democracy Aid on Democracy?”, *Democratization*, Vol. 20, No. 4, pp. 642–667.
- Creswell, J. W. (2014), *A Concise Introduction to Mixed Methods Research*, Sage.
- Dames, M. (2015), “Armee Im Einsatz”, *Bundeswehr*, https://www.bundeswehr.de/portal/a/bwde/start/streitkraefte/grundlagen/geschichte/umbruch/!ut/p/z1/o4_Sj9CPykssyoxPLMnMzovMAfljo8zinSx8QnyMLIzMTMosDAw8DUMCPDy9TloMLIz1WwkpIAJ/KG-AAjgb6wSmp-pFAM8xxmuFvph-sH6UflZVYllihV5BfVJKTWqKXmAxyoX5kRmJeSk5qQH6yloSglDeizKDcUREAvKApoA!!/dz/d5/L2dBISEvZoFBIS9nQSEh/#Z7_B8LTL292256800l1TPHlIj52oO6 (accessed 08.01.2017).
- D’Aquino, L. et al. (2019), “Introducing a Sector-Wide Pooled Fund in a Fragile Context: Mixed-Methods Evaluation of the Health Transition Fund in Zimbabwe”, *BMJ Open*, Vol. 9, No. 6, p. e024516.
- Davenport, C. and W. H. Moore (2015), “Conflict Consortium Standards & Best Practices for Observational Data”, <https://conflictconsortium.weebly.com/standards--best-practices.html>.
- Davison, J. (2004), “Dilemmas in Research: Issues of Vulnerability and Disempowerment for the Social Worker/ Researcher”, *Journal of Social Work Practice*, Vol. 18, No. 3, pp. 379–393.
- De Groeve, T. et al. (2015), *Index for Risk Management - INFORM: Concept and Methodology, Version 2016*, Publications Office, Luxembourg.
- Development Researchers Network (DRN) (2011), “Thematic Global Evaluation of European Commission Support in the Sectors of ESI (Employment and Social Inclusion) in Partner Countries (Including Vocational Training)”, Final Report, European Commission, Brüssel.
- DFID (2016), “Bilateral Development Review: Technical Note”, Department for International Development, London, Glasgow.
- Dietrich, S. (2013), “Bypass or Engage? Explaining Donor Delivery Tactics in Foreign Aid Allocation*: *Bypass or Engage?*”, *International Studies Quarterly*, Vol. 57, No. 4, pp. 698–712.
- Dietrich, S. (2016), “Donor Political Economies and the Pursuit of Aid Effectiveness”, *International Organization*, Vol. 70, No. 01, pp. 65–102.
- DiLorenzo, M. (2018), “Bypass Aid and Unrest in Autocracies”, *International Studies Quarterly*, Vol. 62, No. 1, pp. 208–219.

- Doucouliaagos, H. and M. Paldam (2009), “The Aid Effectiveness Literature: The Sad Results of 40 Years of Research”, *Journal of Economic Surveys*, Vol. 23, No. 3, pp. 433–461.
- Dreher, A. et al. (2015), “The Allocation of German Aid: Self-Interest and Government Ideology”, *Economics & Politics*, Vol. 27, No. 1, pp. 160–184.
- Dreher, A. et al. (2008), “Does US Aid Buy UN General Assembly Votes? A Disaggregated Analysis”, *Public Choice*, Vol. 136, No. 1–2, pp. 139–164.
- Easterly, W. et al. (2004), “Aid, Policies, and Growth: Comment”, *The American Economic Review*, Vol. 94, No. 3, pp. 774–780.
- Eriksen, S. S. (2011), ““State Failure” in Theory and Practice: The Idea of the State and the Contradictions of State Formation”, *Review of International Studies*, Vol. 37, No. 1, pp. 229–247.
- Fassin, D. (2012), “Hierarchies of Humanity. Intervening in International Conflicts”, *Humanitarian Reason. A Moral History of the Present*, pp. 223–242.
- Faust (2008), “Are More Democratic Donor Countries More Development Oriented? Domestic Institutions and External Development Promotion in OECD Countries”, *World Development*, Vol. 36, No. 3, pp. 383–398.
- Faust, J. (2011), *Donor Transparency and Aid Allocation*, Discussion paper, Deutsches Institut für Entwicklungspolitik, Bonn.
- Faust, J. et al. (2013), *Foreign Aid and the Fragile Consensus on State Fragility*, Discussion paper, Deutsches Institut für Entwicklungspolitik, Bonn.
- Faust, J. and S. Leiderer (2008), “Zur Effektivität und politischen Ökonomie der Entwicklungszusammenarbeit”, *Politische Vierteljahresschrift*, Vol. 49, No. 1, pp. 129–152.
- Faust, J. et al. (2016), “German Aid from a Partner Perspective: Experience-Based Perceptions from AidData’s 2014 Reform Efforts Survey”, German Institute for Development Evaluation (DEval), Bonn.
- Federal Government of Germany (2017), “Guidelines on Preventing Crises, Resolving Conflicts, Building Peace”, Guidelines, Federal Government of Germany, Berlin.
- Ferreira, I. A. (2016), “Measuring State Fragility: A Review of the Theoretical Groundings of Existing Approaches”, *Third World Quarterly*, Vol. 38, No. 6, pp. 1291–1309.
- Findley, M. G. et al. (2017), “The Choice among Aid Donors: The Effects of Multilateral vs. Bilateral Aid on Recipient Behavioral Support”, *The Review of International Organizations*, Vol. 12, No. 2, pp. 307–334.
- Fjelde, H. (2009), “Buying Peace? Oil Wealth, Corruption and Civil War, 1985–99”, *Journal of Peace Research*, Vol. 46, No. 2, pp. 199–218.
- Fleckenstein, B. (2005), “50 Jahre Bundeswehr”, *Aus Politik Und Zeitgeschichte*, No. 21, pp. 5–14.
- Flick, U. et al. (Eds.) (1995), *Handbuch qualitative Sozialforschung: Grundlagen, Konzepte, Methoden und Anwendungen*, Beltz/Psychologie-Verlags-Union, Weinheim.
- François, M. and I. Sud (2006), “Promoting Stability and Development in Fragile and Failed States”, *Development Policy Review*, Vol. 24, No. 2, pp. 141–160.
- Fukuyama, F. (1989), “The End of History?”, *The National Interest*, No. 16, pp. 3–18.
- Furness, M. (2014), *Let’s Get Comprehensive: European Union Engagement in Fragile and Conflict-Affected Countries*, Discussion paper, Deutsches Institut für Entwicklungspolitik, Bonn.
- Gause, F. G. (2011), “Why Middle East Studies Missed the Arab Spring: The Myth of Authoritarian Stability”, *Foreign Affairs*, Vol. 90, No. 4, pp. 81–90.
- Gehring, K. et al. (2017), “Aid Fragmentation and Effectiveness: What Do We Really Know?”, *World Development*, Vol. 99, pp. 320–334.

- Gibler, D. M. (2017), "Combining Behavioral and Structural Predictors of Violent Civil Conflict: Getting Scholars and Policymakers to Talk to Each Other", *International Studies Quarterly*, Vol. 61, No. 1, pp. 28–37.
- GIZ (2015), "Capacity Development in Situations of Conflict and Fragility. German Approaches and Lessons Learned by GIZ", Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH, Bonn, Eschborn.
- Glass, M.-L. and C. Schönhofen (2018), "Digitaler Fortschritt in der Evaluierung: Eine neue Perspektive oder ein neues Paradigma?", in Sangmeister, H. and H. Wagner (eds.), *Entwicklungszusammenarbeit 4.0 - Digitalisierung Und Globale Verantwortung*, Nomos, pp. 67–82.
- Gleditsch, K. S. (2002), *All International Politics Is Local. The Diffusion of Conflict, Integration, and Democratization*, University of Michigan Press, Ann Arbor.
- Goertz, G. (2006), *Social Science Concepts. A User's Guide*, Princeton University Press, Princeton.
- Gould, L. A. and M. Page (2016), *State Fragility Around the World: Fractured Justice and Fierce Reprisal*, CRC Press.
- Grävingholt, J. (2016), "Entwicklungspolitik Im Gefüge Einer "Neuen Deutschen Außenpolitik"", *Aus Politik Und Zeitgeschichte (APuZ)*, Vol. 66, No. 28–29, pp. 38–43.
- Grävingholt, J. et al. (2015), "Disaggregating State Fragility: A Method to Establish a Multidimensional Empirical Typology", *Third World Quarterly*, Vol. 36, No. 7, pp. 1281–1298.
- Grävingholt, J. et al. (2018), "Constellations of State Fragility v1.0", German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE).
- Grechyna, D. (2018), "Shall We Riot Too? The Geographical Neighbor Impact on Political Instability", *Kyklos*, Vol. 71, No. 4, pp. 581–612.
- Grossman, H. I. (1992), "Foreign Aid and Insurrection", *Defence Economics*, Vol. 3, No. 4, pp. 275–288.
- Hageboeck, M. et al. (2013), "Meta-Evaluation of Quality and Coverage of USAID Evaluations 2009–2012", United States Agency for International Development (USAID), Washington, DC.
- HAP - Humanitarian Accountability Partnership (2013), "2013 Humanitarian Accountability Report".
- Hartmann, H. et al. (2018), *Transformation Index BTI 2018: Political Management in International Comparison*, Bertelsmann Stiftung, Ed., Brookings Institution Press, Washington.
- Hay, C. and M. Lister (2006), "Introduction: Theories of the State", *The State. Theories and Issues*, Palgrave MacMillan, pp. 1–20.
- Hayes, A. F. (2009), "Beyond Baron and Kenny: Statistical Mediation Analysis in the New Millennium", *Communication Monographs*, Vol. 76, No. 4, pp. 408–420.
- Hearn, S. (2016), *Independent Review of the New Deal for Engagement in Fragile States*, Center on International Cooperation.
- Heathershaw, J. (2012), "The Practical Representation of Peacebuilding", in Richmond, O. P. and A. Mitchell (eds.), *Hybrid Forms of Peace: From Everyday Agency to Post-Liberalism*, Rethinking Peace and Conflict Studies, Palgrave Macmillan UK, London, pp. 162–187.
- Hegre, H. and H. M. Nygård (2015), "Governance and Conflict Relapse", *Journal of Conflict Resolution*, Vol. 59, No. 6, pp. 984–1016.
- Heinze, Georg; Schemper, Michael (2002), "A Solution to the Problem of Separation in Logistic Regression - Heinze - 2002 - Statistics in Medicine - Wiley Online Library", *Statistics in Medicine*, Vol. 21, No. 16, pp. 2409–2419.
- Henning, K. (2017), "Was ist „state building“ und wie kann state building in fragilen Ländern gefördert werden?", No. 16, KfW Development Research - Entwicklungspolitik Kompakt, KfW.

- High Level Forum (2005)**, “Paris Declaration on Aid Effectiveness”, OECD/DAC, Paris.
- Hingorani, Y. (2015)**, “The New Deal for Engagement in Fragile States: Where Are We Now?”, *Journal of Peacebuilding & Development*, Vol. 10, No. 2, pp. 87–93.
- Hofmann, B. (2007)**, “Identitätskonflikte als Hindernis entwicklungspolitisch militärischer Zusammenarbeit”, in Weller, C. (ed.), *Zivile Konfliktbearbeitung - Aktuelle Forschungsergebnisse*, Institut für Entwicklung und Frieden, Universität Duisburg Essen, Duisburg, pp. 61–68.
- Holmberg, S. and B. Rothstein (2011)**, “Dying of Corruption”, *Health Economics, Policy and Law*, Vol. 6, No. 04, pp. 529–547.
- Honig, D. (2018)**, *Navigation by Judgment: Why and When Top-down Management of Foreign Aid Doesn't Work*, Oxford University Press, New York, NY.
- Huber, S. et al. (2014)**, “Querschnittsauswertung Bildung: Meta-Evaluierung und Synthese”, Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH, Bonn.
- Imai, K. et al. (2011)**, “Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies”, *American Political Science Review*, Vol. 105, No. 4, pp. 765–789.
- International Dialogue on Peacebuilding and Statebuilding (2011)**, “A New Deal for Engagement in Fragile States”, Fact Sheet, International Dialogue on Peacebuilding and Statebuilding.
- Kahneman, D. (2012)**, *Thinking, Fast and Slow*, Penguin Books, London.
- Kaldor, M. (1999)**, *New and Old Wars: Organized Violence in a Global Era*, Polity Press, Cambridge.
- Kaldor, M. (2013)**, “In Defence of New Wars”, *Stability: International Journal of Security and Development*, Vol. 2, No. 1, p. Art. 4.
- Killick, T. (1998)**, “Principals, Agents and the Failings of Conditionality”, *Journal of International Development*, Vol. 9, No. 4, pp. 483–495.
- King, G. et al. (2004)**, “Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research”, *American Political Science Review*, Vol. 98, No. 01, pp. 191–207.
- Kirschner, A. (2007)**, “Entwicklung, Frieden, Sicherheit – Zivile Konfliktbearbeitung im Kontext deutscher Entwicklungspolitik”, *Zivile Konfliktbearbeitung: Aktuelle Forschungsergebnisse*, Institut für Entwicklung und Frieden, Universität Duisburg Essen, Duisburg, pp. 19–26.
- Klotzle, K. (2006)**, “International Strategies in Fragile States: Expanding the Toolbox?”, No. 1, C-A-P Policy Analysis, Center for Applied Policy, Bertelsmann Stiftung.
- Knack, S. and A. Rahman (2007)**, “Donor Fragmentation and Bureaucratic Quality in Aid Recipients”, *Journal of Development Economics*, Vol. 83, No. 1, pp. 176–197.
- Koehler, D. J. (Ed.) (2007)**, *Blackwell Handbook of Judgment and Decision Making*, Blackwell handbooks of experimental psychology, Blackwell Publ, Malden, MA, 1. ed.
- Krasner, S. D. (2007)**, “Sovereignty”, in Ritzer, G. (ed.), *The Blackwell Encyclopedia of Sociology*, American Cancer Society.
- Krueger, S. and E. Sagmeister (2014)**, “Real Time Evaluation of Humanitarian Assistance Revisited: Lessons Learned and the Way Forward”, *Journal of MultiDisciplinary Evaluation*, Vol. 10, No. 23, pp. 59–72.
- Kuckartz, U. (2014)**, *Qualitative Inhaltsanalyse – Methoden, Praxis, Computerunterstützung*, Grundagentexte Methoden, Beltz Juventa, Weinheim/Basel.
- Lambach, D. et al. (2015)**, “Conceptualising State Collapse: An Institutional Approach”, *Third World Quarterly*, Vol. 36, No. 7, pp. 1299–1315.

- Lech, M. et al. (2019), *Integrating Machine Learning with Remote Sensing to Evaluate Disaster Resilience. The Case of the Philippines as a Proof-of-Concept*, DEval Discussion Paper, Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval), Bonn.
- Lech, M. et al. (2017), "From GIS to M&E: Extracting Value from Geodata for Evaluations", Policy Brief, No. 1/2017, Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit, Bonn.
- Lee, M. M. and N. Zhang (2016), "Legibility and the Informational Foundations of State Capacity", *The Journal of Politics*, Vol. 79, No. 1, pp. 118–132.
- Leeuw, F. L. and L. J. Cooksy (2005), "Evaluating the Performance of Development Agencies: The Role of Metaevaluations.", in Pitman, G. K., O. N. Feinstein and G. K. Ingram (eds.), *Evaluating Development Effectiveness*, World Bank series on Evaluation and Development, Transaction, New Brunswick, pp. 95–108.
- Lieberman, E. S. (2018), "The Comparative Politics of Service Delivery in Developing Countries", in Lancaster, C. and N. Van de Walle (eds.), *The Oxford Handbook of the Politics of Development*, Oxford University Press, New York, NY, pp. 480–498.
- Lindemann, S. (2014), "Was ist ein fragiler Staat?", KfW-Development Research, No. 4, Fokus Entwicklungspolitik, KfW Entwicklungsbank.
- Lührmann, A. et al. (2017), "The Effectiveness of Democracy Aid to Different Regime Types and Democracy Sectors", SSRN Scholarly Paper, No. ID 2896872, Social Science Research Network, Rochester, NY.
- Mac Ginty, R. and O. P. Richmond (2013), "The Local Turn in Peace Building: A Critical Agenda for Peace", *Third World Quarterly*, Vol. 34, No. 5, pp. 763–783.
- Manning, R. and A. Trzeciak-Duval (2010), "Situations of Fragility and Conflict: Aid Policies and Beyond", *Conflict, Security & Development*, Vol. 10, No. 1, pp. 103–131.
- Manuel, M. et al. (2012), "Innovative Aid Instruments and Flexible Financing: Providing Better Support to Fragile States", Overseas Development Institute, London.
- Martens, B. et al. (Eds.) (2002), *The Institutional Economics of Foreign Aid*, Cambridge.
- Mata, J. F. and S. Ziaja (2009), "Users' Guide on Measuring Fragility", Deutsches Institut für Entwicklungspolitik, United Nations Development Programme.
- Mazucca, S. (2017), "States, Regimes and Other Structures of Rule", *APSA Comparative Democratization Newsletter*.
- McCloughlin, C. and I. Idris (2016), "Fragile States: Topic Guide", GSDRC, University of Birmingham, Birmingham, UK.
- Metzger, L. and I. Guenther (2015), "How to Assess the Effectiveness of Development Aid Projects: Evaluation Ratings versus Project Indicators", *Journal of International Development*, Vol. 27, No. 8, pp. 1496–1520.
- Migdal, J. S. (1988), *Strong Societies and Weak States: State-Society Relations and State Capabilities in the Third World*, Princeton University Press.
- Milner, H. V. (2006), "Why Multilateralism? Foreign Aid and Domestic Principal-Agent Problems", in Hawking, D. G., D. A. Lake, D. L. Nielson and M. J. Tierney (eds.), *Delegation and Agency in International Organizations*, Cambridge University Press New York, Vol. 107, pp. 107–139.
- Minasyan, A. et al. (2017), "Does Aid Effectiveness Depend on the Quality of Donors?", *World Development*, Vol. 100, pp. 16–30.
- Minoiu, C. and S. G. Reddy (2010), "Development Aid and Economic Growth: A Positive Long-Run Relation", *The Quarterly Review of Economics and Finance*, Vol. 50, No. 1, pp. 27–39.

- Mitchell, I. and C. McKee (2018), “How Do You Measure Aid Quality and Who Ranks Highest?”, *Center For Global Development*, <https://www.cgdev.org/blog/how-do-you-measure-aid-quality-and-who-ranks-highest> (accessed 11.12.2018).
- Münkler, H. (2003), *Die neuen Kriege*, Rowohlt, Hamburg.
- Nachtwei, W. (2010), “Wachstumsschwäche bei Friedensfähigkeiten: Der 3. Umsetzungsbericht der Bundesregierung zum “Aktionsplan Zivile Krisenprävention” braucht Streit”, http://nachtwei.de/downloads/position/201006_3.umsetzungsbericht-ZKB_kommentar-nachtwei.pdf (accessed 11.12.2017).
- Noltze, M. et al. (2018a), “Evaluation Synthesis of Sustainability in German Development Cooperation”, German Institute for Development Evaluation (DEval), Bonn, p. 72.
- Noltze, M. et al. (2018b), “Meta-Evaluierung von Nachhaltigkeit in der Deutschen Entwicklungszusammenarbeit”, Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval), Bonn.
- Noltze, M. et al. (2018c), “Meta-Evaluation of Sustainability in German Development Cooperation”, Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit, Bonn.
- North, D. C. et al. (Eds.) (2012), *In the Shadow of Violence: Politics, Economics, and the Problems of Development*, Cambridge University Press, Cambridge, doi:10.1017/CBO9781139013611.
- Nozick, R. (1974), *Anarchy, State, and Utopia*, Basic Books, New York.
- Nussbaum, T. et al. (2012), “A New Deal for Engagement in Fragile States”, *Conflict, Security & Development*, Vol. 12, No. 5, pp. 559–587.
- OECD (n.d.), “Official Development Assistance - Definition and Coverage”, <http://www.oecd.org/dac/stats/officialdevelopmentassistancedefinitionandcoverage.htm> (accessed 06.02.2019).
- OECD (2008), “Evaluating Conflict Prevention and Peacebuilding”.
- OECD (2011a), *International Engagement in Fragile States: Can't We Do Better?*, Conflict and Fragility, OECD.
- OECD (2011b), *Supporting Statebuilding in Situations of Conflict and Fragility: Policy Guidance*, DAC Guidelines and Reference Series, OECD.
- OECD (2012), “Evaluating Peacebuilding Activities in Settings of Conflict and Fragility - Improving Learning for Results”, No. 40, DAC Guidelines and Reference Series.
- OECD (2016), *States of Fragility 2016*, OECD Publishing, Paris.
- OECD (2018), *States of Fragility 2018*, Organisation for Economic Co-operation and Development, Paris.
- OECD DAC (2011), “Progress towards More Effective Aid: What Does the Evidence Show?”, OECD DAC, Busan.
- OECD and OECD DAC (Eds.) (2007), “Principles for Good International Engagement in Fragile States and Situations”.
- Orth, M. et al. (2018), “The Future of Integrated Policy-Based Development Cooperation. Lessons from the Exit from General Budget Support in Malawi, Rwanda, Uganda and Zambia”, No. 4/2018, German Institute for Development Evaluation (DEval), Bonn.
- Orth, M. et al. (2017), “What We Know about the Effectiveness of Budget Support. Evaluation Synthesis”, German Institute for Development Evaluation (DEval), Bonn.
- Østby, G. (2008), “Inequalities, the Political Environment and Civil Conflict: Evidence from 55 Developing Countries”, in Stewart, F. (ed.), *Horizontal Inequalities and Conflict: Understanding Group Violence in Multiethnic Societies*, Palgrave MacMillan, pp. 136–161.

- Paris, R. (2010), "Saving Liberal Peacebuilding", *Review of International Studies*, Vol. 36, No. 2, p. 337.
- Paris, R. (2011), "Ordering the World: Academic Research and Policymaking on Fragile States", *International Studies Review*, Vol. 13, No. 1, pp. 58–71.
- Patton, M. Q. (2008), *Utilization-Focused Evaluation*, SAGE Publications, Thousand Oaks.
- Pawelke, A. et al. (2017), "Data for Development: What's next? Concepts, Trends and Recommendations for German Development Cooperation", Deutsche Gesellschaft für Internationale Zusammenarbeit, Institute of Development Studies Institute of Development Studies and World Wide Web Foundation, Eds.
- Pospisil, J. and F. P. Kühn (2016), "The Resilient State: New Regulatory Modes in International Approaches to State Building?", *Third World Quarterly*, Vol. 37, No. 1, pp. 1–16.
- Putzel, J. et al. (2012), *Meeting the Challenges of Crisis States*, Crisis States Research Centre, London School of Economics and Political Science, London, England.
- Raetzell, L. and M. Krämer (2013), "Meta-Evaluation Gesundheit", Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ), Bonn.
- Rajan, R. G. and A. Subramanian (2008), "Aid and Growth: What Does the Cross-Country Evidence Really Show?", *The Review of Economics and Statistics*, Vol. 90, No. 4, pp. 643–665.
- Rajkumar, A. S. and V. Swaroop (2008), "Public Spending and Outcomes: Does Governance Matter?", *Journal of Development Economics*, Vol. 86, No. 1, pp. 96–111.
- Raleigh, C. et al. (2010), "Introducing ACLED: An Armed Conflict Location and Event Dataset", Vol. 47, No. 5, pp. 651–660.
- Rawls, J. (2005), *A Theory of Justice*, Harvard University Press, Cambridge, London.
- Roodman, D. (2012), "An Index of Donor Performance", *Center for Global Development Working Paper*, No. 67.
- Rotberg, R. I. (2002), "Failed States in a World of Terror", *Foreign Affairs*, Vol. 81, No. 4, p. 127.
- Rotberg, R. I. (2004), *When States Fail: Causes and Consequences*, Princeton University Press, Princeton.
- Rotmann, P. et al. (2014), "Major Powers and the Contested Evolution of a Responsibility to Protect", *Conflict, Security & Development*, Vol. 14, No. 4, pp. 355–377.
- Ruppert, L. et al. (2016), "Listening to Communities in Insecure Environments", Resource Paper, Secure Access in Volatile Environments (SAVE), UKaid, Global Public Policy Institute.
- Sartori, G. (1970), "Concept Misformation in Comparative Politics", *The American Political Science Review*, Vol. 64, No. 4, pp. 1033–1053.
- Schliesky, U. (2004), *Souveränität und Legitimität von Herrschaftsgewalt: die Weiterentwicklung von Begriffen der Staatslehre und des Staatsrechts im europäischen Mehrebenensystem*, Mohr Siebeck.
- Schneckener, U. (2007), "Internationales Statebuilding. Dilemmata, Strategien Und Anforderungen an Die Deutsche Politik.", SWP, Berlin.
- Schröder, G. (2002), *Speech by Chancellor Gerhard Schröder at the World Economic Forum 2002 in New York*.
- Schrodt, P. A. and D. J. Gerner (2000), "Cluster-Based Early Warning Indicators for Political Change in the Contemporary Levant", *American Political Science Review*, Vol. 94, No. 4, pp. 803–817.
- Scriven, M. (2009), "Meta-Evaluation Revisited", *Journal of MultiDisciplinary Evaluation*, Vol. 6, No. 11, pp. iii–viii.
- Steets, J. et al. (2016), "Eyes and Ears on the Ground - Monitoring Aid in Insecure Environments", Final Report, Secure Access in Volatile Environments (SAVE), UKaid, Global Public Policy Institute.

- Stengel, F. A. and C. Weller (2008), "Vier Jahre Aktionsplan "Zivile Krisenprävention" - War das alles?", *GIGA Focus*, 11, Hamburg.
- Stewart, F. (Ed.) (2008), *Horizontal Inequalities and Conflict: Understanding Group Violence in Multiethnic Societies*, Palgrave MacMillan, Basingstoke.
- Stoddard, A. and S. Jillani (2016), "The Effects of Insecurity on Humanitarian Coverage", Final Report, Secure Access in Volatile Environments (SAVE), UKaid, Global Public Policy Institute.
- Stollenwerk, E. (2017), "Measuring Statehood on the Sub-National Level in Nigeria", in Heucher, A., L. Linke-Behrens and L. Schettler (eds.), *Measuring Statehood on a Sub-National Level. A Dialogue among Methods*, SFB-Governance Working Paper Series, Collaborative Research Center (SFB) 700, Berlin, Vol. 71.
- Stufflebeam, D. L. (2001), "The Metaevaluation Imperative", *American Journal of Evaluation*, Vol. 22, No. 2, pp. 183–209.
- Sundberg, R. and E. Melander (2013), "Introducing the UCDP Georeferenced Event Dataset", *Journal of Peace Research*, Vol. 50, No. 4, pp. 523–532.
- Svensson, J. (1999), "Aid, Growth and Democracy", *Economics & Politics*, Vol. 11, No. 3, pp. 275–297.
- Taydas, Z. et al. (2010), "Why Do Civil Wars Occur? Understanding the Importance of Institutional Quality", *Civil Wars*, Vol. 12, No. 3, pp. 195–217.
- Trinn, C. and T. Wencker (2018), "Introducing the Heidelberg Approach to Conflict Research", *European Political Science*, Vol. 17, No. 1, pp. 111–121.
- Tversky, A. and D. Kahneman (Eds.) (1983), *Judgment under Uncertainty Heuristics and Biases*, Cambridge University Press, Cambridge.
- UN (2015), *Transforming Our World. The 2030 Agenda for Sustainable Development*, New York.
- UN Foundation (2018), "Unpacking Fragility: Insights from the OECD's New States of Fragility Report", *Unfoundation.Org*, <https://unfoundation.org/blog/post/unpacking-fragility-insights-from-the-oecd-s-new-states-of-fragility-report/> (accessed 04.12.2018).
- Vreeland, J. R. (2008), "The Effect of Political Regime on Civil War: Unpacking Anocracy", *Journal of Conflict Resolution*, Vol. 52, No. 3, pp. 401–425.
- Wagner, L. J. and J. Sattelberger (2017), "Fragility: Dimensions and Classifications", No. 8, KfW Development Research - Entwicklungspolitik Kompakt, KfW.
- Wagner, W. and R. Anholt (2016), "Resilience as the EU Global Strategy's New Leitmotif: Pragmatic, Problematic or Promising?", *Contemporary Security Policy*, Vol. 37, No. 3, pp. 414–430.
- Weber, M. (1978), *Economy and Society: An Outline of Interpretive Sociology*, Roth, G. and Wittich, Claus, Eds., University of California Press, Berkeley.
- Weber, W. (1946), "Politics as a Vocation", in Gerth, H. H. and C. W. Mills (trans.), *From Max Weber: Essays in Sociology*, Oxford University Press, New York, pp. 77–128.
- Widmer, T. (2006), *Meta-Evaluation. Kriterien Zur Bewertung von Evaluationen*, Verlag Paul Haupt, Zürich.
- Winckler Andersen, O. et al. (Eds.) (2014), *Evaluation Methodologies for Aid in Conflict*, Routledge, Abingdon, UK.
- Wittkowsky, A. (2017), "Stabilisierung: Querschnittsaufgabe Zur Überwindung Akuter Gewaltkonflikte", Policy Briefing, Zentrum für Internationale Friedenseinsätze, Berlin.
- Wittkowsky, A. and J. P. Meierjohann (2011), "Das Konzept der Vernetzten Sicherheit: Dimensionen, Herausforderungen, Grenzen", Policy Brief, Policy Briefing, Zentrum für Internationale Friedenseinsätze (ZIF), Berlin.

Ziviler Friedensdienst (2017), “Zahlen Und Fakten”, <https://www.ziviler-friedensdienst.org/de/zahlen-fakten> (accessed 19.01.2018).

Zoellick, R. B. (2008), “Fragile States: Securing Development”, *Survival*, Vol. 50, No. 6, pp. 67–84.