

Ariadne-Background

Governance capacities in reflexive climate policymaking: the scope, role, and institutional arrangements of policy mix evaluation in the German domestic buildings sector

SPONSORED BY THE



Federal Ministry
of Education
and Research

Authors



» Dr. Duncan L. Edmondson
Hertie School



» Oskar Krafft
Hertie School



» Prof. Dr. Christian Flachsland
Hertie School



» Christian van Ballegooy
Hertie School

Published by

Kopernikus-Projekt Ariadne
Potsdam-Institut für Klimafolgen-
forschung (PIK)
Telegrafenberg A 31
14473 Potsdam

April 2024

Cite this paper:

Duncan L. Edmondson, Oskar Krafft, Christian Flachsland, Christian van Ballegooy
(2023): Governance capacities in reflexive climate policymaking: the scope, role, and in-
stitutional arrangements of policy mix evaluation in the German domestic buildings
sector. Kopernikus project Ariadne, Potsdam.
<https://doi.org/10.48485/pik.2023.011>

Corresponding Author: Duncan Edmonson, duncan.edmondson@viablecities.se

This Ariadne report was prepared by the above-mentioned authors of the Ariadne
consortium. It does not necessarily reflect the opinion of the entire Ariadne consortium
or the funding agency. The content of the Ariadne publications is produced in the pro-
ject independently of the Federal Ministry of Education and Research.

Contents

Summary	1
1. Introduction	4
2. Ex-post evaluations processes in reflexive climate governance	8
2.1. Governance framework and institutional configuration of evaluation process	9
2.2. The scope and quality of evaluations.....	12
2.3. The use of evaluations in the policy process	17
3. Research design and procedure	20
4. Institutional configuration of governance of residential buildings in Germany	28
4.1. Ordinance and reporting requirements	28
4.1.1. <i>EU directives and reporting</i>	28
4.1.2. <i>Cross-sectoral ordinance</i>	30
4.1.3. <i>Sectoral programmes</i>	36
4.2. Governance and ministries	38
4.2.1. <i>Federal Funding for Efficient Buildings (BEG)</i>	39
4.2.2. <i>Building Energy Act (GEG)</i>	42
4.2.3. <i>Auxiliary</i>	44
4.3. Consultants	47
5. Ex-post evaluation: procedures, scope, data, and methods	51
5.1. Evaluation approaches and programmes evaluated.....	51
5.1.1. <i>Taxes/economic instruments</i>	52
5.1.2. <i>Federal Funding for Efficient Buildings (BEG)</i>	53
5.1.3. <i>Building Energy Act (GEG)</i>	54
5.1.4. <i>Methodological Guidelines</i>	55
5.2. Scope of evaluative indicators.....	58
5.2.1. <i>Effectiveness</i>	59
5.2.2. <i>Costs/Cost effectiveness</i>	61
5.2.3. <i>Distributional impacts</i>	67
5.2.4. <i>Acceptance</i>	69
5.2.5. <i>Governance</i>	70
5.3. Data.....	71
5.3.1. <i>Data quality</i>	71
5.3.2. <i>Data reporting quality and access in Federal Agencies</i>	72
5.3.3. <i>Data protection and the enforcement of regulatory instruments</i>	74
5.3.4. <i>Establishing a Building database</i>	76
5.4. Methods.....	80

5.4.1. Final and Primary Energy Savings	81
5.4.2. Effects adjustment	84
5.4.3. Interaction effects.....	86
5.4.4. General recommendations.....	88
6. Use of evaluations and effects on the policy process.....	90
6.1. Analytical inputs into ex-ante forecasting	90
6.2. Motivation, commissioning, and use of evaluations for policy recalibration	92
6.3. Timing of evaluations and informal advice	95
6.4. Dissemination of evaluations and transparency	97
7. Key reform options.....	99
7.1. Institutional arrangements and coordination	99
7.2. Scope of programmes evaluated and evaluation criteria	101
7.3. Data access and quality.....	103
7.4. Methods	104
7.5. Use of evaluations and dissemination.....	106
7.6. Compatibility of key reforms with current institutional configuration	107
8. Conclusions.....	109
References.....	113
Annex.....	119

Summary

Residential buildings directly contribute 11% to local greenhouse gas emissions and up to 40% of total emissions when accounting for energy use for electricity generation. In order to achieve the climate targets in line with the Federal Climate Protection Act, increased ambition level of climate policy instruments is required in this sector. In this research, we are interested in the governance of this sector and the role of evaluation: the government-mandated processes used to evaluate policy in terms of the actors, organisations and ministries involved in executing and coordinating these processes; and the metrics and methods as well as the scope and granularity of evaluations.

The report follows a mixed methods research design, utilising multiple sources of primary data for triangulation. We combine 14 expert interviews with content analysis of published reports to investigate the quality and scope of the current evaluation procedures in place. Building on institutional and evaluation literatures, the research offers an enhanced understanding of the content, scope and processes of ex-post evaluation of policy instruments.

We focus specifically on how evaluations effect subsequent policy design and calibration. In this way, we explore the potential impact of an evidence loop of policy instrument implementation on policymaking. Our analysis highlights methodological, scope and institutional limitations that effect the generation and use of evidence in the German domestic buildings sector. We identify procedural and policy options to help improve these processes with an aim to contributing to enabling more effective policymaking and implementation.

Ex-post evaluation quality has direct implications for ex-ante planning, including establishment of targets and strategies. Inaccuracies in evaluating the performance of policy instruments after their implementation can lead to inaccuracies in projecting the future greenhouse gas (GHG) reduction effects of planned policy mixes. This becomes particularly important considering the increased prominence of the “Projektionsbericht” in driving the reform of German climate policy mix, as envisioned in the draft Novelle of

the Federal Climate Laws. Such inaccuracies could have significant implications for steering German climate policy. Reporting requirements under the KSG (Klimaschutzgesetz) are relatively recent, and the evaluation processes have not been sufficiently adapted to incorporate necessary criteria and include all policy programmes.

The scope of current evaluations is limited. Several key indicators need more attention and assessment methodologies to be developed in order to help make informed and strategic policy planning and design choices. The most notable omissions are distributional impacts, governance capacities, and dynamic cost effectiveness.

The GEG is not currently evaluated, nor is there an ex-post evaluation planned. Regulatory costs are not considered as direct costs to the government, and this area of policy has not received the same level of attention as fiscal spending. This overlooks the potential macroeconomic and welfare effects of introducing regulations, as well as the administrative costs required to effectively administer and credibly enforce them to ensure effectiveness.

The evaluation of regulations faces several key challenges related to data availability, enforcement, and accountability. The absence of data on energy use before the implementation of energy efficiency standards, makes it challenging to establish the baseline energy consumption and assess the impact of regulations. Furthermore, there is a lack of data on the effects of regulations after their implementation, mainly due to the absence of reporting requirements. Without comprehensive data on energy consumption patterns and performance indicators, accurately estimating the effectiveness of regulatory measures becomes a challenge.

A major challenge undermining the effectiveness of regulations is the lack of enforcement. Even if reforms are implemented to improve data provision and access, the effectiveness of these measures heavily relies on a robust inspectorate. Existing data protection laws create difficulties for the federal government in accessing regional data, which contribute to a lack of accountability. Consequently, the credibility and effectiveness of the inspectorate regime responsible for enforcing regulations are undermined.

Key recommendations are:

- Development of a publicly accessible anonymised building stock database, including current building envelope efficiency standards, energy carrier and heat source efficiency rating.
- Expansion of local inspectorate training and expertise, and introduction of reporting requirements to federal government.
- Further standardisation of procedures, estimations, and assumptions across agencies and consultancies.
- Expand scope of evaluations and assessments to improve focus on socio-economic impacts, dynamic cost effectiveness, and more explicit treatment of governance and administration.
- Increase transparency of assumptions and parameters in modelling for top-down assessments and evaluations.
- Digitisation of data within funding agencies and increased accessibility, including public access.
- Enhance reporting and accountability for evaluations and assessments, beyond current scrutiny from Bundesrechnungshof of state budgetary spending and economic efficiency.
- Implement more accurate methods of measuring the GHG reduction effectiveness of instruments, especially energy efficiency measures.

1. Introduction

Reducing greenhouse gas (GHG) emissions from the residential buildings sector is a major climate policy challenge. Progress in decarbonising the existing building stock has been slow in Europe and across the world. Germany has ambitious legally binding climate policy sector targets for GHG mitigation and has made significant progress in advancing evaluation procedures for domestic building policy in the past decade. Despite which, the decarbonisation of buildings has made less progress than other sectors. The effectiveness of previously implemented policy in Germany has been underwhelming, having missed the Climate Action Plan (KSG) sector targets twice previously (ERK, 2022).

Unprecedented climate policy ambition is needed for transformation and meeting GHG mitigation objectives. Key reforms include: the expansion of policy mixes to target multiple market and systemic challenges; the ratcheting-up of existing policy instrument stringencies; the removal of barriers to entry for emerging technologies; resolving distributional conflicts; minimising negative interactions between instruments; and addressing unintended outcomes. Critically, adaptive policy mix design needs to evolve to tackle changing conditions and uncertainties over time (Edmondson et al., 2023, 2022). To respond to these multi-faceted challenges requires reflexive governance processes: Reliable and timely production of evaluative data is needed to update and adapt policies dynamically, otherwise policymakers are constrained to drawing from a limited evidence-base, estimation, and may employ normative approaches of decision making (Fishburn, 1988).

Evaluation processes are needed to effectively manage energy transitions and update policy mixes. As numerous instruments are implemented to address these challenges, large numbers can accumulate through layering which adds governance challenges (Berneiser et al., 2021; Meyer et al., 2021). Accounting for instrument interactions is a key issue in the evaluation of policy mixes. At the most fundamental, resource constrained decision-makers need to know how to most effectively allocate funds to achieve most significant impacts on GHG abatement.

Improving capacities for reflexive governance can help reduce the likelihood of governance failures and can increase adaptability as expected changes in conditions arise.

While there is an inherent and unavoidable amount of ex-ante uncertainty and complexity given the sheer scale and speed at which sector transformation has to occur in order to meet mitigation targets (Edmondson et al., 2022), reliable monitoring of policy effects can help improve the reliability of ex-ante approximations. This is particularly needed when scaling up previously implemented policy instruments to unprecedented stringency levels, or when trialling novel instruments or design features. The scope of the criteria included in the evaluations is also significant for the prospects for reform. If important effects are excluded from the evaluation of policies (e.g., distributional impacts) there is a lack of reliable evidence on which to base assessment and underlying assumptions about policy options. In these instances, more discursive or political narratives may play more significance, despite whether these are evidence-based. Consequently, what is included or excluded in evaluations can critically affect the framing of policy alternatives.

Institutional perspectives help identify structural dimensions of key dynamics in the policymaking process. Institutional perspectives relate to capabilities affecting the design, implementation, monitoring, and evaluation of policies. The capacities of governmental departments, and their relationships with key actors involved in the evaluation of policies are particularly important (Meckling and Nahm, 2018). These procedural arrangements and coordination with governmental bodies, agencies, and preferred contracted consultancies, largely determine the quality, scope, and subsequent usefulness of policy evaluations (Schoenefeld and Jordan, 2017). Better understanding these linkages is central for unpacking the role of evidence in policy reconfiguration and policy-making outcomes. To date, there has been limited conceptual work explicitly focussing on how institutional arrangements may enable or constrain the production and the use of evidence in climate policy process (Hildén, 2011). By drawing from institutional literatures our paper contributes towards bridging the gap between public policy and public administration in the climate policy literature (Peters, 2012). We focus on: (i) the formal, structural and procedural arrangements and coordination of policy evaluation, (ii) the quality and scope of evaluations, and (iii) how relevant evaluations are for decision making and policy reform.

The building sector is complex due to heterogeneity of the building stock and split incentives. The heterogeneity of building stock (types of dwelling, retrofit and new buildings), and actors (renters, house owners, landlords, energy companies, installers, component manufacturers etc.) create multiple interrelated complexities in targeting decarbonisation (Moore and Doyon, 2023). Multiple policy interventions are used to target different abatement options (Edmondson et al., 2020). Further, behavioural characteristics such as rebound effects, mean the policies targeting buildings do not act uniformly and makes assumptions difficult and effects challenging to predict (Galvin and Sunikka-Blank, 2016; Sunikka-Blank and Galvin, 2012). From an energy transitions perspective, the building sector has historically been less well researched than other sectors such as electricity generation and transport (Köhler et al., 2019). Recent contributions have started to fill this gap (Edmondson et al., 2020; Moore and Doyon, 2023)

Sectoral institutional perspectives remain under researched, particularly in the building sector. Most of the existing contributions in the institutional literature focus on “climate policy” more broadly and explore variation of national climate institutions (Dubash et al., 2021; Finnegan, 2022; Guy et al., 2023). Consequently, little attention has been paid towards institutional configurations and capacities for climate policymaking processes at the sectoral level. Some contributions in the sustainability transitions literature with a sectoral focus have sought to better integrate the role of institutions, but these are often considered as contextual factors, rather than an explicit focus of the policy process (Brown et al., 2013; Gillard, 2016; Kern, 2011). In particular, there has been limited research directing attention to the institutional configurations which structure policy processes at the sectoral level. To address this gap, we map the institutions corresponding to the policy design and implementation processes in German residential buildings sector, paying particular attention to the role of ex-post policy evaluations in these processes.

Our research design combined mixed methods and multiple sources of primary data sources for triangulation. We combine institutional mapping, content analysis of published reports, and expert interviews, to analyse the development and influence of insti-

tutional arrangements on policy evaluation procedures and reforms. We present our assessment approach in [Section 2](#), while our research design is detailed in [Section 3](#). The institutional configuration of the governance of residential building policy is outlined in [Section 4](#). [Section 5](#) then combines content analysis and interviews to assess the current evaluation processes in terms of scope and quality. [Section 6](#) focusses on the use and dissemination of evaluation processes in policymaking processes. [Section 7](#) proposes key recommendations for reform, before [Section 8](#) draws conclusions.

We discuss implications for policymaking in the German residential buildings sector and make policy recommendations for attaining climate policy ambitions moving towards 2030 and beyond. The report explores the broader question of how institutional configurations may facilitate or constrain the production of reliable ex-post policy evaluations and the use of evidence in dynamic and transformative policymaking processes. Germany has well-established and transparent institutional arrangements in this sector, making it an empirically rich case. In doing so, our research contributes to the salient discussion on the need for better evidence in the German building sector (Singhal et al., 2022), with a comprehensive examination of the institutional configuration which structures evaluation, and key recommendations to improve the quality and use of evidence.

2. Ex-post evaluations processes in reflexive climate governance

This section outlines our procedure for a systematic analysis of policy evaluations processes in the German residential sector. We introduce an analytical heuristic to study the role of evaluation in policy processes. The heuristic develops three frames which guide our line of enquiry throughout the research. First, a categorisation and utilisation of institutions for mapping the configuration of governance and evaluation. Second, policy design challenges and methodological considerations for assessing the scope and quality of current evaluation practices. Finally, key factors relating to the use of evaluations in the policymaking process.

Effective monitoring, evaluation and adjustment of policy is necessary for anticipatory policy mix design and reflexive governance. Successfully recalibrating policy design over time is needed to adapt to changing conditions and learning from previous implementation (Morrison, 2022). Similarly, revision of mitigation policies is needed given inherent uncertainties in carbon budgets (Michaelowa et al., 2018). The recent IPCC AR6 report indicates that the remaining carbon budget for the 1.5 target is only 50% of the previously anticipated target (IPCC 2023). The European Scientific Advisory Board on Climate Change's recent report recommends EU wide 2040 climate targets (ESABCC, 2023), which may require further amendment to the German KSG targets. This highlights the need for adaptation, and potential acceleration of ambition and policy stringency.

Recalibration over time is assisted by reliable and timely production of evidence. Without evidence, decisions become inherently assumption-driven estimations based on limited data. While there are some inherent and unavoidable limits to data provision, especially in a rapidly changing environment, even estimates need to be based on what evidence is available. Evidence can improve several functions of reflexive governance: (i) monitoring and adjustment of instrument stringencies; (ii) monitoring compliance/evsion effects; and (iii) reducing negative interactions and layering of complex (and potentially conflicting) instrument mixes. Although there has been an increased recognition of the importance of policy evaluations in the broader academic literature (Fujiwara et al., 2019), there are only a few studies that systematically compile and evaluate the effects

and outcomes of climate policy evaluations in practice. Some studies have conducted systematic reviews of ex-post climate policy evaluations, (Auld et al., 2014; Fujiwara et al., 2019; Haug et al., 2010; Huitema et al., 2011), but these studies primarily focus on the evaluation outcomes, whereas the quality of state-mandated evaluations is not considered to a large extent.

This report examines the institutional configuration, quality, and use of policy evaluations. We do so by focussing on three analytical frames related to evaluation processes: (i) the governance framework and institutional configuration of evaluation processes in the German residential building sector; (ii) the scope and quality of publicly accessible evaluations, (iii) the use of evaluations in the policy process.

2.1. Governance framework and institutional configuration of evaluation process

We categorise the institutional configuration of sectoral governance to consist of formal and structural elements. Formal institutions include major ordinance (laws and acts), policies and regulations (Kaufmann et al., 2018). Establishment of major ordinance often necessitates the establishment of programmes, and policy instruments to meet enshrined objectives. Structural institutions include ministries/governmental departments tasked with design and implementation (Thelen, 1999) and supportive institutions which assist delivery and evaluation of policy instruments (Edmondson 2023). Structural institutions are configured towards the attainment of policy objectives enshrined in formal institutions (Steinmo and Thelen, 1992). This may involve recalibration of government capacities and responsibilities within the existing structure (Hacker et al., 2015), or formation of new arrangements which are dedicated toward the attainment of a particular objective (or function, e.g. monitoring commission). These institutional elements interact through procedural rules and practices (Skogstad, 2023), such as delegation and commissioning (Kuzemko, 2016; Tosun et al., 2019).

Evaluation processes play a key role within the institutional configuration of climate policymaking. Institutions both shape and are shaped by policymaking (Peters 2000). Policymaking is a constantly evolving, non-linear process (Edmondson et al. 2019), which

plays out within the institutional configuration (Howlett and Ramesh, 2003). We consider three distinct phases of policymaking: (i) agenda setting; (ii) policy formulation and design; and (iii) policy implementation. National climate institutions have recently been demonstrated to affect phases of the policy process (Guy et al., 2023), yet the 'climate institutions' considered lack a clear definition in terms of being structural, formal, or procedural. Evaluation, monitoring, revision, and the input of evidence in decision making plays a key role in each of phases. Yet, the explicit role of evaluation within formal and structural institutional arrangements remains unexplored. We focus on this gap in the current literature, engaging substantively with the policy design and implementation phases, and the role which evaluations can play in shaping outcomes.

Agenda setting concerns how climate policy is understood as a policy problem by the state. This phase is largely dominated by political actors, and interests of competing political parties and stakeholders play out in decision making processes (Howlett, 1998). These processes commonly have a lower representation of bureaucrats, technical experts, or sectoral specialists, which can politicise the process of updating programmes at planned revision steps (Lockwood et al., 2017). Actors may influence these processes by increasing the visibility and salience of issues and framing of interests through public discourse or media (Tversky and Kahneman, 1981). We do not engage in these dynamics for the purpose of this report. We deem it sufficient to focus on the role that assessment and reporting can have on influencing this stage of the policy process, without engaging with causal links between evaluation outcomes and agenda effects.

Policy formulation combines political actors, bureaucrats, public bodies, and stakeholders. A policy subsystem is defined by a substantive issue and geographic scope, and is composed of a set of stakeholders including officials from all levels of government, representatives from multiple interest groups, and scientists/researchers (Howlett et al., 1996; Sabatier and Weible, 2014). The policy subsystem in the context of the residential sector in Germany is a configuration of stakeholders that coalesce around the objective of decarbonizing buildings. It encompasses officials from various levels of government, representatives from interest groups, and scientists/researchers (Mukherjee et al., 2021). While ministries play a significant role in overseeing key ordinances, departmental

representation is not limited to specific mandates, as other policy objectives compete for support and resources (Öberg et al., 2015). Sector-specific policy programmes are typically designed and calibrated under the responsibility of ministries, but input from advisory committees, research offices, consultancies, and think tanks also influences the calibration and updating of instruments and programmes over time (von Lüpke et al., 2022). Reliable information is crucial for effective recalibration and updating of policy instruments and programmes, necessitating comprehensive evaluations that cover a broad range of indicators to inform design and recalibration choices.

The implementation phase plays a crucial role in determining the rate and direction of socio-technical change resulting from policy formulation and design. The policy design and institutional literature often overlooks the importance of implementation beyond the effective design of policy elements. In practice, policies often fail to achieve their intended effects due to institutional factors that influence (or limit) the impact on the target group (Patashnik, 2009). While implementation has recently been considered in relation to variation of national climate institutions (Guy et al., 2023), it corresponds only to a broad conceptualisation of “state capacities”. More specifically, administration and enforcement are critical capacities, which shape policy outcomes and effects (Edmondson 2023). These are typically carried out by delegated actors such as federal agencies, public bodies, devolved administrations, or local authorities (Cairney et al., 2016; Hendriks and Tops, 2003). Delegation is sometimes necessary for effective implementation, particularly when policies require local administration (Jordan et al., 2018). Ideally, delegated actors should be coordinated or supervised by centralized federal departments to enhance accountability, enforcement, and reduce evasion. Accordingly, evaluating the governance capacities for effective policy implementation is essential. The institutions responsible for policy evaluation should be efficient, effective, and well-coordinated. Therefore, the evaluation of instruments and programmes should explicitly consider the governance capacities for implementing policies effectively.

Evaluations may be susceptible to selection bias in their scope of measured indicators and their methodologies. Since evaluations can reveal significant issues with policy and its delivery, there is a risk of biased or selective evaluations (Bovens et al., 2008;

Mastenbroek et al., 2016; Schoenefeld and Jordan, 2017). For example, governmental bodies with specific policy agendas may be less likely to challenge established policy goals during evaluations (Huitema et al., 2011). This may limit the generation of evidence which would otherwise motivate reform. Yet, independent evaluations may be constrained by limited data access (i.e., data protection laws), and even when published may have less influence in the policy process than more officially conducted or commissioned evaluations (Hildén, 2011).

2.2. The scope and quality of evaluations

Having outlined the institutional configuration of evaluation processes, we now outline factors related to the scope of indicators that should be considered in more comprehensive evaluations, and factors affecting their quality.

Scope

The scope of policy evaluation should extend to a wider set of evaluation criteria to enable sufficient data for effective design. To effectively govern the process of transition, policymakers need to have sufficient evidence to draw on (Edmondson et al., 2023). Planning processes are commonly reliant on estimations, limited data, modelling exercises and assumptions which fail to capture the complexity of real-world socio-economic conditions. When evaluating implemented policies, considered effects should extend beyond the energy savings and the cost of the programme. Multiple other issues important for political feasibility and welfare outcomes arise including distributional inequalities. Evaluation should consider the impacts of policy implementation on different income groups, and how fiscal subsidies are distributed across society (Zachmann et al., 2018). To improve these assessment and planning processes, a wide range of empirical data should be collected from existing policy implementation, which extends to potential barriers to implementation, political acceptance, and governance challenges. We build on evaluation criteria developed by Edmondson et al. (2022) as a framework for the scope of policy evaluations for the residential buildings sector (Table 1).

Challenge	Components	Analytical elements for residential building sector
Effectiveness	<i>Energy Use</i>	<ul style="list-style-type: none"> - Energy before implementation. - Energy use after implementation. - Rebound effects.
	<i>GHG abatement</i>	<ul style="list-style-type: none"> - Carbon intensity of energy carrier (emission factor). - Change over time (composition of electricity/energy mix).
	<i>Interaction effects</i>	<ul style="list-style-type: none"> - Positive interactions – synergies between instruments. - Negative interactions – conflicts which reduce energy savings. - Necessary interactions – conditions/interventions needed for other measures to achieve assumed effects (i.e. minimum efficiency standards for operation of heat pump).
Cost Effectiveness	<i>Dynamic cost effectiveness</i>	<ul style="list-style-type: none"> - Market failures, including: consumer myopia, learning by doing spillovers, R&D spillovers, network externalities. - Systemic failures: coordination, strategic and supply chain failures. - Investment effects. - Cost effectiveness over time under changing conditions (i.e., composition of energy mix). - Macro-economic effects over time. - Additionality of the programme/instrument.
	<i>Static efficiency</i>	<ul style="list-style-type: none"> - Marginal abatement costs.
Fiscal burden	<i>Costs/revenues to state</i>	<ul style="list-style-type: none"> - Fiscal costs/revenues generated from policy/programme.
Distribution	<i>Population</i>	<ul style="list-style-type: none"> - Distribution of costs and benefits among population. - Targeting of subsidies. - Direct distributional effects (subsidy). - Indirect distributional effects (energy use). - Market price of energy. - Cost allocation to landlord/tenant.
	<i>Firms</i>	<ul style="list-style-type: none"> - Distribution of costs among firms and impacts on national competitiveness. - Creation of jobs.
Acceptance	<i>Population</i>	<ul style="list-style-type: none"> - Acceptance among population groups.
	<i>Firms</i>	<ul style="list-style-type: none"> - Acceptance among industry interest groups/stakeholders.
	<i>Political</i>	<ul style="list-style-type: none"> - Support by governing political parties. - Coherence between federal government and devolved authorities.
Governance	<i>Administrative/information requirements</i>	<ul style="list-style-type: none"> - Monitoring and enforcement capacities. - Compliance rates. - Information and data provision.

Table 1 - Policy mix design challenges and respective analytical components for evaluation. Adapted from Edmondson et al. (2023)

Effectiveness is the most fundamental consideration for policy evaluation. At its most basic, policy evaluation is needed to assess the effectiveness in attainment of goals, in terms of energy savings and GHG abatement. How effectiveness is measured can be in terms of absolute metrics (such as the total energy saved, or GHG abatement), or in

terms of proxy indicators such as the number of heating systems installed as a percentage of the total stock (Edmondson et al., 2023). The first category gives a more quantifiable value for input into impact models and assessment tools, but the calculation methodology to estimate these values becomes very important. For example, inaccuracies in the baseline energy use prior to the implementation of policy and measures to target GHG reduction can overestimate their impact, i.e. 'pre-bound' effect (Galvin and Sunikka-Blank, 2016; Sunikka-Blank and Galvin, 2012).

The cost effectiveness mitigation is another significant consideration. Policies incur costs, either to fiscal budgets, or socialised and distributed throughout the economy. Cost effectiveness of policy options has been a main focus of historical German evaluations (Rosenow and Galvin, 2013). How these costs are calculated will significantly affect the viability of some policy options over others. While regulatory policies have no direct costs, they incur administrative costs (Baek and Park, 2012). Compliance also incurs short-term costs for firms and/or households, but sunk costs pay back over a longer period through reduction of heating costs (Qian et al., 2016). This creates split incentives between landlords and tenants, based on who bears the cost of compliance and who benefits (Melvin, 2018). Economic policies do not have direct costs on fiscal budgets but incur costs across population groups (Wang et al., 2016). Accordingly, to effectively evaluate a broad range of policy instrument types and meaningfully compare options, requires evaluation of a wider range of factors including: macroeconomic costs, distributional effects, and governance requirements (administrative/enforcement costs).

Evaluations should explicitly consider dynamic effects and societal impacts. To consider cost effectiveness from a more dynamic perspective requires a consideration of innovation, market and systemic failures, and policy instrument interactions over time. For example, the GHG abatement of very high standards of renovation is greatest for a household which still uses a gas-based energy carrier for space heating. However, as the energy carrier decarbonises, the GHG abatement from the renovation decreases. Since the highest rates of efficiency renovation already have steep marginal abatement costs (Galvin, 2023), these would become even higher considered dynamically over time. Similarly, market and systemic failures (Weber and Rohrer, 2012) should be considered to

assess potential issues and bottlenecks which would hinder delivery or cost effectiveness (Edmondson et al., 2023). These effects need to be anticipated and mitigated (to the extent possible) through policy design. However, these should also form a key aspect of policy evaluations. The wider impacts of policy design have societal implications. For example, the targeting of policy subsidies may lead to regressive distribution, where funding is only made available to homeowners. Equally, incentivising landlords to improve the efficiency and energy carrier of housing, whilst not incurring adverse costs on tenants is an important issue (George et al., 2023).

The governance requirements and capabilities of the state to implement and administer climate policy are important considerations to ensure the effectiveness. Assessment should not focus only on the estimated impacts of the instrument itself and draw broad assumptions about implementation and administration. Instead, evaluations should extend to the capacities of the state to effectively implement and enforce policy. Some policies are critically dependant on the monitoring and enforcement to ensure that they achieve the anticipated levels of energy savings and GHG abatement. In particular, regulatory measures need to be credibly enforced, otherwise non-compliance and evasion are likely (Garmston and Pan, 2013; Hovi et al., 2012; Lu et al., 2022). Effective implementation requires sufficient capacities (technical skills and training, budget allocation, retention/turnover of staff), coordination across ministries and agencies (horizontal), and between different levels of government (vertical). Evaluations should seek to assess these aspects of policy performance and governance capacities explicitly (Edmondson 2023).

Quality

The quality of evaluations is highly dependent on the availability and quality of input data. Unreliable input data necessitates the use of estimation and approximation for the effects of instruments and programmes. This can result in over-estimates of the effectiveness of the current policy mix (van den Bergh et al., 2021). In some instances, a poorly performing policy mix may not be recognised as such. Accordingly, ineffective instrument designs and programmes may persist and undergo small incremental changes (Jacobs and Weaver, 2015). More accurate evidence which highlights undesirable effects

or cost (in)effectiveness of the existing programme may motivate more radical reforms. Accordingly, evaluation processes should identify the quality and reliability of the input data. This should be indicated as uncertainty in evaluation outputs. More importantly, identified issues in data availabilities should be addressed to enable more reliable and robust outputs. These should be identified explicitly and motivate institutional reforms to enable more informed policymaking.

Evaluation procedures have been demonstrated to vary in quality, which limits the reliability of evidence generated. To ensure evidence-based decision making, evaluations must have sufficient methodological quality and legitimate analyses. Research has shown that there usually is a disparity between evaluation theory and practical implementation (Huitema et al., 2011). These differences can be attributed to several factors, which include the transparency, replicability and reliability of the methodologies employed and the data used.

Interaction effects need to be considered. Policy evaluation needs to account for the interaction effects of individual energy saving measures and of the instruments which make up programmes (Rosenow et al., 2017). Governance involves evaluating and calibrating large numbers of interacting instruments. At minimum, the consideration of interactions should extend to instruments which may have negative interactions and reduce the anticipated energy saving/GHG reductions, both in the short term (static) and over time (dynamically). A more sophisticated integration of interactions would also elucidate: (i) positive interactions, in which two instruments have synergistic effects greater than the sum of their parts; and (ii) necessary interactions, in which a complimentary instrument is required to create the conditions necessary for the primary instrument to achieve its goals.

Standardisation is needed to ensure reliability and comparability of evaluation outputs. Without standardisation of processes evaluators often use differing methodologies, and apply divergent assumptions or approximations (Huitema et al., 2011). This makes the dissemination of results more difficult, and comparisons across evaluations more dif-

difficult. Good evaluation governance should seek to establish comprehensive and standardised procedures of practice, with formal disclosure requirements (Magro and Wilson, 2017; Schoenefeld and Jordan, 2017).

The production of evaluations should prioritize the generation of outputs that are reliable, robust, transparent, and publicly accessible. This is essential for enhancing the transparency and accountability of policy design options. To achieve this, it is important that the outputs of evaluations demonstrate consistency across various publications. Standardized assessment methodologies and reporting requirements should be established to ensure uniformity and comparability. The data used in evaluations should be transparent, allowing for a clear understanding of the sources and methodologies employed. Furthermore, any gaps in the available data should be acknowledged and clearly identified. It is crucial that evaluation outputs are made publicly available to facilitate external scrutiny and validation. This enables a broader range of experts and civil society to participate in the evaluation process, contributing to its credibility and robustness.

2.3. The use of evaluations in the policy process

Effective evaluations of both policy instruments and state capacities are needed to motivate and inform reform. Over time policy implementation can produce feedback effects, which may motivate reform. These include perceptions of if a policy is working (cognitive effects), which contribute to perceptions of if incremental or more radical reforms to existing programmes (agenda effects) are needed (Edmondson et al. 2019). These cognitions are highly dependent on the reliability and evidence produced from evaluation processes, including how the evidence is presented and disseminated. Another element of feedback is perceptions of how well administrations tasked with policy design and implementation are performing (Oberlander and Weaver 2015). Administrative feedback can motivate the expansion of state capacities (Pierson 1993) or other forms of institutional reforms (Edmondson et al. 2019). Without supportive evidence, feedback is purely discursive and normative, lacking objective information on which to base assessments (Schmidt, 2008). In these situations, ineffective policy mix elements may remain in place without scrutiny. Additionally, more effective reforms and policy options

may not be considered, since current policy evaluations indicate the current programmes to be performing sufficiently (Weaver 2010). Accordingly, funding and human capital may be allocated to supporting programmes which do not deliver at the scale and speed needed to meet GHG abatement targets.

Ex-ante assessments and informed strategic planning of policy options necessitate the availability of reliable evidence. Ex-ante planning exercises rely on assumptions and the existing evidence base derived from previously implemented policies. Given that meeting climate targets is a long-term endeavour, it involves establishing policy trajectories and making adjustments over time. Consequently, the accuracy of these planning exercises is heavily reliant on the quality and accessibility of ex-post evaluations. Ex-post evaluations play a crucial role in assessing the outcomes and impacts of past policies, thus providing valuable insights for future planning. The availability of comprehensive and reliable ex-post evaluations is essential for enhancing the effectiveness and precision of strategic planning processes in achieving climate targets.

The integration of evidence into decision making processes could be more formally established. Decision makers could be obliged to draw on more reliable evidence in order to justify their choices, ensuring that decisions are based on a solid foundation of information and enhancing their quality and effectiveness. This formal integration of evidence can contribute to depoliticising decision-making processes by relying on evidence-based approaches, making decisions less influenced by subjective biases and personal interests. Furthermore, the inclusion of evidence allows for increased engagement from civil society, enabling them to scrutinize decision makers and hold them accountable. This participatory approach promotes transparency and democratic governance. Formal integration of evidence is crucial for holding policymakers accountable, as clear criteria for decision making and the requirement to use reliable evidence can be established. This makes it possible to evaluate and assess the outcomes and impacts of policies, ensuring that policymakers are answerable for their actions. In particular, under the Climate Protection Act, production of better evidence is needed both for review and scrutiny of progress and for the projection report (section 4.1.2.). Robust and comprehensive evidence is

essential in identifying areas where policies may be falling short in achieving their objectives. This enables targeted interventions and course corrections to be made.

Motivation for conducting evaluations and their commissioning influences their eventual utility. The use of evaluative evidence also has path dependent elements, since ministries commission these, if not conducted in-house (i.e. UBA, dena). Therefore, the specification of these reports can largely determine the scope, quality, and use. The motivation behind commissioning the reports is important. If these evaluations primarily serve to justify how a ministry has allocated funding to support programmes, then methodological bias might exist that shows cost effectiveness more favourably. Accordingly, additional factors which might impair cost effectiveness may be excluded from the evaluations, or considered in a more symbolic manner.

3. Research design and procedure

The research design of this report followed three main procedural steps. The first step included literature review and exploratory and descriptive desk-based research to characterise the governance framework in the German residential building sector. The second step included content analysis of methodological guidelines, and published publicly available ex-post evaluation reports. Finally, step 3 involved conducting expert interviews which helped corroborate findings and identify less codified procedural aspects of conducting evaluations and their impact.

Step 1: Governance framework for residential buildings in Germany and institutional configuration

Preliminary analysis followed a top-down approach to identify the governance framework in the Germany residential building sector. The first step followed a top-down approach to identify the governance framework for energy policy in the German residential building sector, focussing on the role of ex-post evaluation. It began by examining relevant EU directives and Germany's current obligations. The analysis linked the governance framework to the institutional configuration, mapping interactions across ordinances and reporting requirements. The objective was to categorise laws, strategies, and programmes, identify evaluation procedures and changes over time, and observe key themes and trends. Additionally, the institutional configuration of ministries, government bodies, agencies, and consultancies involved in programme implementation and evaluation was mapped, focusing on formal evaluative inputs and publicly available reports.

Step 2: Content analysis of evaluation guidelines and publicly accessible ex-post evaluations published after 2020

Content analysis was employed to evaluate the scope and quality of existing state-mandated evaluation processes. The coding process encompassed three main aspects: the scope of evaluation metrics utilized; the methodologies employed to assess the included metrics; and data quality, transparency and replicability. The coding approach incorporated both deductive and inductive elements for each category.

Regarding the scope, an extended set of indicators outlined in Table 1 (Section 2.2) was used as a normative set of evaluation criteria for “good practice”. Reports were coded according to indicators to determine which were accounted for, and if so, how they were defined and applied. For the measured indicators, the coding process included identifying the metrics used to measure the variable and the methodologies employed for their calculation. These metrics and methodologies were then coded in terms of their transparency, replicability, and quality.

We first coded the methodological guidelines and then assessed how these were used in practice in published evaluations. The procedure began with first coding the methodological guidelines, and then using these as a benchmark for what should be included in publicly accessible published evaluations. Published evaluation documents were then coded to identify disparities between recommended practices and actual implementation. This comparison sheds light on the differences between ideal practices and application. It also allows us to assess the extent to which procedures are standardized across ministries and consultancies.

We focus on formal reports published by ministries and contracted consultants as part of the formal evaluation procedures, after the publication of the methodological guidelines. We limit the scope to these formal reports published after 2020, since prior to that date the guidelines were less transparent and standardised. The publication of the guidelines was a response to this phenomenon, and thus detailed extrapolation of evaluations prior to their publication was not considered productive. Instead, we focus on the evaluation procedures introduced through the guidelines and evidence of their use in practice by focussing on the evaluations published after their circulation.

To enhance the traceability of our research design, we concentrate on formal and publicly accessible reports. This ensures a higher level of transparency in our analysis. The impact of un-commissioned policy evaluations and external reports is less traceable. Ministries are not obliged to respond to or acknowledge these publications, making it speculative to evaluate their use or impact. Furthermore, there are ongoing evaluations that are less formalized and not subject to reporting requirements, creating challenges in accessing data for independent research, including our own.

To identify relevant documents, we employed a systematic web-based approach aimed at amassing a comprehensive collection of materials. First, we searched official government portals, including ministries and agencies, to gather all available policy evaluations. This ensured that we captured a wide range of relevant documents directly from authoritative sources. Second, by expanding our research beyond government portals, we explored websites of consultancies, institutes, and non-governmental organizations (NGOs). Next, to further enhance the scope of our search, we utilised general search engines, thereby enabling us to explore a broader range of resources and gather additional relevant documents (Table 2). Finally, we carefully reviewed the additional collected documents for any potential references to other ex-post evaluations, ensuring a thorough examination of the available material.

Institution type	Frequency	Examples
Federal Government/ ministry	5	Bundesregierung, Bundesministerium für Wirtschaft und Klimaschutz (BMWK), Bundesministerium für Wohnen, Stadtentwicklung und Bauwesen (BMWSB)
Federal agency	6	Bundesamt für Wirtschaft und Ausfuhrkontrolle (BAFA), Umweltbundesamt (UBA), Bundesamt für Bauwesen und Raumordnung (BBR)
Consultancy	9	Prognos, Guidehouse, PwC
NGO/ Institute	10	Fraunhofer, Öko-Institut e.V., Dena
Other	4	Expertenrat für Klimafragen, KfW

Table 2 - Websites screened for published evaluations for content analysis. Total n screened =34

During the screening process, we identified a wide range of documents. The selection of screened documents (Annex I) extended to: (a) strategies specifically related to energy efficiency; (b) ex-ante evaluations, which provide insights into the assessment of policies before their implementation; and (c) national monitoring reports that dedicated sections to energy efficiency in the building sector, thereby offering valuable information and analysis.

We selected documents which directly related to the generation of ex-post evaluations. From the screened documents, we narrowed down our selection to those directly related to the generation of ex-post evaluations (Table 3). These included:

- i. The methodological guidelines for ex-post evaluations and ex-ante assessments. These resources helped us understand how ex-post evaluations are conducted and feed into ex-ante assessments following standardised procedures.
- ii. Ex-post evaluations that were conducted subsequent to the publication of the methodological guidelines. These evaluations offered insights into the outcomes and impacts of implemented policies.
- iii. Sections within the national monitoring reports that specifically focused on residential buildings' energy efficiency. These sections provided valuable data and analysis in the context of our research.

Content analysis employed framework coding as the main data evaluation procedure.

Content analysis involves the coding of selected pieces of text according to specific coding categories (Krippendorff, 2004). Deductively, coding started with a list of a priori policy mix design challenges defined by Edmondson et al. (2023), which relate to key assessment metrics which need to be targeted to ensure success of climate policy (Table 1). We followed (Mayring, 2000) and first created codebooks in MAXQDA with a standardised definition, indication, exemplar recommendation and coding rules for each category. This improved inter-coder reliability and the replicability of our method.

Additional codes were added inductively through application of the codebook. New sub-codes were added inductively while coding the methodological guidelines to capture how the broad categories were applied in practice (where appropriate). This incremental process combined deductive and inductive elements. Procedurally, an inductive code book was created which included metrics, methodologies, relating to the use of the assessment criteria in the evaluation documents. Reports were double coded to improve reliability and replicability. Throughout, codes were iteratively reviewed, merged, or aggregated to avoid duplication, as per Krippendorff (2004).

Type	Title	Publication Year	Authors	Client
Ex-post evaluation	Evaluation und Perspektiven des Markt-anreizprogramms zur Förderung von Maßnahmen zur Nutzung erneuerbarer Energien im Wärmemarkt im Förderzeitraum 2019 bis 2020	n.d.	ifeu, Fraunhofer, Fichtner GmbH	BMWK
Ex-post evaluation	Abschlussbericht zur Evaluation der Richtlinie über die Förderung der Heizungsoptimierung durch hocheffiziente Pumpen und hydraulischen Abgleich	2022	Arepo, Wuppertal Institut	BfEE
Formative Ex-post evaluation	Evaluation des Förderprogramms KfW 433	2022	Prognos	BMWK
Formative Ex-post evaluation	Evaluation der Förderprogramme EBS WG im Förderzeitraum 2018	2022	Prognos, FIW München	BMWK
Formative Ex-post evaluation	Förderwirkungen BEG EM 2021	2023	Prognos, ifeu, FIW München, iTG	BMWK
Formative Ex-post evaluation	Förderwirkungen BEG WG 2021	2023	Prognos, ifeu, FIW München, iTG	BMWK
Guidelines	Methodikpapier zur ex-ante Abschätzung der Energie- und THG-Minderungswirkung von energie- und klimaschutzpolitischen Maßnahmen	2022	Prognos, Fraunhofer ISI, Öko-Institut e.V.	BMWK
Guidelines	Methodikleitfaden für Evaluationen von Energie-effizienzmaßnahmen des BMWi (Projekt Nr. 63/15 – Aufstockung)	2020	Fraunhofer ISI, Prognos, ifeu, Stiftung Umweltenergie-recht,	BMWK
National monitoring	Klimaschutz in Zahlen	2022	BMWK	n.a.
National monitoring	Klimaschutzbericht 2022	2022	BMWK	n.a.
National monitoring	Zweijahresgutachten 2022 Gutachten zu bisherigen Entwicklungen der Treibhausgasemissionen, Trends der Jahresemissionsmengen und Wirksamkeit von Maßnahmen	2022	Expertenrat für Klimafragen	n.a.
National monitoring	The Energy of the Future 8 th Monitoring Report on the Energy Transition – Reporting Years 2018 and 2019	2021	BMWK	n.a.
National monitoring	Energieeffizienz in Zahlen	2021	BMWK	n.a.
Other	Begleitung von BMWK-Maßnahmen zur Umsetzung einer Wärmepumpen-Offensive	2023	Dena, Guidehouse, iTG, Öko-Institut, Prognos, EY, pwc, bbh, FIW München, ifeu, heimrich + hannot	BMWK
Strategy	National Action Plan on Energy Efficiency	2014	BMWK	n.a.

Table 3 - Documents included in content analysis. n = 15

Step 3: Interviews

The final stage involved conducting expert interviews. A part of our research motivation lies in examining the impact and utilisation of evaluations in policymaking processes, as well as the broader application of evidence in decision-making. It is worth noting that these processes often lack formalisation and are commonly not codified, making them inaccessible through content analysis. Hence, we rely on interviews as a means to thoroughly investigate these themes. The interviews aim to qualitatively explore the influence of evaluative processes on decision-making, as well as their connections to ex-ante assessment and forecasting. While decision-making is a crucial element of dynamic policymaking, we acknowledge that systematically dissecting these typically opaque processes was beyond the scope of our analysis. Furthermore, the interviews serve to complement the findings derived from content analysis and descriptive mapping of data.

We interviewed a range of different actor groups to represent different viewpoints. The sample of interviewees chosen for our study comprised various stakeholders, including ministries, agencies, public bodies, consultancies, experts, think tanks, and academia. In addition to scoping interviews, 12 interviews were conducted, following a semi-structured format (Annex II) with a duration of approximately 60 minutes each (Table 4). Among the interviewees, we engaged with key representatives from major consultancies (Prognos, Fraunhofer ISI, Öko Institut e.V., Guidehouse), as well as civil servants from the BMWK and BMWSB. However, it is important to note that one significant omission was the absence of federal funding agencies (KfW and BAFA) in our interviews. Despite multiple attempts to secure their participation, our requests were ultimately declined. We acknowledge this omission as a limitation in our research design since an important actor group was underrepresented. Nevertheless, we were able to address the coordination aspect with these funding agencies through discussions with other interview participants, thus partially mitigating the gap created by their absence.

Interview	Actor group	Role/position	Organisation/Affiliation	Length (mins)
1	Scoping interview			45
2	Scoping interview			50
3	Policymaker	Bureaucrat	Federal Ministries	60
4	Policymaker	Bureaucrat		45
5	Policymaker	Bureaucrat	Federal Agency	57
6	Consultancy	Evaluator	Evaluation Consortium	58
7	Consultancy	Evaluator		56
8	Consultancy	Evaluator		115
9	Consultancy	Evaluator		57
10	Consultancy	Evaluator		59
11	Consultancy	Evaluator	Independent evaluator	58
12	Consultancy	Programme lead	Think tank	65
13	Expert	Academic	Monitoring Commission	55
14	Expert	Academic	University	71

Table 4 - Interview Participants

Interviews were transcribed, stored and coded using MAXQDA. Interviews followed a similar procedural logic in coding practice while also following a more inductive approach to creating a codebook. The purpose of the interviews was to explore views on procedural aspects of the evaluation process, including potential gaps, issues, or challenges in the production of evaluations, and their dissemination and use in policymaking. Accordingly, the codebook extended to institutional arrangements, including organisational structures, procedural logics, and constraining rules. Codes were iteratively reviewed, merged or aggregated (Krippendorff 2004).

Limitations

Access to interview participants reduced pluralism of groups represented in our interview sample. Our study encountered limitations due to restricted access to interview participants. Notably, we were unable to include representatives from BAFA and KfW, both crucial Federal agencies responsible for administering financial support programmes in the BEG. Despite reaching out to individuals from these organisations, our invitations to participate were declined. While this absence is a gap in our sample, we include multiple participants from other actor groups who directly interact with these agencies. By triangulating information from various sources, along with our interview

findings, we believe that our research remains valid, even without the representation of BAFA and KfW.

Relatively limited availability of publicly accessible published evaluations. We examined the availability of publicly accessible published evaluations, focusing on reports published after the release of the methodological guidelines in 2020. The guidelines were developed to address the lack of standardisation across consultancies. Therefore, our analysis did not involve a detailed assessment of reports prior to 2020. Instead, we concentrated on: (i) the content of the guidelines to identify areas that still required standardisation, and (ii) determining the extent to which the guidelines were being followed in practice. For the second question, we analysed the content of published evaluations from 2020 onwards. However, our sample size was relatively small (n=14) due to a lower number of reports published during this period.

Generalizability of a single case study. Although this analysis is based on a single case study, the findings have implications beyond the specific context and can contribute to institutional learning and comparisons within other jurisdictions. The identified themes align with the broader literature on policy evaluation and monitoring, adding to the existing knowledge base. These themes include the limited enforcement of energy efficiency regulations, which is a commonly reported challenge that undermines the credibility and effectiveness of energy standards. The study also highlights the need for improved vertical coordination between different levels of government and emphasises the importance of robust methodological and analytical approaches to accurately monitor policy progress and assess the impact on greenhouse gas emissions. These themes provide valuable insights for on the generation and utilisation of evidence in climate policy processes. Further attention is needed to enhance the quality of measurements for effective policy design and calibration, as well as to enhance decision-making transparency and accountability.

4. Institutional configuration of governance of residential buildings in Germany

We map the configuration of formal ordinance for the residential building sector in Section 4.1., paying attention to reporting requirements. In Section 4.2. we map the structural configuration for two main programmes for delivery in this sector the BEG and the GEG. Section 4.3. then examines the relationship between the state and the consultancies commissioned to conduct evaluations in this sector.

4.1. Ordinance and reporting requirements

The ordinance and reporting arrangements for residential buildings in Germany spans multiple interacting levels of governance. The EU sets regulatory frameworks and has reporting requirements for member states (section 4.1). Germany has a climate law which establishes building sector GHG targets and is delivered through policy programmes under the mandate of Federal Ministries (Section 4.2.1.). Under the Climate Action Programme 2030, funding programmes are administered through Federal Agencies (4.2.1.), while regulatory are delivered and enforced at the regional level (section 4.2.3.). More information of on the content of respective ordinance is included Annex III.

4.1.1. EU directives and reporting

The EU Sets the overall framework for member states through Directives. The Energy Performance Buildings Directive (EPBD), the Energy Efficiency Directives EU/2018/2002 and 2012/27/EU (EED) establish many command-and-control polices for the energy performance of the building stock. The Renewable Energy Directive EU/2018/2001 (REDII) provides a framework for the expansion of renewable energy in the heating and cooling market at European level. Among other things, Article 14(1) EED obliges Member States to produce a report on heating and cooling efficiency every five years.

EPBD Article 7 (existing buildings) requires member states to explain where they deviate from EPB standards. Although the new EPBD does not force the Member States to

apply the set of EPB standards, the obligation to describe the national calculation methodology following the national annexes of the overarching standards will push the Member States to explain where and why they deviate from these standards. This reporting is intended to drive implementation across member states *“[it] will lead to an increased recognition and promotion of the set of EPB standards across the Member States and will have a positive impact on the implementation of the Directive.”*

National energy and climate plans (NECP) reporting requirements draw primarily from national ex-ante assessments. The national energy and climate plans (NECPs) were introduced by *the Regulation on the governance of the energy union and climate action* (EU)2018/1999. The last reporting was included in the NECP in June 2020, and the next NECP will be due in June 2023.

Reporting requirements will be further consolidated under the EU Governance Directive. In the future, the European Union (EU) will consolidate the reporting obligations at the European level through the EU Governance Regulation (Directive 2018/1999), which was adopted at the end of 2018. This consolidation encompasses various obligations, such as the National Energy and Climate Plans (NECPs), the EU Energy Efficiency Directive (EED; Directive 2012/27/EU and revised 2018/844), the EU guidelines for evaluating state aid, and the European standards for CO₂ monitoring and reporting.

EU reporting influences national policy implementation but does not carry penalties for non-compliance. EU reporting requirements are influential for policymaking and decision making and have to a large extent shaped the reporting procedures. They do not, however, carry penalties for non-compliance, and thus result in less accountability for the national Government than if clear penalties were established, or if commitments were legally binding.

4.1.2. Cross-sectoral ordinance

Climate Protection Act (KSG)

The KSG includes a legally binding sector target: GHG emissions in the building sector must be reduced by 68% by 2030 compared to 1990, with linearly declining annual targets throughout the 2020s. The revised Federal Climate Protection Act (KSG) establishes the commitment of the Federal Government to decrease greenhouse gas emissions by 65% by 2030 compared to 1990 levels and achieve greenhouse gas neutrality by 2045. The KSG also sets specific targets for different sectors. Regarding buildings, the aim is to reach 67 million t CO₂eq. in 2030, which translates to a 68% reduction from the 1990 levels.

The climate goals are reviewed through continuous monitoring. Every two years, the German Council of Experts on Climate Change presents a report of the goals achieved, as well as measures and trends. The first Biennial report was prepared in 2022 (ERK 2022a). The report is subject to reporting requirements in the Bundestag, the EU and the UN. National reporting requirements carry more weight than the EU level, which are non-binding. If instruments cannot be proved effective, they must be reformulated or replaced. The council also prepares assessment on other aspects of climate mitigation. The council audited the Sofortprogramm's effects on the buildings and transport sectors (ERK 2022b). More recently, the Review Report of the Germany Greenhouse Gas Emissions for the Year 2022 was published 17.03.2022 (ERK 2023).

Ordinance	Category (instruments)	Sector(s)	Monitoring	Frequency/ Year	Reporting	Format (methodology)	
EU	National Energy and Climate Plans (NECPs)	Directive	Cross-sectoral	NECP reporting	3 years	EU	Ex-ante
	Energy Efficiency Directives (EED)	Directive	Buildings	Article 14	5 years	EU	Ex-ante
	Energy Performance Buildings Directive (EPBD)	Directive	Buildings	Article 7	Biennial	EU	Ex-ante
National	Climate Protection Act (KSG)	Climate Law	Cross-sectoral	Section 12 (1)	Annual	Bundestag	Hybrid (aggregation)
				Projections report	Biennial	UN, EU, Bundestag	
	Climate Action Programme 2030	Strategy	Cross-sectoral	Monitoring report	Annual	Bundestag, BRH	Hybrid (aggregation)
				Progress report	3 years		
	Energy Efficiency Strategy 2050	Strategy	Cross-sectoral	Interim report	2020	Bundestag, BRH	Hybrid (aggregation)
				Monitoring report	2022		
	National Action Plan on Energy Efficiency (NAPE)	Programme (aggregate)	Cross-sectoral	NAPE monitoring	Annual	Bundestag, BRH	Ex-post (aggregation)
	The Fuel Emissions Trading Act (BEHG)	Programme (CO ₂ tax)	Buildings/Transport	Progress report	Biennial	Bundestag	Ex-post (top-down)
Federal Funding for Efficient Buildings (BEG)	Programme (funding)	Buildings	Reporting	Annual	Bundestag, BRH	Formative Ex-post (bottom-up)	
			BEG evaluation	3 years		Ex-post (bottom-up)	
			BEG evaluation	4 years	Bundestag, BRH	Ex-post (bottom-up)	
Building Energy Act (GEG)	Programme (regulatory)	Buildings	N/A	N/A	N/A	N/A	

Table 5 - Formal institutions (ordinance) in the residential buildings sector, along with monitoring reporting and evaluation requirements. The red emphasised cell indicates that BEG is the only ordinance which is evaluated through a bottom-up ex-post methodology. The orange colouring indicates there are no published evaluations available.

The German government sees the projection report (Projektionsbericht), as a central monitoring mechanism for its climate protection policy. The impact assessment is mandated by the KSG [Section 9 (2) KSG], and is reported to the Bundestag and the Bundesrechnungshof (BRH) (BRH 2022: 38). In November 2021, the Federal Government adopted the Projection Report 2021 for the year 2020. For the first time, this report contains a forecast (ex-ante assessment) of the expected mitigation effect of the current climate protection measures (BRH 2022: 39). The next report (Projektionsbericht 2023) is still currently being developed.

Review by the Bundesrechnungshof (BRH) indicates the Climate Protection Report falls short of its monitoring objectives. BRH claim the report lacks important information such as the GHG reductions that the federal government expects from the individual climate protection measures or has achieved with them so far (BRH 2022: 37). According to the BRH, the previous climate protection reports did not contain any information on the effects achieved by the current climate protection measures, but claim the corresponding data was available to the Federal Government.

Current projections indicate that the existing climate protection measures are insufficient to meet these ambitious legal targets. Consequently, Germany is projected to face a reduction gap of 195 million t CO₂eq in 2030, which accounts for 27% of the total emissions in 2020. In the buildings sector, emissions are estimated to only decrease by approximately 57%, a shortfall of 24 million t CO₂eq (Umweltbundesamt 2022).

National Strategies/Programmes

The Climate Action Programme is the national cross-sectoral climate protection strategy. First introduced in 2014, there have been two subsequent strategies released in 2021 and 2022. The strategy covers all sectors, including multiple policy instrument types including support programmes, tax relief and regulatory measures.

The annual monitoring report is the core of the monitoring process for the energy transition. Every three years, instead of the monitoring report, the more detailed progress report on the energy transition is presented. On 3 December 2014, the Federal Government published such a progress report for the first time. With this report, the

Federal Government simultaneously fulfilled its reporting obligations under Section 63 (1) of the Energy Industry Act (EnWG), Section 98 of the Renewable Energy Sources Act (EEG), and Section 24 of the Core Energy Market Data Register Ordinance (MaStRV), as well as under the National Action Plan on Energy Efficiency (NAPE) and the Energy Efficiency Strategy for Buildings (ESG). This reporting process condenses a large amount of available energy statistical information. Measures that have already been implemented are included in the analysis, as is the question in which areas efforts will be required in the future.

A commission of independent energy experts oversees the monitoring process. Based on scientific evidence, the members of the commission subsequently give their opinions on the Federal Government's monitoring and progress reports. The report is prepared within the framework of two research projects supervised by the Federal Center for Energy Efficiency (BfEE) at the Federal Office of Economics and Export Control (BAFA) and the Federal Environment Agency (UBA).

National strategies have auditing requirements. National spending strategies are subject to reporting and assessment from the Bundesrechnungshof (BRH), the supreme federal authority for audit matters in the Federal Republic of Germany. For all instruments measures which are relevant for public financing at the national level, BRH oversee auditing the success of these programmes, including how funding is spent and what is the efficiency of the subsidy programmes. They require figures for all financing relevant programmes every year, which is collected through commissioning ex-post evaluations [Interview 7, 8]. This has been conducted since the climate Action Programme 2020, which also included the National Action Plan energy efficiency in 2014, when these processes were established [Interview 7, 8].

Issues in monitoring were identified by BRH in previous assessment periods. The BRH's assessment of the 2020 programme, identified significant problems in monitoring, as the GHG savings effects of the individual measures are not specified, and a large part of the measures do not directly contribute to a reduction. The BMUV justified its po-

sition by arguing that the projection report (every 2 years) did not consider recent developments such as the Immediate Climate Protection Programme 2022, or the increase in the price of certificates in the EU-ETS since the beginning of 2021.

Climate Action Programme 2030

Climate Action Programme 2030 enshrines the target to reduce emissions from the building sector to 72 million tonnes of CO₂ per year (BReg2019). The instrument-mix will consist of: increased subsidies, CO₂ pricing and regulatory measures (BReg2019). Tax deductions for energy-efficient building renovations have also been implemented. Energy-efficient renovation measures such as replacing heating systems, installing new windows, and insulating roofs and exterior walls are to receive tax incentives from 2020. Building owners of all income classes will benefit equally through a tax deduction. The funding rates of the existing KfW funding programmes were increased by 10 % (BReg2019).

The Climate Protection Programme 2030 includes 96 sectoral and cross-sectoral measures to reduce emissions. The programme does not contain target values for the GHG reduction for the individual measures (BRH 2022: 17). The programme continues some measures of the 2020 programme that have demonstrably not contributed to a GHG reduction (BRH 2022: 17).

Energy Efficiency Strategy 2050

Germany's Energy Efficiency Strategy 2050 serves as a framework to enhance the country's energy efficiency policies. By doing so, it aligns with the European Union's energy efficiency target of reducing primary and final energy consumption by at least 32.5% by 2030. The strategy establishes an energy efficiency target for 2030 and consolidates the required measures in a new National Energy Efficiency Action Plan (NAPE 2.0). Moreover, it provides guidelines on how the dialogue process for the Energy Efficiency Roadmap 2050 should be structured, promoting effective stakeholder engagement and collaboration.

Energy Efficiency Strategy is reported to the EU in fulfilment of Germany's NECP requirements and Art. 7 EED. In 2023, the EU will consider whether the Europe-wide reduction targets need to be increased. There are also plans to draft a monitoring report in Germany. The report will look at whether the efficiency target for 2030 is still appropriate in view of the long-term goal of achieving greenhouse gas neutrality or whether it needs to be tightened. The Federal Agency for Energy Efficiency (BfEE) supports the BMWK in the implementation of the Roadmap Energy Efficiency 2050. The implementation of the Roadmap process is supported by a consortium around Prognos. A scientific support group appointed by the BMWK ensures the integration of the BMWK's research platforms into the roadmap process. The administrative tasks are carried out by the German Energy Agency (dena).

National Action Plan on Energy Efficiency

The NAPE (National Action Plan on Energy Efficiency) was designed as a comprehensive set of measures aimed at improving energy efficiency in Germany. NAPE was first implemented in 2014, alongside the climate action plan 2020. NAPE 2.0, was implemented in 2019 as part of the Energy Efficiency Strategy 2050 (BMWi 2019). It incorporated policy learnings and adapts to new developments, particularly focusing on the timeframe from 2021 to 2030. NAPE 2.0 will be again updated this year in alignment with the new roadmap for energy efficiency "Energieeffizienz für eine klimaneutrale Zukunft 2045".

NAPE has annual monitoring requirements. NAPE has been monitored annually since 2014 under the mandate of BfEE. However, the last publicly available report was published in 2021 for the reporting period of 2018-2019 (NAPE-monitoring-2021). It is unclear if this lack of publication is due to the change in government, COVID, the Ukraine conflict, or a combination of all of these factors.

Aggregate evaluation informs the NAPE reporting. NAPE reporting is the main mechanisms for evaluation energy efficiency measures. This is a large multi-sectoral report which includes the building sector. The aggregate NAPE reporting draws from bottom-up ex-post evaluations, modelling, and statistical extrapolation.

4.1.3. Sectoral programmes

The Fuel Emissions Trading Act (BEHG)

The Fuel Emissions Trading Act (BEHG) established a carbon price mechanism for buildings and transport until 2026. BHEG incorporated all fuel emissions outside the EU Emissions Trading System (EU ETS) into the national emissions trading system (nationalen Emissionshandel -nEHS) since 2021. From 2021 to 2025, the nEHS operates as an emissions trading system with fixed CO₂ prices that increase annually, starting at 25 euros/t in 2021 and reaching 45 euros/t in 2025. From 2026 onwards, a price corridor of 55 to 65 euros/t CO₂ will be implemented, however the future of the BHEG beyond 2026 remains undefined.

The BEHG has reporting requirements and revision steps every two years, starting in 2022. The Federal Government is obligated to conduct evaluations of the Act (BMUV 2021). These evaluations are required to be submitted as progress reports to the Bundestag by November 30, 2022, and November 30, 2024. Subsequently, evaluations are to be conducted every four years thereafter. The progress reports must focus on the implementation status and effectiveness of the national emissions trading scheme. They should also address the impacts of fixed prices and price corridors outlined in Section 10, Subsection (2) of the Act. Based on these findings, the government proposes necessary legal amendments to adapt and refine the emissions trading scheme. Additionally, the government is required to consider the annual climate action reports specified in Section 10 of the Federal Climate Change Act (Bundes-Klimaschutzgesetz) during this process.

At the first revision step the price trajectory was paused due to gas price volatility and pressures on the costs of heating driven by trade shocks arising from the Ukraine conflict. As part of the German government's third relief package in the beginning of 2022, the coalition committee decided to postpone the planned price increases in the nEHS by one year (at a time), starting from 2023. This decision was enacted through an amendment to the BEHG, which became effective in November 2022.

Federal Funding for Efficient Buildings (BEG)

An important building block of the Climate Action Programme 2030 is the Federal Funding for Efficient Buildings (BEG). BEG bundled the existing building funding programmes (CO₂ Building Modernisation Programme, Market Incentive Programme (MAP), Energy Efficiency Incentive Programme (APEE) and Heating Optimisation Funding Programme (HZO)) commencing in 2021 in a new system that aims to meet the needs of the target groups (BMWi 2021: 92).

The BEG is structured into three sub-programmes: the BEG Residential Buildings (BEG WG), Non-residential Buildings (BEG NWG) and Individual Measures (BEG EM). For further details on the distinction between these sub-programmes see the Annex. The providers of the promotional programmes remain KfW and BAFA (Federal Office of Economics and Export Control) (section 4.2.1.).

The first formal evaluation of the BEG was recently released (06.2023). Evaluation was coordinated by Prognos, and reports on the three aggregate sub-programmes were released independently. Aggregated data and performance metrics were last published on the BEG through the BMWK website in Q3 2021 (BMWK 2021).

Building Energy Act (GEG)

The Building Energy Act (GEG) makes up an important building block of the Climate Protection Programme 2030. The Buildings Energy Act (GEG) entered into force on 1 November 2020 (BMWi 2021: 92) and was revised on 01.01.2023. Assessment will determine if, and to what extent, the goals of the Energy Concept will be achieved in the medium to long term. It will also help indicate what new measures need to be taken.

The GEG imposes requirements on existing buildings for retrofitting and for refurbishments. The GEG imposes retrofitting requirements for certain parts (replacement of certain old boilers, insulation of certain pipelines, insulation of top floor ceilings, installation of certain control technology of heating and air-conditioning systems) independent of measures. It also establishes standards for renovation work that would otherwise be carried out to comply with energy regulations.

The responsible authority under state law may also grant exemptions from the conditional requirements upon application. These apply in particular in the event of a lack of economic viability in an individual case. However, the guidance on how these exemptions are applied are loosely defined.

4.2. Governance and ministries

This section outlines the institutional configuration of ministries, federal agencies and consultancies which contribute to the current procedures for evaluations and assessments in the German residential buildings sector (Figure 1).

Ministerial mandates, coordination, and role in the evaluation procedures are outlined while discussing these organisations. While there is coordination across the institutional structure, for the purpose of this report we arrange the institutions as corresponding to the delivery of the main two programmes: (i) Federal Funding for Efficient Buildings (BEG); and (ii) the Building Energy Act (GEG). These structural arrangements are categorised as national ministries and agencies, and sub-national agencies and devolved local authorities. Other relevant actors are also outlined in respect to their role in the evaluation of these programmes.

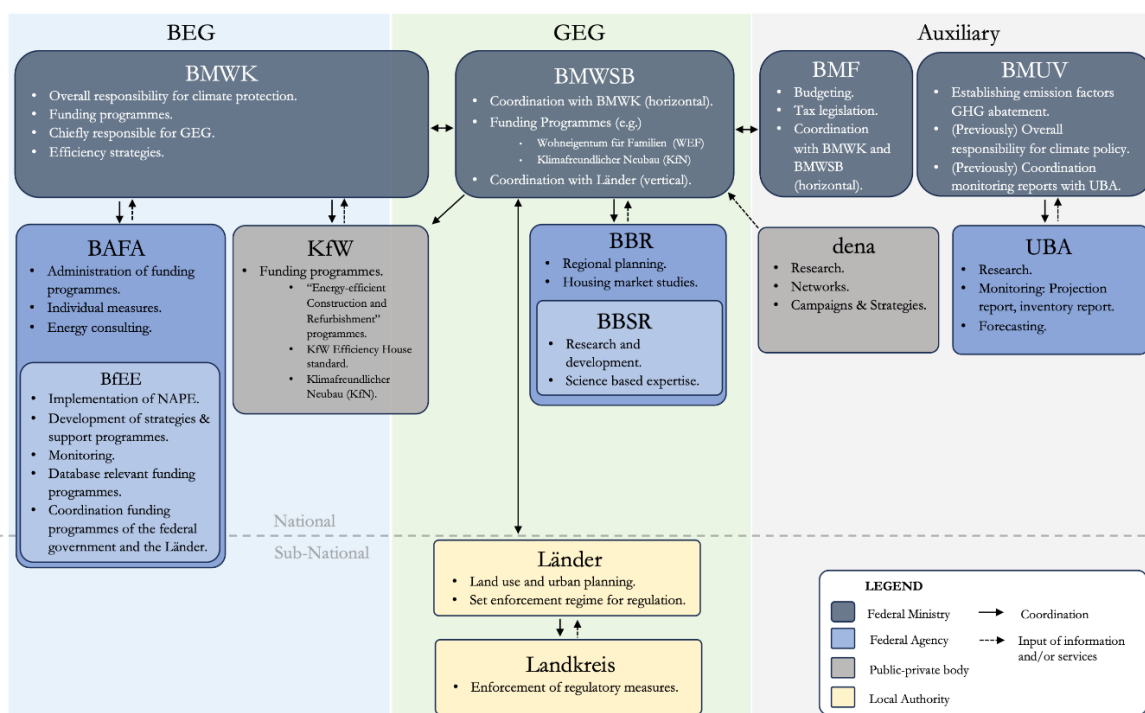


Figure 1 - Structural bureaucratic institutions for delivery and evaluation of the BEG and GEG.

4.2.1. Federal Funding for Efficient Buildings (BEG)

Federal Ministries

The Federal Ministry for Economic Affairs and Climate Protection (BMWK) has overall responsibility for energy and climate policy, and oversees delivery of the BEG. The BMWK has a broad mandate that includes various responsibilities. These include implementing the Energy Efficiency Strategy, overseeing multiple funding programmes including the BEG, managing the *EnEff.Gebäude.2050* funding initiative for significant projects, and monitoring the progress of the energy transition. In this capacity, the BMWK collaborates with federal funding agencies and holds primary responsibility for initiating evaluations and assessments through a network of consulting firms (section 4.3.). Additionally, the BMWK oversees the monitoring commissions dedicated to the energy transition ("Energiewende"). The overall BEG programme was recently evaluated (June 2023), for the period Q3 2021, published on the BMWK website. As of 2022, a survey-based evaluation of the BEG is currently ongoing¹, coordinated by consultancy Prognos².

The BEG evaluation is designed as an accompanying/formative evaluation. The main reason for this is that the complete impact can only be ascertained with a delay of three to four years (provision/call-up period until submission/audit of the proof of use). This means that at the time of the evaluation, some of the approved/funded projects have not yet been implemented; only the application data is available. Methodologically, this aspect is addressed by including the cancellation/waiver rate. The situation is similar for the KfW 433 evaluation. In both evaluations, the respective programme is or was not yet completed [Interview 7]. A special feature of the BEG was the support of the BMWK from the ongoing evaluation work in the development of the Summer 2022 guideline amendment and for the 2023 guideline, as well as other policy processes in the subject area in question [Interview 7]. In the EBS WG evaluation, on the other hand, the information on the funding output was more "reliable", as the delay was much shorter. This evaluation

¹ <https://www.energiwechsel.de/KAENEf/Redaktion/DE/Meldungen/2022/2022-03-31-befragung-fuer-evaluation-der-beg-startet.html>

² <https://www.prognos.com/de/FAQ-BEG>

therefore corresponds more to an ex-post evaluation; also because EBS WG was completed and transferred to the BEG at the time of the evaluation [Interview 7].

Federal agencies

BEG funding is coordinated through Federal Agencies. BMWK, along with the ministries for Housing, Urban Development and Construction (BMWSB), and Finance (BMF) have agreed on a joint approach to federal funding for efficient buildings (BEG). This is coordinated with and delivered through Federal Agencies: the Federal Office for Economic Affairs and Export Control (BAFA), and the Kreditanstalt für Wiederaufbau (KfW). These ministries and agencies work together to provide funding initiatives and programmes aimed at promoting energy efficiency in the construction sector. Evaluation of individual branches of the BEG is coordinated through the respective Federal Agencies and commissioned consultants.

The Federal Office for Economic Affairs and Export Control (BAFA) is responsible for administration of funding subsidies for renovation of residential buildings, new buildings and energy consulting services. *Department 5 (Energy Efficiency, Renewable Energies, Special Equalisation Scheme)* is responsible for tasks and support programmes in the field of renewable energies and energy efficiency. *Department 6 (Climate Protection Buildings, Energy Info Center, Adjustment Allowance)* is responsible for promotion programmes in the field of renewable energies. Energy consulting activities include handling of the funding procedures (application, proof of use); Federal funding for energy consulting for residential buildings (EBW) including iSFP-programme, conducted by PricewaterhouseCoopers GmbH; and approval of energy consultants.

The Federal Energy Efficiency Center (BfEE) supports the BMWK conceptually covering all aspects of energy efficiency. BfEE is a subdivision within BAFA responsible for the implementation of measures of the National Action Plan on Energy Efficiency (NAPE), the development of Efficiency Strategy on Buildings initiative, and the development of new aid programmes. The centre also runs PR campaigns promoting Energy Efficiency. Under the monitoring framework, it is responsible for the determination of Germany's energy savings and the related reporting, as well as for the monitoring and assessment of energy services markets with the objective of developing them. The monitoring process for

the energy transition “Energy of the Future” serves to review the implementation of the Energy Concept and the Federal Government’s programme of measures and to take countermeasures if targets are not met. The area of energy consumption and energy efficiency is a main topic of the monitoring process. In addition, reports are submitted on the expansion of renewable energies, GHG emissions, security of supply, grid infrastructure and the energy transition in an international context.

Public-private bodies

The Kreditanstalt für Wiederaufbau Bankengruppe (KfW) is jointly responsible for administering the BEG funding programme. Since the implementation of federal funding for energy-efficient residential buildings (BEG WG) and new buildings (BEG NWG) on July 1, 2021, the building standard has been referred to as the “KfW efficient house” (*KfW Effizienzhaus*). When it comes to refurbishment, KfW provides funding for various efficiency house standards, including Denkmal (heritage), 100, 85, 70, 55, and 40 (Table 6). For new buildings, funding was available for the efficiency house levels 55 and 40.

KfW Efficiency Standard	Energy consumption
KfW Efficiency House 100	Consumes the same as a new GEG reference building.
KfW Efficiency House 85	Requires only 85 percent of the energy of a new GEG building.
KfW Efficiency House 70	Requires only 70 percent of the energy of a new GEG building.
KfW Efficiency House 55	Requires only 55 percent of the energy of a new GEG building.
KfW Efficiency House 40	Requires only 40 percent of the energy of a new GEG building.

Table 6 - KfW Efficiency standards for renovations.

Funding for the KfW 55 standard for new buildings was recently removed, and replaced by the Klimafreundlicher Neubau (KfN) programme. The funding for the Efficiency House 55 standard for new buildings has been completely cancelled. Currently, the funding for Efficiency House 40 for new buildings is not available either. The government is planning to reorient the programme to align it more closely with potential CO₂ savings. The funding for Efficiency House 40 for new buildings continued until the end of the 2022, albeit with lower subsidy rates and a cap of one billion euros. Since 1st March 2023, funding for new buildings now available through the Klimafreundlicher Neubau (KfN) programme, overseen by BMWSB. This new funding, for which a total of 750 million

euros is available, is provided in the form of low-interest loans. The KFN programme does not provide repayment subsidies, but interest rate reductions.

Regional level (Länder)

In January 2019, a coordination office for the energy efficiency funding programmes of the federal government and the Länder was established at the BfEE. The coordination office maintains a database of relevant funding programmes and intensifies the early coordination of funding activities of the federal government and the Länder. In this way, funding agencies from the federal and state governments are supported in identifying overlaps and duplications of funding at an early stage and in better coordinating their funding programmes.

4.2.2. Building Energy Act (GEG)

Federal Ministries

Federal Ministry for Housing, Urban Development and Building (BMWSB) currently holds overall responsibility for overseeing the delivery of the GEG. The ministry was established in December 2021 following the formation of the new coalition government. Currently, there are no existing ex-post evaluation processes in place for the GEG, and the departmental website has not published any reports publicly. This can be attributed, at least in part, to the typical timeframe of 2-3 years required for conducting and publishing formal evaluations. The current process of updating the GEG is jointly coordinated between BMWK and BMWSB.

Federal agencies

The Federal Office for Building and Regional Planning (BBR) supervises federal building measures in Berlin, Bonn and abroad. BBR supports the Federal Government in various policy domains such as regional planning, urban planning, housing, and building. Its responsibilities encompass a wide range of tasks, including overseeing significant construction projects, implementing model initiatives, addressing building culture and monument preservation, handling matters related to European cooperation, organizing architectural competitions, and conducting studies on the housing market.

The Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR) is a research institution within the BBR. Since 2009, the research division of the BBR merged with the Institute for Rehabilitation and Modernization of Buildings (IEMB) to form BBSR, a departmental research institution. BBSR and provides research and development services as well as science-based expertise to the Federal Ministry.

Regional level (Länder)

The Länder may act as veto player for the implementation of Federal legislation. State representatives in the second chamber (Bundesrat) have to approve all federal legislation affecting their financial and administrative matters. This makes the Länder an important veto player in federal building policies.

States have some degree of autonomy for implementation of climate policy. States have the competences for overseeing land-use and urban planning as well as in training and education; they can initiate pilot projects and funding programmes for renewable energy, building renovation, housing promotion, or urban development; they engage in research promotion; and they can act as role model by enhancing the energy performance of state-owned buildings or adopting a climate-friendly procurement policy a considerable discretion for own climate related building policies (Jacob & Kannen 2015: 5).

A significant enforcement deficit exists in Germany for regulatory instruments. The enforcement and implementation of federal laws, including building standards, falls within the remit of the states, who are free to choose the administrative structure of their enforcement system [Interview 14]. There is little accountability at the Federal Level, since data protection laws prevents sharing of records and information [Interview 4]. Earlier research into the EnEV indicated non-compliance was estimated to reach at least 25 percent, but presumably even more (Weiß and Vogelpohl, 2010: 18). This lack of enforcement results from staff shortages in the Länder (Jacob & Kannen 2015: 11).

A systematic treatment of implementation and enforcement of regulations at the Länder level is beyond the scope of our analysis. This report considers the federal level and the reporting and evaluation processes in place. However, consideration of these signifi-

cant issues in the implementation of regulatory measures is notable, since non-compliance will have significant and adverse impacts on the estimations of policy effectiveness of the German policy mix.

4.2.3. Auxiliary

Federal Ministries

The Ministry of Finance (BMF) decides on the provision of financial resources for climate protection and energy efficiency measures. BMF allocates certificates for tax incentives for energy-efficient building refurbishment, and provides ordinance on the determination of minimum requirements for energy efficiency measures in buildings used for own residential purposes (§ 35c Income Tax Act) à (Energetic Renovation Measures Ordinance - ESanMV). The BMF oversees general budgeting (e.g. emergency programme for more climate protection 2021: 5,5b €), which means it plays a key role in assessment of fiscal spending and cost effectiveness of programmes.

BMF is responsible for determining the allocation of financial resources to support climate protection and energy efficiency measures. It plays a crucial role in deciding how funds are distributed to initiatives aimed at addressing these issues. In terms of budgeting, there have been notable initiatives to allocate funds towards climate protection. For instance, in 2021, an emergency programme (Sofortprogramm) was introduced with a budget of 5.5 billion euros, dedicated to bolstering climate protection efforts. This programme aimed to provide financial resources for various projects and initiatives aimed at combating climate change and promoting sustainability.

BMF also administers a certificate for tax incentives, but this is yet to be evaluated. The tax incentive is targeted at energy-efficient building refurbishments. This certificate serves as a documentation of compliance with energy efficiency standards and enables individuals or companies to qualify for tax benefits when conducting building renovations that meet the prescribed energy-efficient criteria. In order to establish minimum requirements for energy efficiency measures in residential buildings, the Energetic Renovation Measures Ordinance (ESanMV) has been implemented under § 35c of the Income

Tax Act. This ordinance outlines the specific criteria and guidelines that must be met to ensure energy efficiency in buildings used for residential purposes. However, as a recently adopted measure, there have been no evaluations of the tax incentive since implementation [Interview 9].

In the past, the Federal Ministry for Environment, Nature Conservation and Nuclear Safety (BMUV) held the responsibility for climate policy and all matters related to renewable energies. The BMUV played a crucial role in driving ambitious energy standards and promoting renewable energy, while the BMWi (now BMWK) and BMWBS showed less inclination towards these issues (Jacob and Kannen, 2015; Michaelowa, 2008; Wurzel, 2010). The BMUV and BMWi (now BMWK) were often seen as having contrasting regulatory ideas, interests, cultural identities, and political affiliations (Jacob and Kannen, 2015).

BMUV plays a limited role in current evaluation processes. Since the change in government, the BMUV has been stripped of its responsibilities. BMUV no longer plays a leading role in the legislation of the Building Energy Act during the current legislative period. The specific involvement of the BMUV in shaping the legislation has been significantly diminished. The main role of BMUV in the current evaluation processes is establishing the emission factors and primary energy factors for energy sources, which are used to extrapolate and calculate GHG emission savings from energy savings (section 5.4).

Federal agencies

German Environment Agency (UBA), supports the Federal Environment Ministry and coordinated environmental research. UBA's scope covers matters relating to emission control and soil conservation, waste management, water resources management and health-related environmental issues. UBA provides central services and support for environmental research by the Federal Environment Ministry and for coordinating environmental research by federal authorities.

UBA coordinates the projection report (Projektionsbericht). The projections report aims to identify the concrete climate protection instruments which can be utilised to attain the sector climate objectives outlined in the Federal Climate Protection Act by 2030. The

identification of gaps in the through the projection report can be addressed through implementation of more effective measures. The HIS-2030 scenario, which employs instrument-based modelling akin to the Projection Report, presents specific actionable alternatives. The framework data assumptions utilised in the modelling of the projection report are established as a standard practice before commencing the modelling process. Research participants collaborate with the UBA (German Environment Agency) and coordinate with relevant departments. To increase transparency, for the first time UBA published the framework data and assumptions prior to the release of the upcoming 2023 report (UBA 2022).

Public-private bodies

Dena - the German Energy Agency (Deutsche Energie-Agentur GmbH) - is a German private body owned by the federal government, and consists of multiple stakeholders.

It is listed as a lobbying organization in the EU's transparency register. Dena's main objective, as stated in its articles of association, is to provide services at national and international levels to shape and implement the energy and climate policy goals of the German government, with a particular focus on energy transition and climate protection.

One of dena's primary activities is the development of pilot projects aimed at testing the feasibility and effectiveness of energy efficiency measures. These projects serve as practical experiments to assess the viability of different energy-saving techniques. Additionally, dena works to improve the quality of planning, implementation, and monitoring of efficiency measures by establishing standards and guidelines to ensure their effective execution.

Dena plays a role in information provision by developing strategies, roadmaps, and communication platforms. It creates and manages various communication platforms, strategies, campaigns, political initiatives, and events to facilitate engagement and communication among stakeholders. These initiatives aim to raise awareness about energy transition and climate protection while encouraging participation and collaboration. In collaboration with the ifeu research institute and the Passive House Institute, dena has developed a comprehensive methodology for individual renovation roadmaps specifically tailored to residential buildings. This roadmap serves as a standardized tool used during

energy consultations for both complete refurbishments and step-by-step renovation processes. It is applicable to different types of residential buildings, including single-family houses, two-family houses, and multi-family houses. The purpose of providing this roadmap is to support and guide homeowners and professionals in making informed decisions and taking effective actions towards energy-efficient building renovations.

4.3. Consultants

Evaluations and assessments are commissioned to consultancies. The German configuration of evaluations and assessments for domestic building policy, ministries and federal agencies commission a group of consultancies to conduct the ex-post evaluations and ex-ante assessments.

Commissioning takes place through an open tendering process. In principle these are open tenders and anyone can apply [Interview 6]. In practice, there is a group of consultancies which work in this area and are commissioned to conduct both the ex-post evaluations and the ex-ante assessments (section 6.1 and Annex I). Within this consortium, tendering for certain projects results in some variation in project leads and which consultancies collaborate on certain evaluations. There also exist some smaller or independent consultancies. Between 2018-2022 Fraunhofer ISI/Prognos co-conducted the NAPE forecast (Fraunhofer lead), but since 2023 the NAPE is conducted exclusively by Fraunhofer ISI [Interview 7].

Evaluations are conducted by fairly constant consortium of evaluators. A consortium of consultancies (with some changes over the years) consistently conduct evaluations (Annex I). They have developed competencies and expertise over time, which has led to path-dependent aspects, including cumulative tacit knowledge and experience of the sector which cannot be readily replicated due to lack of publicly accessible data. For specific topics some consultancies built up expertise and partnerships are established, but it is not a closed circle [Interview 6]. These consultancies follow established practices, applying the same protocols across different sectors. However, the complex and heteroge-

neous nature of the building sector poses challenges for evaluations. This requires bespoke solutions to tackle issues such as rebound effects. The extent these challenges are accounted for in evaluation practices will be further explored in section 5.2.

Relatively small group of consultancies limits the possibility of procedural change.

The evaluation and assessment processes are primarily conducted by a limited number of consultancies. Since both ex-post evaluations and ex-ante assessments are conducted by the same group of consultants [Interview 7, 8, 9, 12], only large and completed publications are publicly available [Interview 3, 6, 7], and more technical aspects are often excluded from the reports [Interview 7] partially due to data protection laws [Interview 6], this vertical integration restricts the scope for validity checking from other sources of expertise.

The evaluations are commissioned through public tenders, with the commissioning specification defining the aims and scope of the evaluation. The commissioned scope is to assess the efficiency and effectiveness of the previously implemented programmes, and is inherently incremental. It is unlikely that these evaluations would produce radical results, for example that indicate the failure of the entire programme [Interview 12]. It depends on the client what appears in the public report and in what level of detail [Interview 7, 12]. Often there are much more harshly formulated background papers/non-public report sections [Interview 7]. The consultants in the consortium establish parameters and methodologies for conducting evaluations within the defined scope [Interview 10, 11].

Ministries specification extends to the overall aspects which should be evaluated, but not to parameters or methodologies. The consultancies are largely responsible for establishing the methodologies evaluations are conducted and the parameters of how the results are reported. This is usually set by the consultancies, and then proposals are discussed with the ministries who formally have the final decision.

You talk to the ministry and agree or check with them. So, you discuss, but in the end, it is your own decision to set the parameters. [Interview 11]

Consultants have influence over defining parameters. Due to limited time and capacities within ministries to validate evaluations, or parameters, the consultants are granted some autonomy in how the evaluations are conducted and reported.

But of course, I must also say that this is, I mean, not our main work here. So if we have like a bunch of consultancies and also some good people evaluating, then we also leave some things to them. I mean if we don't want to do it ourselves or we can't do it ourselves. So, it wouldn't really be useful to go into every, every detail, and then we just see what they come up with and sometimes, of course, we discuss these things and maybe sometimes also we ask for changes, but from the main points I think we're usually rather happy with what we get. [Interview 3].

Standardization has increased in the past ten years. These consultancies have increased efforts to standardise procedures, producing a set of guidelines [Interview, 7, 8, 9, 12], which have since been adopted by BMVK [Interview 9]. Prior to the standardisation of procedures, the calculations used by various consultancies to adjust for final energy demand, and the emission factors used in calculating GHG abatement varied across consultancies and were not transparent. This makes the existing evidence base collected in previous years hard to interpret and leads to uncertainties and limits the reliability of past data. This is particularly problematic when conducting ex-ante work since the evidence base from which to draw on is impacted. The guidelines establish a set of procedures for calculating effectiveness and costs of measures, instruments, and programmes [Interview 7, 8, 9]. One major contribution was to establish a common definition of terms that helped establish a common lexicon across consultancies [Interview 7]. Interviews have indicated that has further increased recently - especially in the wake of the Ukraine war and its impact on energy supply/prices etc. In the recent BEG evaluation (2023), for example, deviation from methodological guidelines was made transparent [Interview7].

Further standardization of protocols across consultancies is needed. While there has been much progress in the last 10 years, and the publications of the "Standard procedure" there is still much outside of the scope of these guidelines. Consultancies apply different methods to approximate energy adjustments and performance gap [Interview 9, 12]. The lack of standardization of methods means that results are not comparable.

The guidelines are also limited in their scope [Interview 12], are relatively short term and static [Interview 9], and still are applied differently across different consultancies [Interview 11]. We further discuss these points in section 5.

5. Ex-post evaluation: procedures, scope, data, and methods

This section covers factors related to the scope and quality of the current evaluations: (i) procedures and types of programmes evaluated; (ii) scope of indicators; (iii) methods; and (iv) data.

5.1. Evaluation approaches and programmes evaluated

Programmes are evaluated through top-down and bottom-up evaluation methodologies. Top-down approaches involve modelling and statistical extrapolation and are used to assess the effects of economic instruments such as taxes and carbon pricing (Figure 2). Bottom-up approaches involve the evaluation of individual instruments, and these evaluations are then aggregated through extrapolation. These methodological differences mean that the potential scope of top-down and bottom-up approaches is quite different. Our focus here is on the evaluation methods for bottom-up approaches which require evaluation of ex-post measured data.

Background data informs the statistical extrapolation of measures and of bundling of measures and programmes. Overall “framework data” established for the top-down processes. This is mainly energy prices (but also includes economic development), and usually draws from statistical data from the statistical office or other official data sources [Interview 8]. Transparency in these background conditions data is expressed as being essential for validation and comparability of results [Interview 8]. Formal data reporting protocols could be implemented to increase replicability (section 5.4.4.)

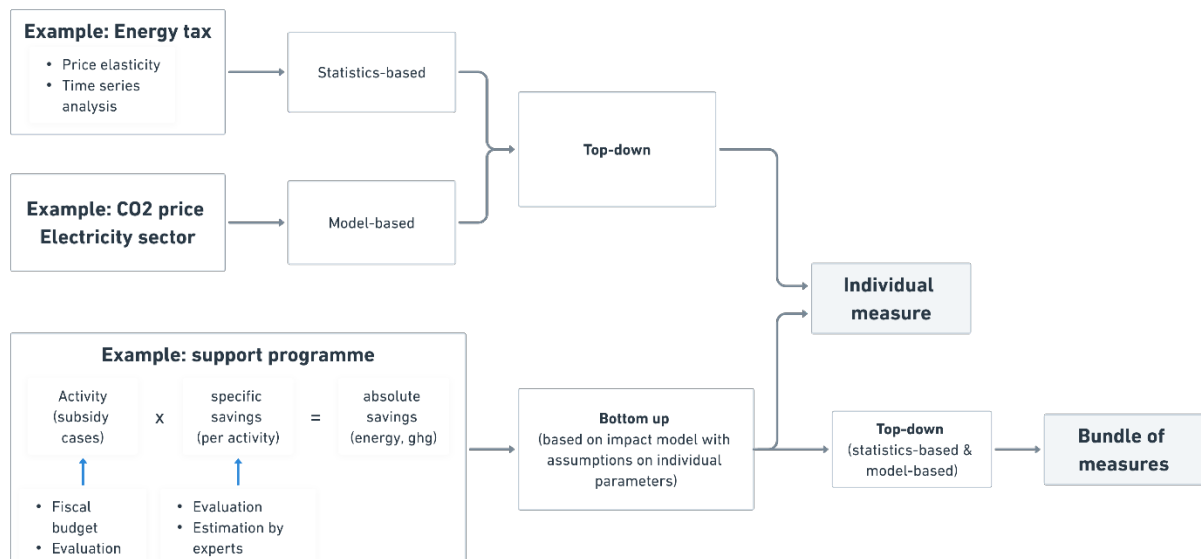


Figure 2 - Representation of bottom-up and top-down approaches. Adapted from Schlomann et al. (2020).

5.1.1. Taxes/economic instruments

The impacts of energy tax adjustments are estimated through a top-down procedure.

The CO₂ price (BHEG) is included in the evaluation methodology as top-down statistical extrapolation and modelling using price elasticity assumptions (KfW-EBS-WG-2018-2021, p. 57) (e.g., rather than empirically estimating price elasticities of the BEHG introduction). However, due to the limited scope of evaluation criteria, important considerations, notably distributional effects, are not considered. This presents a significant omission from the current procedures. Moreover, relatively static evaluation, and exclusion of more complex system interactions including innovation effects, or market failures limits the reliability of the current consideration of the effectiveness of energy tax and CO₂ price instruments.

Data availability also hinders more comprehensive evaluation. Counterfactual analysis is limited by gaps in data related to the existing building stock, and a counterfactual control group of citizens who renovate without subsidy support (KfW-EBS-WG-2018-2021, pp. 61-61). Existing delays in producing timely sectoral evidence at the level of resolution needed to look at market effects, socioeconomic impacts, and effectiveness, does not currently exist. Accordingly, taxes are only considered in terms of top-down modelling. However, effectiveness of energy and CO₂ pricing instruments in the building sector is

less predictable due to high heterogeneity. Moreover, without consideration of distributional effect in the current evaluation framework, there is not a sufficient scope for assessing what more significant increments in the fuel prices would be on the broad population:

We really tried to get data on the heating market and the building market when the CO₂ tax was introduced. To see [effects] from one month to another. It's only the market bodies, the associations collecting that data, and there's no governmental data [Interview 9]

Income tax deductions through BMF for renovations are not included in the current ex-post evaluations. This form of subsidy mechanism is not included in the evaluations [Interview 9]. This instrument was recently implemented, and the lag between policy implementation and evaluations being published may account for this omission.

5.1.2. Federal Funding for Efficient Buildings (BEG)

In practice, only the Federal Building Energy Subsidy (BEG) programme undergoes evaluations using a bottom-up approach. Evaluations are necessary for the BEG programme due to its direct costs to the government, as mandated by the Federal Audit Office (BRH). Additionally, data collection is a requirement for receiving funding under the programme.

The BEG is the only policy programme included in published ex-post evaluation processes. The ex-post guidelines focus extensively on the subsidy programmes which are administered by the federal funding agencies. The federal spending programmes are the published evaluations which following the bottom-up methodology (KfW-EBS-WG-2018-2021, KfW-433).

Data on the energy usage of recipients is collected by federal agencies as part of the application process for support mechanisms. Information is collected through funding programmes as a condition of the subsidy. This collects the energy usage before and after the measures are installed. In addition, a survey of a sample of grant recipients is carried out. With these: a) the information from the applications is validated, and b) details for the modelling of the savings ("building stock") are carried out [Interview 7].

It's always easier to evaluate when you give money because then people have to answer your questionnaire, otherwise they have to give the money back. [Interview 9]

5.1.3. Building Energy Act (GEG)

Regulatory policy, in general, is largely excluded from the current evaluation processes in Germany. The GEG does not have individual evaluation processes or reporting requirements to parliament. Within NAPE, the GEG is reported in relation to its implementation status only (Figure 3). The status is descriptive and does not include evaluative metrics. Within NAPE report, under the GEG, there are no previously conducted evaluations listed and none planned to be conducted (unlike other instruments and programmes).

Instrument	Implementation status
108. Energy Conservation Legislation for Buildings – Buildings Energy Act (GEG)	<p>Goal: On 18 June 2020, the German Bundestag adopted the Buildings Energy Act (GEG), which was put forward by the Federal Government at the proposal of the BMWi and the Federal Ministry of the Interior, Building and Community (BMI). This was confirmed by the Bundesrat on 3 July 2020. It entered into force on 1 November 2020.</p> <p>Scope: The Buildings Energy Act brings the Energy Conservation Act (EnEG), the Energy Conservation Ordinance (EnEV) and the Renewable Energies Heat Act (EEWärmeG) together and creates a new standardised and coordinated set of regulations to cover the energy requirements for new buildings, for existing buildings and for the use of renewable energy sources to supply buildings with heating and cooling. This will facilitate implementation and completion. The Building Energy Act implements the European requirements for the overall energy performance of buildings and integrates the regulation on nearly zero-energy buildings (nZEB) into the unified Energy Conservation Legislation. The Act retains the current energy standards for new buildings and modernisation projects.</p>

Figure 3 – Reporting of the GEG programme in most recently published NAPE report.

The NAPE monitoring also includes some individual regulatory programmes and instruments, but the analytical inputs or methodologies for calculating effects are not transparent. Regulations are included alongside funding and informational programmes in its reporting format. In the NAPE reports no ex-post evaluations are drawn-on for regulations, and none are planned to be conducted (BMWK 2021, p.80). Funding and informational programmes make up the substantive elements of the reports. The uniform reporting format across all types of instruments does not adequately capture the impact of regulatory instruments, or identify issues or necessary reforms within these programmes.

Where there is definitely still a gap is in the regulatory law. The Building Energy Act, Energy Saving Ordinance, etc., there is no evaluation. So the estimation of what effect a tightening of regulatory law will have on emissions in the long term is all very, very model-based and assumption-driven. There is simply no data on this. [Interview 6]

There are currently no specific requirements to evaluate the programme GEG. This is potentially since it does not incur direct costs to the government, and has not historically been scrutinized to the same extent as fiscal spending, despite incurring significant governance requirements and administrative costs [Interview 13]. It could also be attributed to the implementation regulation at the Länder level, and a lack of vertical coordination with Federal Ministries [Interview 4]. As a result, the evaluation focus has primarily centred on the BEG programme, while evaluations for the GEG programme remain limited or non-existent. There are other commissioned studies which look at specific issues such as the performance gap in refurbishment of buildings (e.g. see Jagnow and Wolff, 2020), but these do not form part of formal evaluation procedures.

5.1.4. Methodological Guidelines

The current guidelines for evaluation appear specifically tailored for assessing subsidy programmes and grants. The guidelines theoretically have applicability to other types of programmes or instruments and set out an assessment logic methodology that can incorporate them (ex-post-guidelines, p. 28). The practical guidance for indicators, however, has a clear focus of the guidelines is for evaluating subsidies. While regulatory measures are included in the intervention logic, they receive limited guidance and have the smallest number of appropriate indicators compared to other types of measures (ex-post-guidelines, pp. 49-51). Moreover, using efficiency indicators for regulatory instruments is not recommended due to their limited comparability and potential misrepresentation of funding efficiency (ex-post-guidelines, p. 62). Interviews have indicated that the guidelines reflect the focus of development at the time, and the methodology guide must be seen as a "work in progress" [Interview 7].

Guidelines differentiate between different “logics of intervention”. Regulatory measures and taxes are considered as separate logics to the subsidy programmes (ex-post-guidelines, p. 29). Measures with regulatory intervention logic usually involve the specification of binding rules in laws or regulations (legislative measures). This defines obligations (e.g., regarding behaviours, market access) or standards, the non-compliance of which is sanctioned in specific ways. Practical examples of measures with regulatory intervention logic include: regulatory measures (e.g., EnEV requirements for buildings) and taxes or levies (e.g., KWK-G).

The Federal Budget Code (Bundeshaushaltsordnung -BHO) distinguishes between regulatory and fiscal instruments. The framework for the further development of the measure, i.e., the implementation of the political directive, is provided by the Federal Budget Code (Bundeshaushaltsordnung, BHO) and its administrative regulations. Essentially, a distinction is made between two different types of measures: on the one hand, legislative regulatory measures, and on the other hand, financially effective measures. Due to the subject matter of the BHO, this differentiation is made from a budgetary perspective/logic, i.e., it is distinguished whether budget funds are provided for the implementation of the measure or not (ex-post-guidelines, p. 24).

There is a discrepancy between the evaluation focus of the BHO and the methodological guidelines. The BHO establishes an input-focused evaluation (how much do we have to spend to get a certain outcome) whereas the guidelines propose an output-focused approach that makes the effect of subsidies and regulations more comparable than simple efficiency measures:

The advantage of the outcome-oriented evaluation perspective lies in the fact that the comparability of different types of measures is made possible through the standardization of impact models and their intervention logics. This allows for the creation of measure packages as needed and the systematic preparation of funding offers, or potentially occurring overlapping effects can be identified and considered individually (e.g., in the “handover” of different measures in the individual steps of the impact model, such as the subsidized investment following a funded consultation) (ex-post-guidelines, p. 32).

The guidelines advise against using efficiency indicators for regulatory instruments.

The guidelines specify that the type of intervention logic associated with a measure should also be taken into account when interpreting the “funding efficiency” indicator (methodik-leitfaden-fuer-evaluationen-von-energieeffizienzmassnahmen, p. 62). Specifically, the guidelines state that measures with regulatory intervention logic may initially appear to have high funding efficiency compared to financially effective measures with economic or informational intervention logic. However, this perception is primarily because regulatory measures do not involve the provision of government financial aid. It is important to note that the calculation of the funding efficiency indicator often does not include other public and private costs associated with regulatory measures, such as administrative costs for control, or costs for personnel and investments required to comply with specific limits or minimum standards. Therefore, from an economic perspective, a regulatory measure may have significantly lower funding efficiency overall than initially indicated by the indicator. In the case of measures with regulatory intervention logic, it is recommended to consult with contractors to determine whether it is appropriate to report the funding efficiency indicator at all. This decision should be based on the ability to include relevant cost factors in its calculation. If the decision is made to report the indicator, it is necessary to provide a clear classification and a note highlighting the limited comparability of funding efficiency with financially effective measures.

Current evaluation framework is unsuitable for the evaluation of regulations. Costs incurred from regulations are macro-economic and administrative, which are underserved in the current evaluation framework. The current evaluation processes have primarily served as a procedure to demonstrate the cost-effectiveness of federal funding programmes (section 6.2). Since regulatory costs are not considered a direct cost to government, this area of policy has not incentivised the same level of attention as fiscal spending, is not subject to scrutiny by the BRH. This overlooks the considerable potential macroeconomic and welfare effects of introducing regulation and the administrative costs to effectively administer and credibly enforce these to ensure effectiveness. Reporting requirements under the KSG are relatively recent and the evaluation processes have not been adapted to accommodate necessary criteria for effective evaluation of regulatory measures.

5.2. Scope of evaluative indicators

This section focuses on the scope of ex-post evaluations using an extended set of criteria as a normative benchmark (Table 7).

Challenge	Components	Ex-post Methodological Guidelines	Published Evaluations
Effectiveness	<i>Energy use/savings</i>	Final energy savings (lifetime) <ul style="list-style-type: none"> • Building model (energetic balancing). 	Calculated in all accessible evaluations.
	<i>GHG abatement</i>	GHG abatement (lifetime) <ul style="list-style-type: none"> • Based on emission factor. • Static. 	Calculated in all accessible evaluations.
	<i>Stock composition</i>	Number of heating systems <ul style="list-style-type: none"> • Number of heating systems as a representation of the total stock. 	Not included in accessible evaluations.
	<i>Interaction effects</i>	Synergies between policy measures on a bundle level.	Synergies are assessed based on survey results, but they do not affect final savings.
	<i>Rebound</i>	Not included <ul style="list-style-type: none"> • Too difficult to calculate (p. 94) 	Not included in accessible evaluations.
Cost Effectiveness	<i>Dynamic cost effectiveness</i>	Investment effects <ul style="list-style-type: none"> • Based on expert assessment • Caused investments (p. 66) • Leverage effect (p. 67) • Jobs created/ saved (pp. 67-68) 	Included in all evaluations. None of the coded evaluations calculates them.
	<i>Static efficiency</i>	Marginal abatement cost (p. 65) Marginal energy saving cost (pp. 63-64)	Calculated in all accessible evaluations.
Fiscal burden	<i>Costs/revenues to state</i>	Aggregate fiscal budget costs	Calculated in all accessible evaluations.
Distribution	<i>Impacts on population</i>	Socio-economic impacts <ul style="list-style-type: none"> • Not included. 	Recent BEG evaluation includes “social aspects” which includes target groups by income, age and education.
		Job creation <ul style="list-style-type: none"> • Investment factor applied. • Expert assessment. 	Calculated in all accessible evaluations.
		Targeting/recipients of funding <ul style="list-style-type: none"> • Target group definition is a crucial element of system of objectives (pp. 41-42). 	<ul style="list-style-type: none"> • KfW 433 specifically targets “early adopters” • Other evaluations review group characteristics of recipients without explicit targeting. • BEG evaluation includes “social aspects” which includes target groups by income.
	<i>Impacts on firms</i>	Impact on heating costs <ul style="list-style-type: none"> • Reduction of energy costs (pp. 65-66) 	Calculated in 2023 evaluation of BEG.
		Job creation	Calculated in all accessible evaluations.
Acceptance	<i>Population</i>	Satisfaction with measures	Calculated in all evaluations except MAP.
	<i>Firms</i>	Satisfaction of recipients with administrative procedure	Calculated in all evaluations except MAP.

	<i>Political</i>	Not included	Not included.
Governance	<i>Administrative</i>	Administrative costs <ul style="list-style-type: none"> • Cost to Federal Agency for administration of BEG. 	Not included in accessible evaluations.
		Administration of measures <ul style="list-style-type: none"> • Satisfaction of recipients with administrative procedure. 	Calculated in all evaluations except MAP.
	<i>Information requirements</i>	Data quality	Published evaluations mention their data sources. Quality is not explicitly reviewed.

Table 7 - Scope of ex-post evaluations. Methodological guidelines are represented alongside the application in published evaluations to enable a comparison with their application in practice. The yellow colour indicates that the indicator is partially included, while red indicates that it is not included.

5.2.1. Effectiveness

The evaluations primarily assess the effectiveness of measures by establishing energy demand before and after implementation of measures. However, there are challenges in the evaluation process, such as the lack of standardisation in calculating energy demand and the absence of clarity regarding the certainty and uncertainty of these calculations (section 5.4).

The main indicators for effectiveness of measures implemented is energy savings and greenhouse gas (GHG) abatement (ex-post-guidelines, pp. 50-51). The evaluation process involves a combination of bottom-up evaluations of individual measures, which are then aggregated to determine programme-level savings.

GHG abatement

Primary energy savings are used to calculate GHG abatement. It appears that the primary energy factors are applied statically and are especially sensitive to the primary energy factor of electricity which must be updated on a regular basis (ex-post-guidelines, p. 18). Interviewees also indicated that in practice consultants can apply different values to these.

“we also take other carbon emission factors that we believe are more precise”. [Interview 11]

The calculation of GHG abatement is applied for the estimated lifetime of the measures installed, However, the primary energy factors applied should be coordinated with the client how to take into account changing CO₂ and primary energy factors. A current forecast of these factors should be used at the time of evaluation (ex-post-guidelines, p. 19 -

20). However, there is little guidance on how to implement these forecasts and can therefore be applied differently in practice.

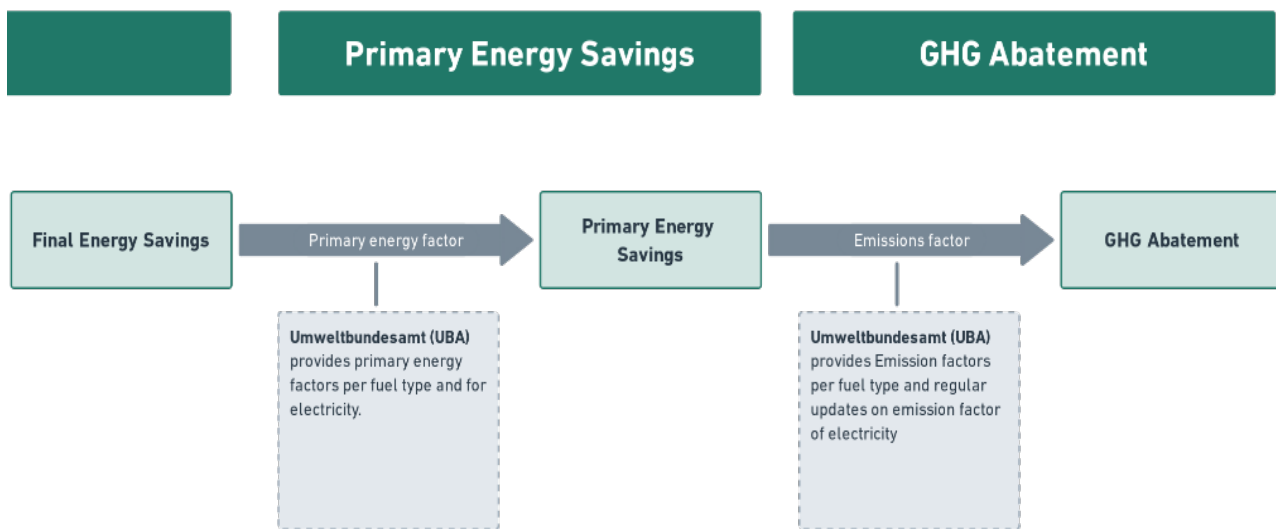


Figure 4 - Metric adjustments applied to final energy savings to determine GHG abatement in methodological guidelines.

Application of emission factors lacks transparency in published evaluations. In the reports analysed, is not transparently described how emission factors were applied for forecasting future energy saving effects. Based on our assessment, it appears that the emission factors were applied statically (i.e. an updated reference value at the time of evaluation), yet cost savings from installed measures are estimated across time. Energy savings are based on the expected life cycle of the measure, but the application static of PEF leads e.g. to an underestimation of GHG abatement from heat pump installation. Since the electric supply is expected to decarbonise over the next 15-20 years, the expected annual CO₂ abated would increase over its life cycle. Omitting this makes heat pumps appear potentially less efficient at GHG abatement. The converse is true for renovations. Renovations will deliver the most potential GHG abatement under current conditions, when combined with a gas boiler, or while the electricity mix is more carbon intensive (if coupled with a heat pump). This calculation should be calculated with potential confidence intervals since it is highly dependent on rate of decarbonisation of the supply of electricity. Work is being done to amend this in the ongoing BEG evaluation [Interview 7].

Rebound

The guidelines recommend against calculating rebound effects for ex-post or ex-ante evaluations. Rebound effects are counteractive impacts to an action. They include direct and indirect types. Direct rebound, such as energy-saving LED technology, leads to more consumption within the same system. Indirect rebound results in increased consumption in a different system due to savings in another, like more travel due to a fuel-efficient car. Separating rebound effects from other behavioural effects is considered challenging and subject to continuing academic debate. Thus, the guidelines advise against incorporating them primarily because of the complexity and lack of robust methods for quantifying these effects, thus making their measurement and inclusion often speculative and less reliable (ex-post guidelines, p. 94). In the ongoing BEG evaluation, it is being tested how the consumption comparison according to IWU, i.e. an empirically analysed influence of individual behaviour, can be included in evaluations. This has been included in the recently published report. While this is not directly the rebound, it is partially taken into account indirectly [Interview7].

5.2.2. Costs/Cost effectiveness

The cost efficiency is derived from the calculated energy savings and the fiscal costs of the programme. The cost-effectiveness of energy efficiency programmes is determined by considering the final energy savings achieved and the associated fiscal programme costs (ex-post-guidelines, pp. 60-61). This is a relatively static perspective on effectiveness. A more dynamic perspective on cost effectiveness could include projections for the future costs and demands of electricity and energy, compared against the energy saved per measure [Interview 7, 9]. This could also take into account cost trajectories for sustainable technologies, including potentially limiting factors such as supply chains, availability of materials, skills and capabilities to install measures.

The calculation of greenhouse gas (GHG) abatement relies on application of primary emission factors. The current methodology, which due to the application of energy factors, seems favourable to renovation support (see above – GHG abatement). As a result,

the marginal costs of abatement for higher-cost renovations may appear lower than they actually are, when considered dynamically over time. This caveat highlights a limitation of the current evaluation procedure, as it may support renovations that would not be considered cost-effective if future costs were taken into account. It emphasizes the need to assess and consider the long-term costs and benefits of energy efficiency measures to ensure a comprehensive evaluation of their cost-effectiveness.

I should point at one danger of evaluations as well. I mean, you tend to define indicators and one of the most prominent ones that policy always asks for is the greenhouse gas efficiency (tons saved per euro). I think, in very short period of time, it'll lead into a dead end, because when we decarbonize our electricity and our district heat and even the gas, etc., you will not save greenhouse gases in the future with more efficient devices, for instance, or buildings, or other energy saving measures, or other model shift measures. Because it's all decarbonized, so you save renewable energy. And I think there tends to be a very mechanical view. It was the same discussion we had in the evaluation of the market incentive program [...] So the evaluation doesn't really look into how much is the individual subject that is funded or supported or regulated really needed for the overall energy system. So you always need to be very aware of the limitations of an evaluation and of the indicators used, especially these coefficients, euro per ton, because you have uncertainty both in the nominator and the denominator. And so, and it's very unstable and leads to some wrong conclusions if you only follow these lines. [Interview 9].

Dynamic effects are underplayed, including potential market and systemic failures.

The current evaluation processes mostly overlook the dynamic aspects of cost-effectiveness, as well as the behaviours of market actors and the individual. Factors such as myopia, bounded rationality, or barriers to entry that may influence actor/consumer behaviour are underdeveloped. Although research has started to explore these aspects (e.g. IREES), integration into evaluation processes is limited. Some considerations of dynamic effects are considered though the definition of “effect adjustments” (section 5.4.2.), but these are commonly not applied in published evaluations due to lack of quality data. Understanding and incorporating these behavioural factors into evaluations can provide valuable insights into the effectiveness of policies and programmes [Interview 5].

Why don't people decide rationally? What obstacles are there? Of course, if you look at the individual models, they already have an infinite number of findings from the conglomerate of analyses. But I think there is still a lot of research to be done. [...] I've been more interested in the decision-making models of the individual actors lately [Interview 5]

Limited dynamic perspective has potentially favoured very high standards of renovation with steep marginal costs over targeting a larger volume of the housing stock.

The current procedural logics favour a single house focus for renovations. This incentivises high efficiency standards in a single housing unit [Interview 14]. Subsidy support for KfW efficiency standards, subsidise a higher proportion of the costs for higher efficiency standards (Table 8). The maximum subsidy for 'Efficiency house 40' is 24,000 euros, and up to 37,000 if coupled with a renewable energy installation. Heat pumps, by comparison, receive a maximum of 40% subsidy (BAFA 2023). Since the purchase costs are lower (10,000-30,000 euros) this equates to a maximum subsidy of 12,000 euros, which will reduce as the costs decrease through innovation and economies of scale. The marginal costs for the additional energy saved from renovation are relatively steep. An alternate logic would target a larger number of houses to an adequate efficiency level for effective functioning of a heat pump. This has been suggested to be around KfW 70 [Interview 14] but needs further research to substantiate a reliable benchmark. If accurate, the same fiscal budget could potentially renovate double the number of housing units, and facilitate more widespread acceleration of heat pump installation. When considered dynamically, the current approach may have lower cost-effectiveness than when considered statically, particularly as the supply of electricity is decarbonised.

KfW standard	Grant in % of total cost (per housing unit)	Total (euro per housing unit)
Efficiency house 40 + renewable energy	25 % of max. EUR 150,000	37,500
Efficiency house 40	20 % of max. EUR 120,000	24,000
Efficiency house 55 + renewable energy	20 % of max. EUR 150,000	30,000
Efficiency house 55	15 % of max. EUR 120,000	18,000
Efficiency house 70 + renewable energy	15 % of max. EUR 150,000	22,500
Efficiency house 70	10 % of max. EUR 120,000	12,000
Efficiency house 85 + renewable energy	10 % of max. EUR 150,000	15,000
Efficiency house 85	5 % of max. EUR 120,000	6,000

Table 8 - Funding available for KfW standards. Source: KfW (2023).

A more dynamic perspective on cost effectiveness needs to combine anticipated cost trajectories for the decarbonisation of electricity generation with demand reduction (energy savings). To evaluate if the energy savings are cost effective for higher standards of renovation, requires integrating with projections for the costs of decarbonisation of electricity generation per kWh, and anticipated cost trajectories for the roll-out of heat pumps. Heat pumps have been demonstrated to be the most cost-effective method to reduce German gas consumption (Altermatt et al., 2023), and meeting decarbonisation targets necessitates the decarbonisation of electricity generation. Modelling indicates that a balanced approach between demand reduction and decarbonisation of supply is needed [Interview 9]. This requires a more integrated strategic approach which combines sector coupling and energy system demand estimation. The lack of a comprehensive treatment of interaction effects in the evaluation of programmes (section 5.4.3.), would suggest a potential lack of foresight and highlights an area where more methodological attention is required.

Macro-economic indicators

Reduction of Energy Costs. This indicator evaluates the goal of reducing energy costs and is part of the goal achievement (gross) and impact analysis (net including adjustment) (ex-post-guidelines, pp. 65-66). Energy cost savings are calculated based on the computed energy savings over the entire effective period of the measure. The reduction

of energy costs is considered in the sample of coded ex-post evaluations in the recently published evaluation of the BEG.

Triggered Investments and Leverage effect. For an economic consideration, value-added effects through the investments triggered by the measure play a significant role. The leverage effect is a ratio between the used resources and the triggered investments (ex-post-guidelines, pp. 65-66). This indicator does not take into account administrative costs. The leverage effect represents how much investment is triggered per euro used. In practice, both indirectly and directly triggered investment are calculated. In a comparative context, the KfW 433 evaluation primarily emphasises the leverage effect of the funding, determining the extent to which it encourages additional private investments (KfW-433, pp. 101-102). In contrast, the KfW-EBS-WG-2018-2021 (pp. 9-10) and MAP 2019 (pp. 27-28) evaluations prioritise input-output models to calculate the direct and indirect value-added effects, respectively. This involves assigning the investments to different sectors based on their nature and analysing the corresponding increase in demand and production within those sectors. In the evaluations of national monitoring reports and the high-efficiency pumps evaluation, these effects are not considered.

We always try to do that a little bit, but of course there's a limited time, so you cannot follow that a hundred percent through. On the other hand, these macroeconomic impacts and these kind of things are typically investigated in separate projects. So they're not, they're not really part of the formal evaluation. But for instance, we had in 2010 a project on macroeconomic impacts of different efficiency measures, and looking at all the different effects on jobs and of energy security, etc. So this is typically done in separate activities, one does not really expect that from the actual evaluation of the policy instrument.

[Interview 9]

Inclusion of more advanced methods could be explicitly included in the tendering specification. Modelling requires resources from the evaluator, and if the client does not explicitly ask this question ("What are the gross value-added effects associated with the funding?"), it will not be included. Otherwise, additional costs incurred would make the process more expensive - which would possibly have negative effects in the competitive

process of securing the work, and could be awarded to a competitor. From the evaluators' point of view, it therefore makes economic sense to offer only the scope of services specified in the terms of reference for the tender [Interview 7].

Job creation is considered, and is included in the assessment methodology as a factor to scale. How this is applied actually contradicts socio-economic considerations, since the indices favour higher cost measures which have a higher job creation index. This therefore favours higher cost renovations which are currently supported by the KfW, rather than lower cost renovations which are more likely to support lower income households.³ This may lead to a bias in the cost effectiveness of measures, which is potentially. A second consideration is that Germany is currently facing a labour and skills shortage to deliver the measures, which is currently considered a bottleneck towards achieving more rapid decarbonisation in the sector. Given this skills shortage, emphasising job creation as a positive metric, without supportive measures to support upskilling and training for tradespersons, may lead unrealistic implementation strategies being devised ex ante.

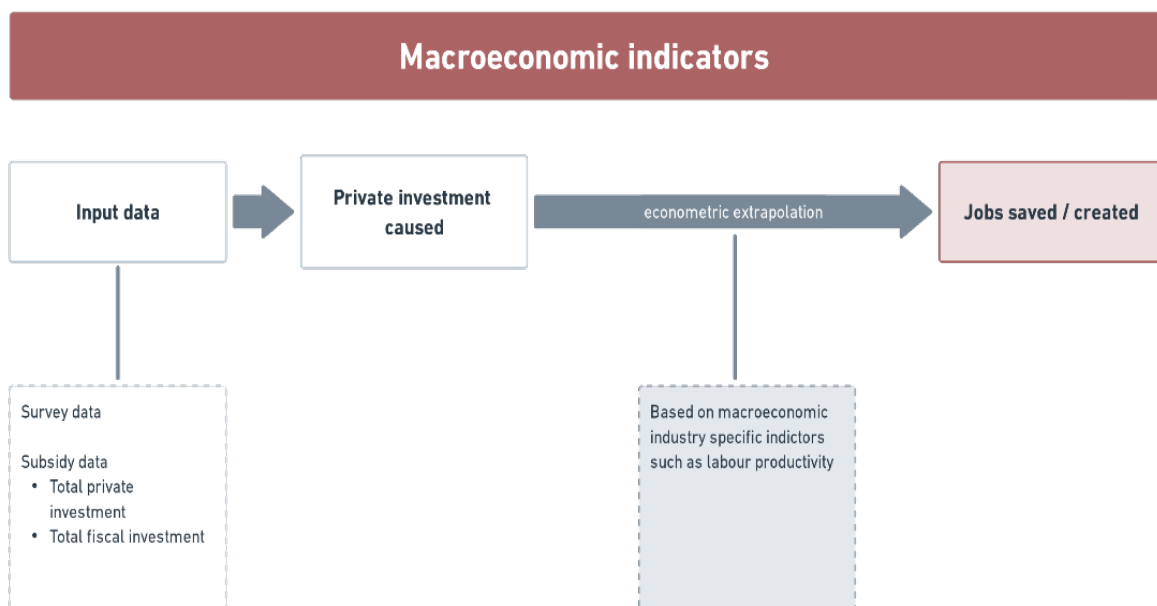


Figure 5 - Calculation for private investment caused and jobs created.

³ Current heat pump subsidy rates go up to 40%, whereas small single measures like exchanging windows only cover 15% of the costs (20% when done with an energy consultant).

5.2.3. Distributional impacts

In the published methodological guidelines, there is very little attention to socio-economic impacts. The distributional impacts of policy options are currently mostly absent from the currently published evaluations and assessments [Interview 6, 8, 11, 12]. The most recent BEG evaluations have included “social aspects” which includes some information on the recipients of funding, including net monthly salary, age and education level. How this information is collected and analysed is currently unclear. The recent addition is encouraging since targeting and income groups which have benefited from fiscal support represents a major gap in the evaluation of the federal spending programmes. Historically, information has not been recorded by the federal agencies which collect the data, and therefore this evaluative dimension has been excluded [Interview 6]. The recent BEG evaluation indicates that the largest beneficiaries are the highest income groups. There was no uptake of the funding in the lowest income group (up to 1000 euros net monthly salary) and only 8% uptake in the 1000-2000 euros group.

The indirect effects of this methodological bias likely favour regressive outcomes.

There is no direct subsidy targeted at socio-economic groups who spend a large proportion of their income on heating. This has major implications for policy design. Due to poor current thermal insulation, not all of the current housing stock is suitable for the installation of heat pumps without further energy efficiency investment. Without support, this means that those who cannot afford to meet the renovation requirements could be excluded from receiving heat pump support. This may result in the heat pump subsidy support being unintentionally regressively targeted [Interview 6], even though it is currently designed as neutral mechanism. Secondly, carbon pricing has much more impact on these worst performing buildings, and can therefore be more regressive in its impacts than for affluent households who have installed high cost renovations. Third, without targeted support for low-income groups to increase the efficiency of their homes, the increasing regulatory standards in the GEG will impose costs on those least able to pay, incurring further regressive outcomes.

This reflects a more general approach in German bureaucratic and policy communities in which climate policy and social policy have been treated separately. Climate policy

has focussed primarily on reducing energy demand and GHG abatement, and achieving these goals cost-effectively. Social policy has been considered a separate issue, and only recently have there been more efforts to integrate these policy areas [Interview 8].

The EU level has recently increased requirements to focus on social aspects of energy policies across sectors. This emphasis from the EU has led to a more explicit consideration of socio-economic impacts (energy poverty etc.), and this indicator is required to be included in the next round of reporting. Other countries, including France and the UK have paid more attention to energy poverty and social impacts of policy design, including targeting and schemes aiming at most vulnerable groups. This is likely driven by higher levels of income inequality and lower quality of the building stock, which have meant that the social impacts of energy policies have been more salient for a longer period. Germany, like Sweden, had not focused on this area but with rising energy costs and the implementation of the BEHG these considerations have recently become much more visible [Interview 8]. Moreover, introduction of the BEHG and planned implementation of minimum efficiency standards will incur economy wide effects have distributive implications for socio-economic groups. Accordingly, updating the evaluation procedures to comprehensively include these dimensions is essential.

We suggest further consideration of socio-economic impacts in the next round of reporting and explicit targeting of support to those most adversely impacted. The EU has recently mandated the inclusion of reporting requirements for socio-economic aspects to be added to assessments [Interview 8]. The most recent BEG evaluations include some targeting aspects such as the net income of recipients, but this should be extended to estimate reduction of energy costs in proportion to income groups. Recent reports indicate that the current targeting of BEG subsidies generally favour more affluent groups. These programmes should be reconsidered to increase progressivity. Currently, there has not been an update to the existing methodological guidelines with an expanded scope to include distributive impacts. Consequently, as yet, there is limited evidence of how distribution will be calculated. It appears to present a significant challenge to expand the existing evaluation methodology, due to the aforementioned limitations in assessment procedures, and data availabilities (further discussed in section 5.3).

5.2.4. Acceptance

Acceptance is considered only narrowly as the satisfaction of the recipients of subsidies with the installed measures. Through survey the recipients of the subsidy are questioned about the satisfaction with the installed measures. While this provides some information on one dimension of acceptance, this only includes a sampled population group of adopters. This does not generate any applicable information on the acceptance of the general population.

Population acceptance is not an explicit evaluative dimension of either ex-post evaluations or ex-post assessments. Acceptance is considered in the guidance in terms of soft recommendations, but not explicitly in evaluative dimensions. For example, guidelines make recommendations for using craftspeople/as advocates of measures, due to citizens having more trust in tradespersons than in government.

Acceptance is excluded from the ex-ante methodology. Whilst difficult to evaluate some consideration of acceptance in the ex-ante assessment methodology could be included. Doing so could help identify targeted measures which might be needed to promote measures, i.e. targeted information campaigns.

Acceptance is not easy to evaluate due to lack of counterfactual group. Ex-post evaluation of a sample of adopters does not represent the wider population of those who are unlikely to accept or implement measures. Even in broader academic research, acceptance is commonly estimated based on respondents to national survey questions and polling (Levi, 2021). These types of research are often based on a single policy instrument type, but do not consider instrument stringencies, interactions with other instruments, or practical issues such as the inconvenience and disruption of having measures installed. Further research in this area is a critical gap in current knowledge and better understanding is needed to enable faster adoption and more progress on climate policy.

Acceptance in this sector is not always linked to policy instruments or programmes, but also to the complementary measures needed to enable more effective governance. As previously discussed, data protection and the utilisation of smart metering and real

time measurement of energy usage are key reforms which could greatly increase the reliability of evaluations of policy in this sector. However, acceptance issues related to lack of trust in government and use of data have commonly been barrier to more widespread adoption of smart metering in households (Bugden and Stedman, 2019).

5.2.5. Governance

The methodological guidelines for evaluations currently do not encompass wider aspects of governance requirements, such as capacities, enforcement, and compliance rates. Consequently, these aspects are excluded from the scope of the evaluations. However, considering the potential significance of governance deficits on the effectiveness of instruments and programmes, it would be beneficial to explicitly include them in the evaluation process. Independent reports that specifically investigate these issues, e.g. with regards to effectiveness of enforcement, can provide valuable insights. Therefore, incorporating governance deficits and their potential implications into the evaluation framework would enhance the understanding of the overall effectiveness and efficiency of instruments and measures. It is recommended to adopt this practice to ensure a more comprehensive evaluation of programmes.

Governance requirements are partly considered in terms of the administrative costs incurred on federal agencies for BEG measures in ex-post evaluations. The guidelines outline that administrative costs to the Federal Agencies tasked with delivery of the BEG are to be recorded and evaluated. Notably, the description of administrative costs does not include administrative or enforcement of regulatory measures for the Federal States (Länder) or the Landkreis.

In our sample of published evaluations administrative costs were not calculated in the reports. While the guidelines for evaluations include the consideration of administrative costs, in practice, these costs are often excluded from ex-post evaluations due to the challenges associated with accurately calculating them. While administrative costs play a significant role in the overall cost-effectiveness assessment of energy efficiency pro-

grammes, their exclusion from ex-post evaluations can limit the comprehensive understanding of the programme's efficiency. It is important to recognize this limitation and explore ways to improve the methodology for capturing and incorporating administrative costs in the evaluation process to ensure a more accurate assessment.

Recipients of subsidies are required to give feedback on their satisfaction with the administrative process. This provides some information on the efficiency of Federal Agencies on their administrative processes, and the installer. The information produced is however, subjective from the perspective of the recipient. Data may be subject to a positive bias since the responses are higher from those who completed the process, as opposed to those who did not. As a counterfactual, those who applied but did not complete the installation of measures "Storno-Befragung" are also surveyed (KfW-433 p.5-7), but in some instances the response rate is low. If the n is too low, extrapolation is not always possible and is therefore excluded from the evaluations.

5.3. Data

The accuracy of the evaluation depends on the reliability of input data. Data availability largely influences the reliability of the evaluations. Reliable input data related to the building stock, the current energy carriers used, and the efficiency of the measures implemented are vital to ensure that energy savings and GHG abatement are calculated correctly (section 5.4). This section reviews key issues related to data: quality, reporting, protection, enforcement, and establishment of a database.

5.3.1. Data quality

Subjectivity of application of data quality standards in the assessment process can introduce inconsistency and inaccuracies in the results. While guidelines outline different categories of data quality, there is a lack of specific guidance on how to assess the data quality for a concrete evaluation. This ambiguity leaves room for subjective judgments and decisions regarding the quality of the data, which may ultimately lead to inconsistent and inaccurate outcomes. The evaluation process combines the use of technical

data and survey responses, but several important criteria lack technical data, potentially affecting the comprehensiveness and reliability of the evaluation.

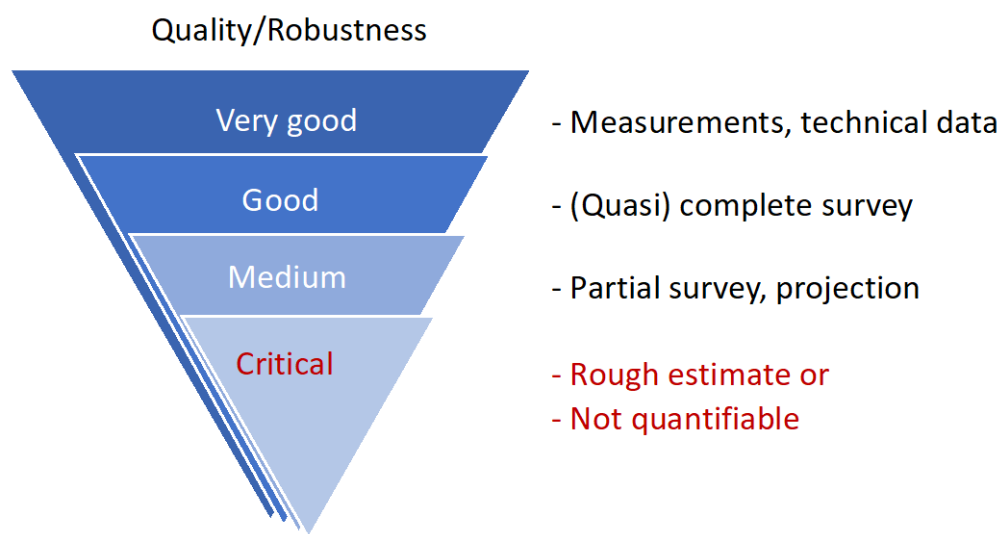


Figure 6 - Visualisation of data quality. Adapted from Schlomann et al. (2020).

Another potential source of bias stems from the sampled group involved in the evaluation. Self-reporting introduces the possibility of subjective interpretations and biases in the collected data. Additionally, a low response rate from non-completed applicants can further impact the representativeness and reliability of the data.

Data access and issues pertaining to quality of record keeping are not explicitly discussed in the evaluation guidelines. While quality is assessed for data in terms of availability, there is not an explicit focus on issues relating to accessibility, and the quality of data which is available. We now discuss some of these issues more explicitly.

5.3.2. Data reporting quality and access in Federal Agencies

Data on the recipient's energy use is collected by the federal agencies for applications for support mechanisms, but not anonymously. Due to the absence of anonymization, this data falls under the purview of data protection laws and is not publicly accessible. As a result, there are limitations on the possibility of conducting accuracy and validity testing and limited administrative accountability. With little validity testing or auditing, the quality of the recorded data can be compromised. For example, the recorded data can be

low quality, with instances of buildings being recorded with implausible sizes as small as 1m² [Interview 7]. This concern has been raised by bureaucrats [Interview 3, 4] and consultants [Interview 6, 7, 10, 12]. This adds a significant challenge to conducting evaluations, since consultancies have to check low quality data for errors. How these errors are corrected, may not be transparent or standardised which may limit replicability (section 5.4).

There is usually high quality of data about the financial aspects, and overwhelming amount on what is done. If you look at the measures in detail, there are lots of details not available. So, most of the time you don't know by refurbishing what was state before, what will be the state after. There are things we don't get from the data. Sometimes this kind of data is in a file but it's not collected and stored as electronic data. [Interview 7].

Data is provided to commissioned consultants in anonymised form. When data is received evaluators conduct an initial data check for missing/implausible data. These are corrected using different methods. What is done and how it is done is made transparent at least to the employer this procedure is common with evaluators [Interview 7]. However, this is not transparent or publicly accessible, which limits replication. In addition, further surveys are often conducted with funding recipients, e.g. to obtain baselines or to be able to model building parks in more detail [Interview 7].

Lack of digitisation of records is a further barrier for increasing the timing of evaluation processes. The collected data is not digitized and is stored in physical files, requiring consultants to request access to view the information. This process can lead to delays and hinder the timely analysis and evaluation of the data *“funding data that you only get after four months and 25 emails”* [Interview 9]. This was a commonly reported issue among consultants [Interview 7, 10, 12].

Recording of metrics and cost structures varies across agencies. Compared to BAFA, KfW's cost structures are generally higher. KfW employs calculation models that are better equipped to handle comprehensive refurbishment projects, which tend to be more complex compared to individual measures covered by BAFA. BAFA's programmes are de-

signed to cater to a broader market and are relatively more straightforward. These differences in approach and complexity are not indicative of preferences but rather reflect the specific focus and scope of each institution's funding activities [Interview 4].

Enabling verification and validation the recorded data can impact the reliability and credibility of the evaluation outcomes. Addressing these issues would involve finding solutions that balance data protection requirements with the need for transparency, accuracy, and validity in evaluation processes. This may include exploring anonymization methods to make data accessible for testing and analysis purposes without compromising privacy. Additionally, digitizing the data and implementing user-friendly platforms for data access could streamline the evaluation process and improve efficiency.

5.3.3. Data protection and the enforcement of regulatory instruments

The evaluation of regulations in achieving energy efficiency goals is hindered by several key challenges related to data availability, enforcement, and accountability. First, the lack of data on energy use before the implementation of energy efficiency standards, as mentioned in the previous paragraph, makes it difficult to assess the baseline energy consumption and establish the impact of the regulations. Moreover, there is a lack of data on the effects of regulations after their implementation, mainly due to the absence of reporting requirements. Without comprehensive data on energy consumption patterns and performance indicators, it becomes challenging to accurately estimate the effectiveness of regulatory measures.

Your core question is about evaluation, and evaluation presupposes data if you don't just want to do it in a bubble. I think that's simply the crux of the matter. We don't have any-, we invest, to put it brutally, but I think I already said in the last round with you: we invest billions and don't really know the actual state and thus of course not the effect. So even if I were to find out afterwards that consumption would now be such and such. I don't even know where I started from. I have to have an actual [control] and then a target and then a fulfilment. [Interview 4].

One of the most significant issues undermining the effectiveness of regulations is the lack of enforcement. Non-compliance with regulations in the building sector is well-known, with estimates suggesting a non-compliance rate as high as 25% or possibly even higher (Lu et al., 2022). This collective set of challenges makes it extremely difficult to assess the true impact of regulations on energy efficiency. Even if reforms are made to improve data provision and access, the effectiveness of these measures relies heavily on a robust inspectorate. However, the current limitations in data provision and sharing, including difficulties in the federal government accessing regional data, contribute to a lack of accountability. Consequently, this undermines the credibility and effectiveness of the inspectorate regime responsible for enforcing regulations.

There's no building inspection in Germany. So the law is, if I have a building and I want to repair ten percent or more of any feature, such as the outside wall, I have to energy efficiently improve that to new build standard, that's the law. But nobody checks up, nobody cares. [Interview 14].

More funding is needed for a robust inspectorate at the Länder level along with better coordination with Federal Government. To enhance enforcement, addressing these challenges requires increased funding, skills, and training at the regional level (Länder), as well as better vertical coordination between different levels of government. However, accomplishing these tasks is not a straightforward endeavour, which may explain the existing shortcomings. Nonetheless, recognizing and addressing these administrative issues, reforming data provision and access, increasing accountability, and conducting evaluations to assess the performance of regulations can have a significant impact on the decarbonization progress of the building sector. By improving data availability, enforcement mechanisms, and accountability structures, policymakers can foster more effective and impactful regulations that contribute to the desired energy efficiency outcomes.

Verification and control of regulation policies is very difficult. The Länder, the states have to do that and not the federal state. There is very little, surprisingly little data. We are doing the project on developing the next generation of the billing code right now, last year and this year, it's a huge project, and there's so little data to build on. We can see the

overall trends, but we cannot see neither regionalized data nor [compliance rates]. So how do you evaluate regulation? That's much more difficult. [Interview 9]

5.3.4. Establishing a Building database

The availability of data poses a significant barrier to better quality evaluation. The limited availability of data on the current building stock poses a significant barrier to accurately evaluating the energy saved from energy efficiency measures. This deficit was recently acknowledged in a Expertenrat für Klimafragen (ERK) report:

The availability of publicly accessible data in the buildings sector is limited. In Germany, there are official statistics as well as scientific and economic statistics such as association figures or funding statistics that provide information on individual characteristics of the building stock. However, changes in the building stock with regard to the structural characteristics of the heating system and the energy quality of the building envelope are not regularly surveyed officially. Therefore, there are data gaps that make continuous observation of structural change difficult. Thus, important indicators or data on energy refurbishment rates and the energy status of the building stock or energy efficiency labels cannot be measured sufficiently. (ERK 2022; p.24)

Germany does not have a buildings database, unlike most EU countries. With the exception of Germany and Romania, all other countries maintain an energy certificate database and utilize the data from it to make assessments and statements regarding their building stock (*BfEE-data-quality-building-sector*, p. 6). These databases serve as valuable resources for gathering information on the energy performance of buildings and enable policymakers, researchers, and stakeholders to analyse and evaluate the energy efficiency of the built environment. Germany's current lack of a centralized energy certificate database in place, potentially limits the ability to make comprehensive statements about the building stock's energy efficiency.

To facilitate better evaluation and analysis, the establishment of a comprehensive database is crucial. Establishment of a database is a priority for enabling better evaluation

procedures [Interview 4, 6, 7, 10, 11, 12, 14]. This database should encompass information on the building stock, including energy performance indicators, efficiency ratings, and other relevant characteristics. Making this database publicly available allows for transparency and facilitates research, policy development, and monitoring of energy efficiency efforts. In line with our findings, the Expertenrat für Klimafragen (ERK) report also made the recommendation of a buildings database (ERK 2022; p.24)

The establishment of a nationwide building and housing register, which has been discussed on various occasions, would help considerably to obtain complete transparency on the refurbishment status and technical equipment of the building sector and thus contribute significantly to an effective monitoring of the causes relevant to the development of emissions. For example, a building and housing register has already existed in Austria since 2004. In addition, there is an energy performance certificate database in Austria. By means of this data, besides questions of housing policy and questions of local spatial planning, it is also possible to develop more targeted measures to achieve climate goals. The findings from such a database can, for example, have a direct impact on subsidy stocks and levels (Statistik Austria 2022).

Introduction and enforcement of minimum efficiency standards would significantly increase the availability and access to data for the existing building stock. Germany will be obligated to introduce minimum efficiency standards as part of new EU commission reform the EPBD included in the Fit-for-55 proposals. Introduction of minimum efficiency standards will require significant changes to be made to the enforcement and sharing of data for regulations in Germany. Effective enforcement mechanisms are needed to ensure compliance with energy efficiency standards and prevent evasion of regulations. Additionally, improving data sharing between Länder and the Federal government is vital for a comprehensive understanding of the energy performance of buildings across the country. Over time, as minimum efficiency standards become more standardized, these can serve as a benchmark for assessing energy savings.

From the EPBD, the European Buildings Directive, we need to introduce a building database. That's very helpful because that would be an automatic, and statistically a wonderful starting point for evaluations, because then you can really see in real time what is happening. It will take a number of years before we really get it. [Interview 9]

Mandatory inspection of current buildings and performance standards could be introduced and reported. One potential solution to address this data gap is to conduct inspections of the housing stock. This would involve systematically assessing the energy performance of buildings to gather relevant data. Similar inspections are currently mandated for gas boilers, so training and skills of the existing inspectorate regime could be extended to assess the whole property. To ensure compliance with data protection regulations, the collected data could be anonymized to avoid any conflicts with privacy concerns.

Under current German data protection laws, the use of energy certificates is not permitted to construct a database. This restriction poses a limitation on the feasibility of methods that rely on energy certification databases. Due to concerns regarding the protection of personal information, individual-level data cannot be accessed or used for analysis. As a result, data can only be stored and utilized in the form of random samples, which hampers comprehensive assessments and analyses of energy efficiency in the building sector.

Alternatively, better data quality could be enabled through real time monitoring, and improved accessibility of data from energy providers. Moving towards more accurate and real time monitoring of building standards and energy usage would help improve this gap in the current input data [Interview 3, 6, 7, 11, 12]. Accelerating the installation of smart metering in housing would help provide more accurate data on household energy usage. Public acceptance remains a key consideration, but recent polling has indicated 41% of Berlin citizens are willing to install smart meters⁴. Reforming data protection laws so that energy usage can be provided anonymously from energy providers could also improve the provision of data for evaluation and research purposes.

⁴ <https://www.bitkom.org/Presse/Presseinformation/Interesse-an-Smart-Metern-steigt>

There are two current options for constructing a buildings database, without reforming data protection laws. Given the limitations on energy certificates for database construction: (i) Remote sensing data analysis; and (ii) Multi-level sampling data analysis. Remote sensing data analysis utilizes exterior features of buildings. It draws on various data sources, including satellite imagery, aerial surveys, cadastral maps, local government data, and open data resources. This approach offers the advantage of leveraging multiple data sources to generate comprehensive and detailed information about buildings. It does not however have detailed information on the internal composition of buildings.

Multi-level sampling data analysis by the IWU (Institute Housing and Environment) has been used to develop a German building typology. They ran previous projects (ENOB:dataNWG) that entailed conducting extensive surveys of the residential building stock (BfEE-data-quality-building-sector, p. 39). Currently, they are engaged in a new project focused on surveying the light residential building stock, where renovation activities are being recorded independently of the funding sources. This provides valuable insights into the dynamics of building renovations and their timing. However, it should be noted that this type of research is both time-consuming and unfortunately not consistently integrated into ongoing practices. The data gathered through these projects is valuable as it offers a deeper understanding of the temporal patterns and trends associated with building renovations. By capturing information on renovation activities independently of funding sources, it provides a better view of the broader landscape and sheds light on the motivations and drivers behind renovation decisions. Conducting such surveys and gathering this wealth of information requires a significant investment of time and resources.

The current frequency of these studies is too long to provide accurate data on a short timeframe. The process of data collection, analysis, and interpretation is not a straightforward task with the previous project spanning five years. Moreover, there is a challenge in ensuring their continuity over time. It is essential to anchor and institutionalize these efforts to enable the consistent monitoring of building renovation activities [Interview 6].

This would allow for a more comprehensive understanding of trends, shifts, and the impact of policy interventions over an extended period. This has not happened yet, but is always project-related over a few years, and then there is a break of a few years. For them, the interval between surveys is almost ten years. Given the urgency of climate protection in the building sector, that is arguably not frequent enough [Interview 6].

Proposals for the Heat Planning and Decarbonisation of Heat Networks Act could introduce improvements to data quality and access. During the redrafting of the Heat Planning and Decarbonisation of Heat Networks Act proposals were made to increase data quality. These included the collection of data on electricity consumption over a period of three years, as well as information regarding the types of heating systems in use. These data collection measures aimed to gather important information related to energy consumption patterns and the types of heating technologies employed. However, the draft law released to municipalities on 21st July indicates the data requirements may have been relaxed (BDEW, 2023). At the time of writing, it is unclear how the Act will be finalised, but the original proposals to increase data collection would both increase the quality of evaluation processes and help compliance with EU requirements for establishment of a database under the EPBD.

5.4. Methods

Having discussed the scope and data of the evaluation processes, we now focus on methods. As previously outlined, the main focus of the evaluation processes is on the environmental effectiveness of instruments and programmes (energy savings and GHG abatement), and their cost effectiveness. Given the prominence of these criteria we focus on the methodologies for calculating these two indicators in more detail. In particular, we pay attention to the transparency, reliability and replicability of these methods. We then focus on instrument interaction effects and how they are accounted for in the current evaluation practices. The section concludes with some recommendations to improve the procedures in these aspects.

5.4.1. Final and Primary Energy Savings

Final energy savings are calculated as the difference between the estimated energy demand before and after the implementation of the measure. To assess the impact of energy-saving measures, the starting point is an estimation of the total energy demand, which takes into account factors such as building characteristics, occupancy, and energy-consuming devices (ex-post-guidelines, pp. 56-57). This estimation provides a baseline against which the effectiveness of measures can be measured. Subsequently, the evaluation calculates the energy demand after the implementation of specific measures, considering the energy carrier mix, efficiency improvements, and changes in user behaviour. By comparing the initial estimation of energy demand with the final energy demand, the impact of the implemented measures on energy savings and GHG abatement can be determined.

Energetic balancing

The current method of calculating the total energy demand is energetic balancing and application of a building model. However, this process is currently opaque, as evidenced by a lack of clarity in the methods applied during coded ex-post evaluations (KfW-EBS-WG-2018-2021, p. 9, 40; KfW-433, pp. 97-98; high-efficiency-pumps, pp. 15-16). A potential concern among interviewees is that these adjustments are not standardised. While there is a simplified equation developed by IWU, it is left to the discretion of the evaluator how to apply this from calculation per m² to a relatively large building space (i.e. 150 m²) [Interview 9, 12, 14]. Consequently, this lack of methodological clarity potentially overestimates the actual energy usage of a building type, with worst-performing buildings seeing an overestimation as high as 40-50% [Interview 12, 14], a finding corroborated by past research on the pre-bound effect (Rosenow and Galvin, 2013; Sunikka-Blank and Galvin, 2012). Recent work on building refurbishment has dealt with this issue well (Jagnow and Wolff, 2020), and could be incorporated into revisions of the evaluation methodology.

Limited standardisation creates complication in the comparability of results across studies. Due to the lack of standardisation, both metrics have to be recorded for comparability across studies [Interview 9]. To interpret these results requires transparent of the

adjustment factors applied by different consultancies. While this may be possible for technically knowledgeable consultancies it reduced the transparency of these results for policy makers or for publicly accessible publications.

What we are doing now is we apply a building model to calculate the savings per building. And then we have normalization factors, the so-called "Bedarf-Verbrauch Faktoren", which were derived from, for all different categories of buildings. So you have factors to be multiplied so that you can convert the calculated demand savings to real demand consumption savings. So we calibrate that, and typically we report both. If you use the evaluation data to see whether the measures that are implemented in Germany to achieve our climate goals are sufficient, it doesn't help to use calculated data, you need to see what is really happening, and so you would use the second data. Whereas if you want to compare it to other studies it might be helpful to also have the calculated data, if people apply other normalization or correction factors [Interview 9].

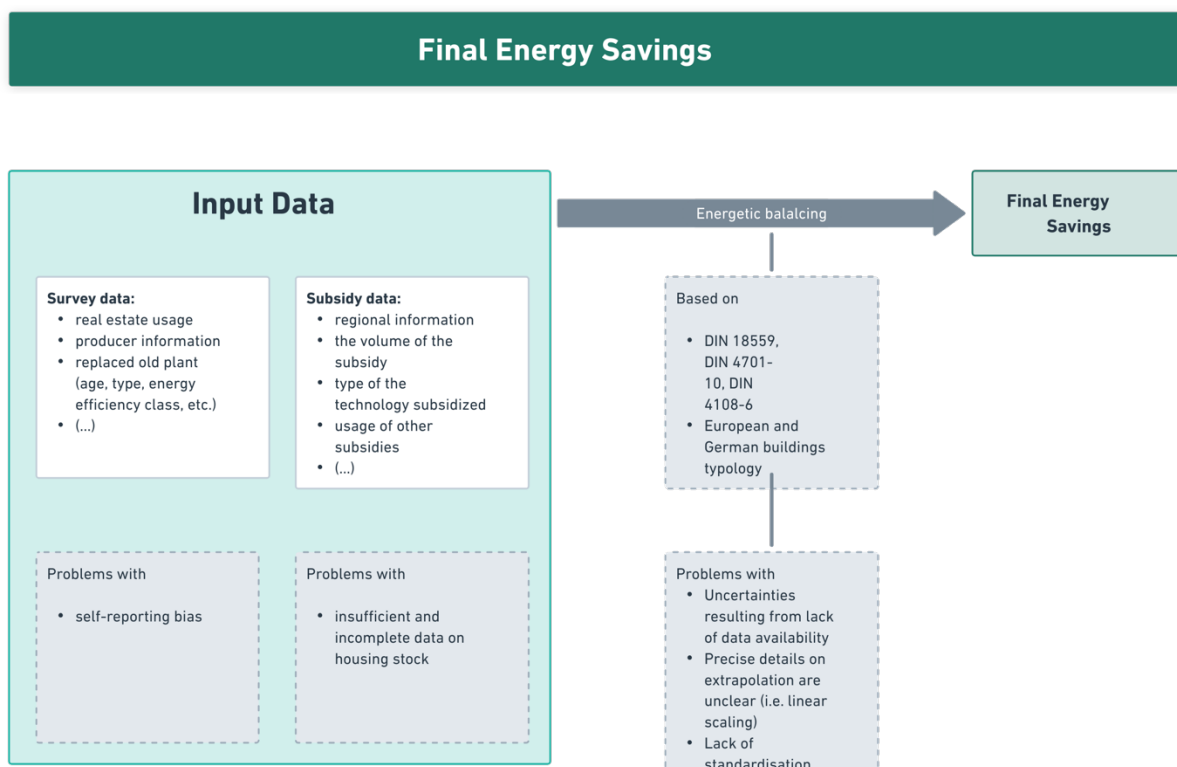


Figure 7 - Calculation for final energy savings.

Existing evaluations do not provide a clear insight into the certainty or uncertainty surrounding the final energy demand calculations. They lack details such as confidence intervals or sensitivity analysis. Instead, they present mostly estimated numbers and references to DIN norms, but fail to provide in-depth information on the reliability or variability of these estimates (KfW-EBS-WG-2018-2021, pp. 8-9). This gap in information hinders a comprehensive understanding of the energy demand evaluations' accuracy and robustness, thus demanding increased transparency in their implementation and reporting.

The application of these evaluation methods shows variability across different consultants. Due to issues with the quality of recorded data consultants are required to apply methods for calculation and adjustments, but how these are applied can vary across consultants [Interview 9, 11, 12]. This discrepancy underscores a need for increased transparency and standardisation, particularly because these calculations significantly impact policy and measure outcomes by affecting the calculation of final energy savings.

In general, there is a guideline how to calculate this energy and carbon savings. And we use this guideline but it is not/ it does not perfectly fit to our project. [...] We have the data about each case. So, as always, the data is not perfect. So, we have to look at the data and if there are always missing numbers and obviously wrong numbers and we try to sort this out. And then we calculate, and we use different methods for the calculation. [...] We do not only take the calculated values, but we try to estimate the real efficiency. But it is usually a little bit different from the standardized calculation [Interview 11].

Final energy savings are transformed into primary energy savings using the primary energy factor. The primary energy factor is a coefficient used to quantify the total amount of primary energy sources (such as coal, natural gas, or renewable energy sources) required to produce a unit of final or usable energy (ex-post-guidelines, pp. 18-19). Final energy savings are converted into primary energy savings through the application of this primary energy factor. The calculation inherits its methodology from the existing energy demand determination process. As such, any inaccuracies within the calculation of energy savings - whether arising from the application of energetic balancing or

the validity of input data - will directly impact the resultant primary energy savings and subsequently the GHG abatement.

5.4.2. Effects adjustment

The guidelines outline effect adjustments to correct from gross to nett effects. This includes the consideration of anytime effects, spill-over effects and pull forward effects (ex-post-guidelines, p. 86). The methodological application of the adjustment effects is illustrated in Figure 8.

Anytime cost effects aim to capture the additionality of the measure. This adjustment aims to assess whether the measures would have been implemented anyway without the subsidy. This concerns what economists regard as additionality and relates to the cost effectiveness of the fiscal subsidy.

Not enough data on the general population for a counterfactual group. Very little is known about groups who do not take fiscal support or who take up measures without support [Interview 6, 5, 7, 12]. This makes it very challenging to build up a counterfactual group to estimate the additionality of the fiscal spending programmes.

The whole area of what happens in buildings that don't apply for funding, etc., that's just something where we don't know anything, where permanent, continuous monitoring would actually be super helpful. [Interview 4].

These effects are estimated based on survey data. All of the adjustment effects included are based on survey Responses from those who took part in the scheme (KfW-EBS-WG-2018-2021, p. 60; KfW-433, p. 84). For anytime effects this is problematic (at least for general extrapolation) since the *sample* does not reflect the wider population. Recipients of subsidies for higher cost renovations are much more likely to have already been already considering the measures. Anytime costs have been estimated as high as 50% from surveyed participations (KfW-EBS-WG-2018-2021, p. 63).

Effect Adjustment

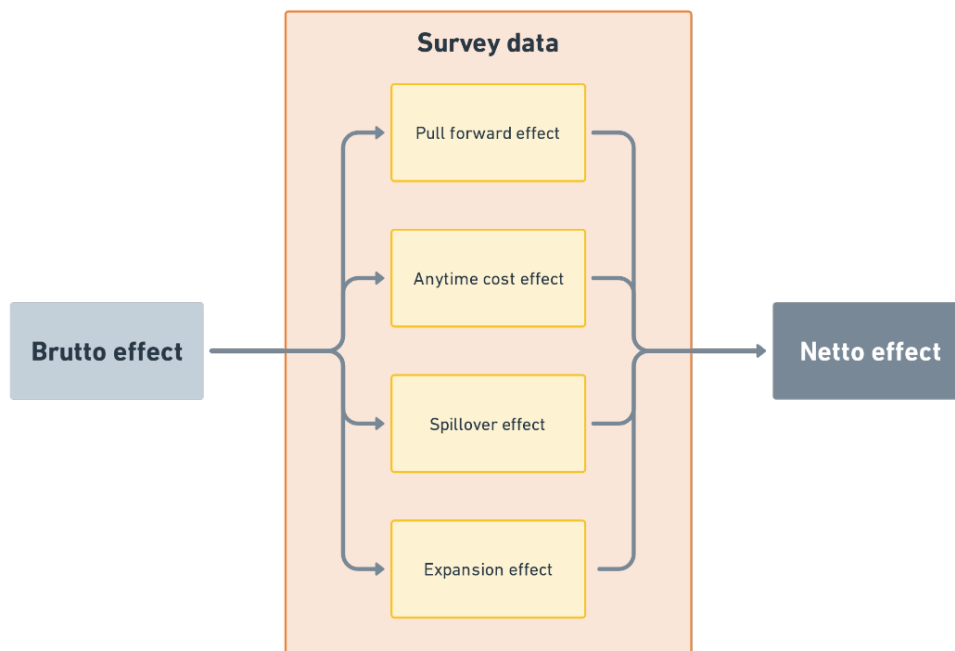


Figure 8 - Effect adjustments to calculate Net effects from Gross.

In practice (in published evaluations) the gross effect is used, not the net adjustment.

The calculations which are related to the use of gross/net effects are written in the guidelines in such a way that either value can be used (Figure 9). Depending on the group considered, the margin of error can be up to 20%, for the total group up to 5% [Interview 7]. This means that considerable uncertainties are to be expected when determining the net effect. Often, therefore, the clients do not determine the net effect. In the sample of published evaluations analysed, in all cases the gross effects were used.

In an evaluation of the KfW 433:

As previously outlined, the effects can be determined. However, from the perspective of the evaluation team, it does not seem useful to use the effects for determining a net effect. Therefore, they will be described and discussed in the following but will not be included in the impact determination. (KfW-433, p. 84)

In the EBS WG ex-post evaluation- Section 5.1.1 – Approach to effect adjustment:

Against this background, effect determination using surveys represents an approximation whose meaningfulness should not be overinterpreted, especially with regard to the differentiated allocation to individual programs/actor groups. Therefore, both gross and net values are always given in the evaluation report. (evaluation-kfw-foerderprogramme-ebw-2018-2021, pp. 61 - 62)

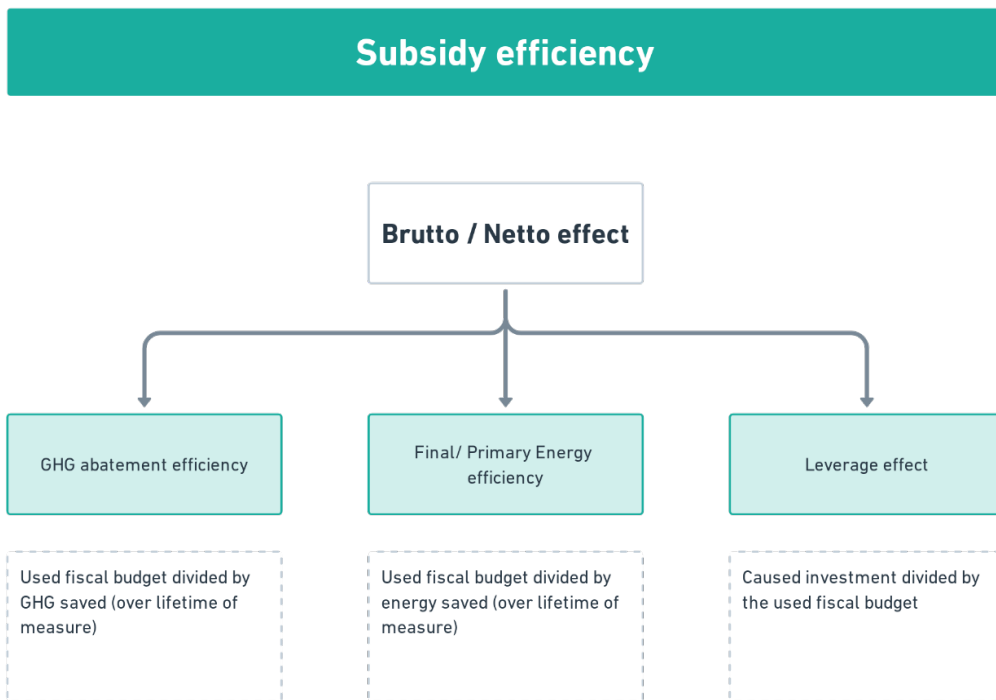


Figure 9 - Use of gross/net effects in methodological guidelines.

5.4.3. Interaction effects

Methodological guidelines highlight interaction effects at the bundle level, rather than individual measures. Both the ex-ante and ex-post guidelines acknowledge the importance of considering interaction effects at the bundle level rather than at the individual measure level (ex-post-guidelines, pp. 99-100; ex-ante-guidelines, pp. 11-13).

For ex-post evaluations the main motivation for considering interactions is to avoid double counting of energy saving effectiveness. In the context of ex-post evaluations, it is emphasized that interactions among energy efficiency measures, especially within measure packages, should be examined to avoid double counting of effects (ex-post-

guidelines, p. 99). The guidelines propose different approaches to address these interactions, including direct descriptive evaluation, analytical evaluation, or expanding the circle of companies. Instrument factors and interaction matrices can be utilized to transparently describe the interactions between different combinations of instruments.

Ex-ante guidelines also stress the need to account for instrument interactions. Similarly, in ex-ante evaluations, the guidelines highlight the significance of accounting for interaction effects within measure packages (ex-ante-guidelines, pp. 11-13). Various methods are suggested for adjusting these interactions, with or without integrated modelling. These methods may involve using instrument factors, interaction matrices, or integrated models. Integrated modelling frameworks implicitly capture interactions between individual measures and enable the redistribution of the savings effect of a measure package back to individual measures using techniques like ranking, linear scaling, or scaling with instrument factors or an instrument matrix.

In practice, published ex-post evaluations did not follow the guidelines when accounting for interaction effects. However, it should be noted that the two ex-post evaluations conducted on funding programmes for energy-efficient buildings in Germany, specifically KfW 433 and KfW EBS-WG, did consider interaction effects on the bundle level. However, neither evaluation explicitly applied the recommended guidelines to account for these interactions. While the KfW 433 evaluation surveyed the utilization of other funding programmes in addition to KfW 433, it did not employ the recommended approaches to address interactions (KfW-433, pp. 108-109). Similarly, the KfW EBS-WG evaluation examined awareness and usage of other funding programmes but did not utilize the recommended approaches for adjusting interactions (KfW-EBS-WG-2018-2021, p. 76).

From our sample of evaluations conducted after the publication of the guidelines, we suggest that further standardisation is needed for interaction effects. While our sample is small due to limited publicly available data, the guidelines were not explicitly applied in either of these evaluations to account for interaction effects. This suggests that further methodological standardisation is needed.

5.4.4. General recommendations

Transparency and reliability are key considerations in evaluation methods. Currently, evaluations primarily focus on static effects and should also incorporate the lifetime effects of measures to provide a more comprehensive assessment. Replicability is crucial for ensuring the credibility and robustness of evaluation results. Further efforts are needed to enhance standardization in evaluation practices, allowing for consistent replication of studies. Sensitivity analysis is recommended, particularly when there are uncertainties in parameters. By varying these parameters, sensitivity analysis helps to understand their potential impact on evaluation outcomes and provides more reliable data for decision-making processes.

Adoption of data disclosure documentation for commissioned evaluators would significantly improve transparency and replicability. It is currently not transparent how the data is corrected to account for missing values, or data which is subjectively considered “obviously wrong”. To improve these issues, supporting data disclosure documents should be implemented and answered by the consultant alongside the policy evaluation. This data is often available as a non-public document (must sometimes also be kept available by the audit department due to internal company requirements) [Interview 7]. However, this should be made publicly accessible. Without widespread accessibility of such a data statement/data documentation, replication is currently not possible.

To better assess acceptance, adoption rates, and additionality, alternative methods are needed to construct a counterfactual group of the wider population. While the Federal Agencies who administer subsidies may collect data from recipients through self-reporting requirements, there are no such obligations for buildings that do not receive subsidies. This creates a self-reporting bias, where the collected sample is not representative of the wider population.

An evaluation is usually biased. We just know about the ones who ask for grant. We don't know about the ones who do similar things without granting or without funding. [Interview 7]

Conducting national surveys can help construct a more representative control group.

Prognos are currently conducting a survey on the effectiveness of the BEG⁵. These are still dependent on response rates and may have a bias towards population groups who are more environmentally conscious or more likely to adopt measures as they are already aware of the support mechanisms. Increased participation could be enlisted through including survey questions into broader national surveys, or incentivising participation through promotional offers or discounts in partnership with energy service providers. Other options for control group design and counterfactual analysis is also recognized in the methodological guidelines as an option but is often not applied due to concerns about cost and data protection. Independent research, such as the Wohnen-Wärme Panel (FrondeI et al. 2021), has started to help fill this gap.

⁵ <https://www.prognos.com/de/FAQ-BEG>

6. Use of evaluations and effects on the policy process

This section outlines the main uses of evaluation evidence in the policy process in Germany. The sections relate to four main themes: (i) analytical inputs in ex-ante forecasting; (ii) motivation, commissioning of evaluations; (iii) timing of evaluations and informal advice; and (iv) dissemination of evaluation results and transparency.

6.1. Analytical inputs into ex-ante forecasting

Ex-post evaluations are needed to inform ex-ante assessment. Ex-post evaluations are crucial aspect of governance since they provide the evidence if policy measures are working effectively and forms the basis of the ex-ante assessments. Reliable and comprehensive ex-post evaluations are needed, at least for the most important programmes, to make good ex-ante evaluations [Interview 8].

If you have a building subsidy program for example, then we take the ex-post evaluation and use the main relevant indicators for this impact model also for the ex-ante evaluations. [Interview 8]

The quality and reliability of these evaluations improves the assumptions for ex-ante impact models

Assumptions are funded by empirical knowledge, and usually we base on the empirical figures from the ex-post evaluations for ex-ante evaluations. [Interview 8].

By extension, inaccuracies in evaluating instrument performance parameters ex-post may translate into inaccuracies in projecting future GHG reduction effects of planned policy mixes. Given the increased prominence of the projections report (Projektionsbericht) in triggering reform of the German climate policy mix as envisioned by the new draft of the Federal Climate Laws, this may have important implications for the steering of German climate policy.

There are of course enormous studies, also very large-scale ones, for example by Dena, which then also include all possible scenarios for different energy sources and of course

also across sectors, everything is very highly complex. But the building stock data is regularly a weak point. [Interview 4]

There is often a lack of transparency regarding how ex-ante assessments work is compiled. The assumptions in modelling exercises is less transparent or traceable than for some of the ex-post evaluations. This lack of transparency raises questions about the extent to which scientific input and evidence are considered in the target formation and agenda-setting processes. The limited scientific input in these stages may hinder the effectiveness and credibility of policy decisions.

Unlike the projection report, where we are somehow involved as an intermediary and coordinate the process, this quantification report always runs without us. So that means that the findings are then used by those who also do the ex-ante modelling and we also come to the fact that the findings are usually a little bit limited. We are not so directly involved though [...] So, when it comes to the detailed curves, I don't know how detailed it is planned. There are of course those per model who do it on our behalf [...] experts who do it. [Interview 5]

Ex-ante guidelines do not differentiate between different instrument types. As a complementary document, the ex-ante guidelines do not extend to the differentiation between types of measures and only refer to the ex-post evaluations.

The ex-ante assessments do not consider administrative costs, and thus exclude the governance criteria completely. Beyond administrative costs, there is no consideration of required capacities, of potential governance failures (such as enforcement), and required reforms (such as conflicts with existing laws) needed to effectively implement. This overlooks important evaluative dimensions such as bottlenecks in the availability of craftsmen to install renovation measures, and the availability of consultants, inspectors and evaluators. Again, notably is the lack of consideration of a robust local inspectorate regime for regulatory measures, which are a known issue in the sector and which undermines the effectiveness of the estimates for impact through regulatory measures.

Addressing these issues would require a shift towards a more comprehensive and inclusive evaluation approach, incorporating a broader range of criteria beyond fiscal

justifications. This would involve greater transparency in the process of target formation, improved scientific input, and enhanced stakeholder engagement to foster more effective and impactful energy and climate policies.

It's difficult to answer. When you look at the landscape of instruments and programs in the building sector, they began the last year's fundamental changes. So, most of the new instruments and programs aren't evaluated properly till now. So, most inputs for the Sofortprogram was done by a mix of results of older evaluations and parts of ongoing evaluations mixed together and just modelled or analysed by tools and to have a view of possibly impacts. So, it's very reduced. One key feature is how to get the so-called further efforts. For example, when you fund one euro GHG emissions, greenhouse emissions or energy is saved. This information we try to put together by different evaluations and get along on the long run-on average. [Interview 7]

6.2. Motivation, commissioning, and use of evaluations for policy recalibration

The motivation behind commissioning evaluations plays a crucial role in their utilization. Historically, evaluations have primarily been used to justify fiscal spending on support programmes. As a result, the main criteria considered in these evaluations are often focused on assessing energy savings and cost-effectiveness. The compliance of ministries with reporting requirements appears to be the main use of evaluations. Meeting reporting obligations is crucial for the allocation of funds and ensuring accountability in the implementation of energy and climate policies.

Interviews suggest commissioning process for evaluations limits the involvement of the scientific community and leads to confirmation bias. The practice of commissioning evaluations can lead to established relationships with key actors involved in the evaluation process. This can result in a closed circle of input and limited opportunities for external stakeholders or “outsiders” to contribute their perspectives or expertise [Interview 14]. As a consequence, valuable research conducted by external parties may be disregarded, potentially leading to the exclusion of important insights and alternative viewpoints. Moreover, due to data protection, the data collected by Federal Agencies is only

made available to the commissioned evaluators. It is not publicly available or accessible to the scientific community, limiting opportunities for independent analysis and validation [Interview 7, 12, 14]. This lack of data sharing creates a barrier to transparency and hinders the potential for collaborative research and cross-validation of findings.

Several actors within the consortium of main consultants conduct both ex-post evaluations and ex-ante assessments, which potentially limits external validity checking.

There is a potential for confirmation bias since the same actors conduct the ex-ante assessments. If an independent group were involved, they might question the quality, robustness, and scope of the ex-post and ex-ante evaluations. The enhanced role of the Projektionsbericht envisaged by the KSG Novelle and the role of the ERK in reviewing it may enhance transparency and feedback. Importantly, potential conflicts of interests of individuals must be avoided. Nevertheless, there has been progress in evaluation procedures, and standardization has increased recently. In 2020, the consortium of main consultants (section 4.3) published methodological guidelines originally meant for a specific project [Interview 9]. These guidelines have now been adopted by the BMWK as the standard procedure, and all evaluations must adhere to them.

Data access is only granted to commissioned evaluators, which limits the scope for independent and academic research.

There is a notable limitation in the accessibility of ongoing work or formative evaluations to the public [Interview 6, 7], despite the potential significance of these reports. This lack of transparency hinders the ability of stakeholders, researchers, and the public to monitor the progress of evaluations and engage in informed discussions about the findings. In order to address this issue, it is essential to prioritize transparency in the evaluation process. Promoting transparency involves making evaluation data available to the scientific community, enabling collaboration and facilitating rigorous scrutiny of the findings. By allowing external experts and stakeholders to have access to the evaluation process, diverse perspectives can be incorporated, enhancing the overall quality and credibility of the evaluations. Transparency in the evaluation process is crucial for the development of evidence-based policies. It enables informed decision-making and fosters public trust in the evaluation outcomes. By ensuring that evaluation reports, ongoing work, and relevant data are accessible to a

wide range of stakeholders, the evaluation process becomes more accountable and effective in driving positive change.

The absence of external input and independent analysis can hinder the validity and scope of the ex-post evaluations. Without external scrutiny, there may be limited opportunities for new ideas, improvement of operational practices, or critique of existing approaches [Interview 12]. The lack of external input and validity testing can restrict the ability to question the accuracy and reliability of the evaluation results, potentially limiting the effectiveness and credibility of the evaluation process [Interview 14].

The specification provided by ministries for evaluation processes can introduce certain biases that impact the outcomes of evaluations. The primary motivation for conducting the evaluation is to demonstrate the effectiveness of programmes and to justify the use of Federal spending. The commissioned evaluations can create a confirmation bias in their specification, meaning the results are not likely to yield radical recommendations or reveal systemic failures in the programmes [Interview 12]. Instead, they tend to focus on small technical suggestions, limiting the scope for transformative changes. Such changes relate to the limited scope of the evaluated criteria (section 6.2.) and the types of programmes which are evaluated (section 6.3.).

The role of evaluations seems to have relatively low influence on policy decisions during the previous government. Based on our interviews, when evaluations indicated issues with programmes (for example highlighting that fiscal spending should not focus on new buildings, or on very high standards of renovations - KfW 45) recommendations did not lead to reforms [Interview 6]. The accumulation of policy feedback that the programme was not functioning well, was insufficient to motivate policy reform without a change government leadership. The relative low salience and use of previous evaluations could help explain the relatively limited scope they consider. Increased accountability through the implementation of the KSG may lead to increased use of evaluations in the future.

Many of the things that were actually consensual in science and also in the institutes that work on the topics were simply not politically enforceable in this way in previous federal governments. [Interview 6]

The impact of ex-post evaluations on decision making is constrained to incremental changes to the current programme. The types of evaluations commissioned are not likely to produce more radical recommendations. They can inform whether the policy is working or not, and are used to justify the costs of the programme. As detailed in section 6.2, the reliability of these evaluations is highly dependent on the adjustments for energetic balancing that are applied, and the assumptions for the original energy use of the building prior to the measures. Currently there is not enough transparency in how these are applied external validation testing suggests that these are over estimated, which has implications for both the effectiveness of the current measures and the cost-effectiveness. Potential over-estimates in the effectiveness of current programmes limits the prospects for evaluations to motivate more radical reform.

While evaluations and assessments do play a role, they tend to have a more significant impact on smaller technical changes rather than larger systemic reforms. In the realm of significant policy changes, political actors often rely more on politically salient topics or intuitions rather than comprehensive evaluations. In some cases, they may selectively seek out evaluations that align with their preconceived ideas or desired outcomes.

So all the adjustments that have now been made to the BEG, also pretty much ad hoc and at short notice, what was the decisive factor, I don't know, but ultimately in the scientific debate, the indications have been coming for a very, very long time, to focus the funding on the building areas where it is really needed [...] whether that was driven by these evaluations and studies or by individual people in the ministries, I can't judge, but in the end some things were implemented that have been criticised again and again in the evaluations and studies for many, many years. [Interview 6]

6.3. Timing of evaluations and informal advice

Timing is a critical factor in the dissemination of evaluation evidence. One prominent issue is the lag in publication, which can result in delays between conducting evaluations and its availability. This delay can hinder the timely utilization of evaluations for decision-

making and policy formulation. The lag between research completion and its integration into policy discussions and decision-making processes.

Large changes in the programmes and the required speed of the transformation does not allow time for evaluations to be conducted. This is most problematic for new measures, which do not have a body of existing evidence to draw on. This means that consultants are asked to generate advise in very short timeframes which does not allow robust evaluative work [Interview 9].

Right now the speed is so fast, of changes and additions of the programs, that sometimes, I mean like a couple of weeks ago, I mean we are involved in meetings where before we have done one little piece of evaluation, we should recommend how the programs should be changed. So I mean, but that's really the situation right now, the pace and the speed of required transformation [Interview 9]

Delays in the production of evidence leads to information provision and consultation being conducted in a formalised or codified manner. The role of experts' advice based on prior experience can apparently become influential in shaping policy options. Established relationships with these sources of advice often take precedence, limiting the scope for considering new ideas and alternative options. The reliance on informal communication and the influence of consultants can undermine transparency in policy formulation. Furthermore, the lack of publicly available data hampers the possibility of independent evaluation that is not directly commissioned by the ministries. This limitation creates a selection bias in the types of actors who are able to inform ministries.

What's actually better, in my opinion, is to talk to the people who do these evaluations because they can, of course, observe the same dynamics and the market and the world, and they know also what's going on and they have a really good idea. And additionally, they know the data that is, of course, then not extremely recent. But they can bring those two together also in a more recent timeframe. And this is then nothing what's written in the reports, or maybe in some of these additional reports that we're asking them to write, but not in these ex-post evaluations. [Interview 3]

Transparency could be increased by participatory consultation processes, which invite evidence from multiple sources of expertise. Enhancing transparency by making relevant data publicly available would facilitate independent evaluations and broaden the range of information considered in policy formulation. Encouraging diversity in expertise and fostering an inclusive decision-making process can also help avoid undue influence and promote a more robust and innovative approach to policy development.

6.4. Dissemination of evaluations and transparency

More explicit specification for the outputs and reporting of evaluations would increase potential impact in decision making. The lack of specificity in the composition of evaluation teams hinders the usefulness of reporting. Internal coordination among consultancy team members responsible for conducting evaluations is often inadequate, and there is often a disconnect with the policy team [Interview 12]. This lack of coordination results in technical recommendations that may be challenging to translate into actionable policies or measures.

To address these issues, it is essential to incorporate diversity in evaluation teams.

This diversity should not only encompass disciplinary expertise but also include individuals with different perspectives and backgrounds. By involving a more diverse range of experts in the evaluation process, the communication of findings can be improved, ensuring that technical recommendations are more accessible and actionable for policymakers. Introducing diversity requirements in the tendering process for evaluation teams can help address the limitations in reporting and enhance the quality and usefulness of evaluations. This would enable a broader range of insights and expertise to be considered, potentially leading to more comprehensive and impactful recommendations.

Some stakeholders have expressed the need for greater inclusion and consultation during the target and strategy formulation stages. Some consultants have expressed the concern that targets have been established without the input from a broad range of expertise [Interview 7, 11]. Similarly, insufficient coordination across different ordinance has led to the establishment of separate objectives and goals across programmes, which

has led to different reporting requirement metrics and makes transparency and dissemination more difficult [Interview 7]. This call for more involvement and input from a diverse range of stakeholders indicates a desire for a more inclusive and participatory approach in shaping energy and climate targets. This could help address consistency across the goals in different targets and strategies. Such inclusion can help ensure that targets are sufficiently ambitious, while being scientifically grounded.

7. Key reform options

Key reform options are now outlined in regard to each section, and summarised in respective tables.

7.1. Institutional arrangements and coordination

Enhancing coordination mechanisms between federal agencies and ministries would improve accountability and data accessibility. Implementing stronger accountability measures to ensure the quality of recorded data is advised. This necessitates increased coordination and collaboration among ministries and agencies. The establishment of a common code of conduct and practices for agencies can promote accountability by setting standardized guidelines to be followed. Key aspects such as digitization of recorded data should be prioritized within this framework. Making the data publicly available and subject to external review and validation, either through government channels or external capacities, further enhances transparency and accountability. By adopting these measures, the government can foster a culture of accountability, improve data accessibility, and enable more effective decision-making processes.

Government can encourage greater diversity in consultants and evaluator groups through increased collaboration with the wider scientific community. To enhance transparency and external scrutiny, the base of expertise within consultants and evaluators could be broadened. This can be achieved by increasing the transparency of data and assumptions, and making data publicly available to enable external expertise and scientific research, thereby opening up the policy subsystem. By providing access to data, transparency is increased, enabling accountability and allowing external experts and civil society to scrutinize the findings. Involving external expertise brings diverse perspectives and independent assessments, strengthening evaluations. This could be particularly valuable for topic areas which are only recently being incorporated into evaluation practices, such as distributional impacts. Encouraging scientific research fosters evidence-based decision making and collaboration with academia, also facilitating formal evaluation reports that might be able to draw on newly developed methodologies. This

could form part of the tendering requirements, while requiring transparency of data and assumptions ensures reliability and allows for external validation.

Coordination between the Federal Government and Länder. Enhancing coordination between the Federal Government and Länder involves several key reforms that can contribute to more effective collaboration and implementation of policies. Firstly, it is recommended to foster increased accountability of the Länder. This can be achieved through establishing clear mechanisms for reporting and monitoring progress, setting performance targets, and ensuring transparency in the decision-making processes of the Länder. By holding the Länder accountable for their actions and outcomes, the coordination between the Federal Government and Länder can be strengthened. Secondly, promoting the sharing of data between the Federal Government and Länder is essential for effective coordination. This includes establishing mechanisms and platforms for exchanging information, data, and best practices. Sharing relevant data and insights can facilitate evidence-based decision-making, enhance policy coherence, and enable better coordination of efforts between the Federal Government and Länder. Lastly, increasing funding for the inspectorate at the Länder, with support from the Federal Government, is crucial to ensure effective oversight and enforcement of policies. Adequate funding can strengthen the capacity of the inspectorate to carry out monitoring, compliance checks, and enforcement activities. This, in turn, helps maintain the integrity of policies and regulations, enhances accountability, and contributes to the successful implementation of joint initiatives between the Federal Government and Länder.

Current issues with institutional configuration and potential reform options		
Issue	Impact	Potential solutions
Limited coordination and accountability of record keeping in Federal Ministries	Variable data quality recorded makes evaluations more difficult and requires evaluators to correct errors and make adjustments.	<ul style="list-style-type: none"> Formal auditing of Federal Agencies in terms of operational procedures from Ministries responsible for line-management. Internal reporting requirements.
Lack of vertical coordination and data sharing between Federal level and local authorities creates accountability gap for Länder on GEG enforcement.	Undermined credibility of enforcement for GEG reduces effectiveness.	<ul style="list-style-type: none"> Data protection and sharing issues between the Federal Government and Länder should be resolved.
Enforcement/compliance is delegated to the Landkreis, which are typically lacking funding and expertise to fulfil this role.	Landkreis are unable to effectively fulfil their role and an inspectorate and enforcement body.	<ul style="list-style-type: none"> A new Federal public body dedicated to regulatory enforcement. This agency should standardize enforcement and inspectorate procedure. Funding should be dedicated to establishing assessment methodologies. Regional subdivisions should be established to fulfil this role, directly line managed by the centrally coordinating inspectorate agency. Funding and training for local inspectorate should be increased.
Relatively closed group and stable consortium of consultants.	Limits innovation in evaluation methods and practices.	<ul style="list-style-type: none"> Broaden group of consultants. Promote collaboration with other sources of expertise in the commissioning specifications.

Table 9 - Key reform options for improving institutional configuration.

7.2. Scope of programmes evaluated and evaluation criteria

The GEG is currently not evaluated, which presents a significant gap in the governance of the residential building sector. The omission of regulatory instruments in evaluation and assessment processes may be attributed to regulatory standards being administered at the regional (Länder) level, along with data protection laws that restrict the sharing of information between local authorities and federal ministries, as previously mentioned. As a result, there is a lack of aggregation and inclusion of regional data in evaluations and assessments. Resolving these issues should be a priority, even if coordination between local and federal governments remains challenging. One potential solution could be granting consultancies access to local records and enabling them to aggregate the data for a comprehensive evaluation. The recent introduction of efficiency standards may contribute to improving the reporting of regulatory compliance. It is

worth noting that evaluation and assessment methodologies might be currently undergoing updates following the implementation of efficiency rating standards in Germany in 2021, although updated methodological guidelines are not yet publicly available.

Expanding the scope of current assessments in energy efficiency policies involves addressing distributive impacts and governance challenges. It is crucial to pay attention to the socio-economic impacts of these policies to ensure a fair distribution of benefits and avoid exacerbating inequalities. The most recent evaluations are a step in the right direction, by including targeting of the BEG subsidies. Currently, social policy programmes pay for the utility bills of residents in state subsidised housing, for disadvantaged citizens who receive state support and are unable to work. The costs spent on these social programmes are not integrated with the fiscal support spent on subsidy programmes to promote energy efficiency. Crucially, evaluating the impacts of the GEG on income groups should also be prioritised, and where possible, conducted as ex ante assessments, thus alleviating potential adverse outcomes. Better integration of social policy and climate policy should also lead to better consideration of cost effectiveness.

Governance challenges play a significant role in the operationalization of policies but are currently not considered in evaluations. This requires specific capacities in administration, including better coordination across funding agencies. Improved coordination can enhance the efficiency and effectiveness of policy implementation by streamlining processes and avoiding duplication of efforts. Addressing governance challenges requires a multi-faceted approach that includes strengthening administrative coordination, enhancing enforcement capacities, and implementing robust monitoring mechanisms. By doing so, energy efficiency policies can be effectively operationalized, ensuring equitable distribution of benefits, and driving meaningful progress towards sustainable and efficient energy systems.

Current issues with scope and potential reform options		
Issue	Impact	Potential solutions
GEG is not currently evaluated.	<ul style="list-style-type: none"> • NAPE reporting and any ex-ante assessments of building sector regulatory impacts are not based on ex-post evidence. • Revision of the GEG is not based on reliable ex-post evidence. 	<ul style="list-style-type: none"> • Introduce reporting requirements. • Standardized evaluation requirements and procedures for all local authorities. • Make data recorded at the Länder accessible to a coordinating evaluation consultancy.
Tax incentives and BEHG not evaluated at high resolution.	<ul style="list-style-type: none"> • Not possible to accurately attribute changes in energy use and GHG abatement to changes in energy prices. 	<ul style="list-style-type: none"> • Higher resolution data collection on energy use. • Improved buildings database.
Scope of indicators is limited	<ul style="list-style-type: none"> • Important information is excluded from current evaluations, which limits evidence to base decisions. 	<ul style="list-style-type: none"> • Evaluations should include more indicators, in particular: distributive impacts, governance, dynamic cost effectiveness.
Evaluations do not explicitly include governance gaps/challenges or specific challenges on data access and quality.	<ul style="list-style-type: none"> • Key issues are not explicitly made known to government about necessary reforms to improve governance. 	<ul style="list-style-type: none"> • Evaluation should be extended to identifying governance challenges, including capacity/coordination/delivery gaps and specific issues with access and quality of data.

Table 10 - Key reform options for improving scope of programmes evaluated and evaluation metrics.

7.3. Data access and quality

Improving access to and the quality of data represents a critical challenge for enhancing the reliability of evaluations. Addressing this challenge requires reforms in data protection laws to enable greater availability and sharing of data. While sensitive data must still be protected, basic information such as the number of houses should be made more readily accessible. Additionally, the digitization of data is essential for achieving greater transparency and accountability, particularly for federal funding agencies. By digitizing data, it becomes easier to scrutinize and fact-check information, leading to higher quality data and more robust evaluations.

To facilitate access to relevant data, the production of a publicly accessible database of the building stock is necessary. This database should include information such as the types of dwellings, the heating systems in place, and U-values or efficiency ratings. By making this data accessible to researchers, policymakers, and the public, it promotes transparency, enables evidence-based decision-making, and facilitates the monitoring

and evaluation of energy efficiency measures. By reforming data protection laws, digitizing data, and establishing a publicly accessible database of building stock information, the evaluation process can benefit from improved access to reliable and comprehensive data. This, in turn, strengthens the reliability of evaluations and enhances the effectiveness of energy efficiency policies and programmes.

<u>Current issues with data access/quality and potential reform options</u>		
Issue	Impact	Potential solutions
Federal Ministries' recorded data is limited in scope, excluding important factors such as targeting (i.e., distributional effects).	Limits what can be robustly evaluated without being survey based or approximated.	<ul style="list-style-type: none"> • Important information should be included, such as income group. • This information should be included as part of the BEG application process which would also improve targeting to reduce additionality of subsidies.
Recording keeping in Federal Ministries stores data with applicant names.	Data falls under Data Protection Laws attributed to personal information, which restricts access and sharing of data.	<ul style="list-style-type: none"> • Data should be recorded anonymously. Some identifying characteristics, such as demographic information should be recorded to help extrapolation. • Alternatively, this data should be anonymized or pseudonymized in post-processing.
Record keeping in Federal Agencies is not digitized.	Restricts access and limits possibility for quality checks on data.	<ul style="list-style-type: none"> • Digitization across Federal Agencies should be prioritized to increase access, allow quality checking of recorded data, and auditing.
Lack of buildings database	Limits ability to conduct more accurate evaluations by providing reliable data on energy use before measures.	<ul style="list-style-type: none"> • Change data protection laws to enable use of certificates. • Increased resolution of multi-level sampling data analysis.

Table 11 - Key reform options for improving data access and quality.

7.4. Methods

Monitor the actual post-retrofit consumption and base evaluations on that. Monitoring is essential for assessing the effectiveness of energy efficiency policies. It should involve real-time monitoring to provide immediate feedback and allow for timely adjustments. Additionally, monitoring over a longer period, typically 3-5 years, provides insights into the long-term impacts of policies and helps identify trends and areas for improvement. Using data on pre-retrofit consumption only when it is actual measured consumption over 3 years increases accuracy, but presents a challenge for current policy implementation in the lack of an existing database to draw from.

Standardization and transparency of adjustments and assumptions in the evaluation process need to be further improved to enhance accuracy and comparability. Currently, estimation of gross and net effects is prone to overestimation, and different consultancies follow varying procedures. Aligning and refining estimation methods can yield more precise assessments. To achieve standardization, better coordination among consultancies is necessary, beyond the existing guidelines' limited scope. Ministries can play a central coordinating role by ensuring consistency in evaluation procedures, methodologies, and extending commissioning terms to enhance transparency. However, ministries may require capacity-building efforts to enhance their technical expertise in evaluation methodologies. Training, resources, and guidance can empower ministries to effectively contribute to standardization. Promoting further standardization, facilitating consultancy coordination, and strengthening ministries' technical capacities will result in robust evaluations, accurate estimations of effects, and informed decision-making for effective policy implementation.

Current issues with methods and potential reform options		
Issue	Impact	Potential solutions
Interaction effects are only partially accounted for and lacks transparency or standardization.	Important synergies or necessary interactions for measures to be effective are not sufficiently measured. This may lead to inefficient planning and distribution of resources.	<ul style="list-style-type: none"> Methodology of policy instrument interactions should be further developed as part of the bottom-up assessment methodology. Further standardization of practices in applied evaluations.
Limited use of Control Group Design and Counterfactual Analysis: This approach is recognized as an option but is often not applied due to concerns about cost and data protection.	Difficulties in determining the effects of policies, additionality, and the impacts of market mechanism like taxes.	<ul style="list-style-type: none"> Funding should be provided to conduct randomized control group research. This could be conducted with leading scientific institutes. Data protection issues would need to be resolved, either as part of more general reforms, or research-specific exemptions/waivers.
Assumptions not stated clearly in evaluations.	Reduced transparency and replicability, and reliability	<ul style="list-style-type: none"> Ensuring transparency and reliability of evaluation methods is crucial. Currently, the evaluation procedures primarily focus on static effects, and there is a need to incorporate lifetime effects of measures. Enhancing replicability in evaluations is important for ensuring the credibility and robustness of the results. Further standardization efforts are necessary.

Table 12 - Key reform options for improving evaluation methods.

7.5. Use of evaluations and dissemination

Consistency between different reporting requirements. Ensuring consistency between different reporting requirements is essential for enhancing external transparency and facilitating the translation of outputs across various programmes. However, inconsistencies often arise due to differing outputs and parameters specified in different pieces of ordinance. To address this issue, greater coordination is needed across the ministries responsible for these programmes, along with the establishment of common goals. The presence of different outputs and parameters for reporting requirements can create challenges in terms of understanding and comparing the outcomes of different programmes. This lack of consistency reduces external transparency, as stakeholders find it difficult to interpret and compare the reported outputs effectively. It is essential to establish clear and standardized reporting requirements that encompass key indicators and metrics relevant to the goals of the programmes. To achieve consistency, coordination efforts across ministries responsible for these programmes are crucial. By aligning their objectives and working together, ministries can streamline reporting requirements and ensure that the same outputs and parameters are used consistently across different programmes. This coordination should include regular communication and collaboration to harmonize reporting frameworks and establish common guidelines for data collection, analysis, and reporting. Furthermore, it is important to establish common goals across the programmes. When programmes have different objectives, it becomes more challenging to achieve consistency in reporting requirements. By defining shared goals and aligning the programmes towards those common objectives, it becomes easier to establish consistent reporting frameworks. This will enable stakeholders to compare and evaluate the outcomes of different programmes more effectively.

Government can actively lead the reform of evaluation procedures, requiring political will to drive reforms, and increased internal capacities to engage with defining scope and methodologies. Better specification of evaluation parameters would enhance the value of outputs, allowing for improved forecasting and ex-ante assessments. To comply with EU requirements, as well as for better evaluation and forecasting, careful attention

should be given to distributional effects to avoid regressive outcomes and political repercussions. Furthermore, expanding the scope of evaluations to include issues related to governance effectiveness, administration, and information requirements is essential for a comprehensive assessment. By taking these steps, the government can drive meaningful reforms and ensure that evaluations provide valuable insights for informed decision-making and policy improvement.

Current issues with use/dissemination and potential reform options		
Issue	Impact	Potential solutions
Ex-ante assessments not as transparent in data sources and evidence.	Targets and programmes may lack reliable evidence base.	<ul style="list-style-type: none"> Formalize disclosure, transparency, and use of evidence in decision making.
Incoherent objectives across strategies and goals result in different reporting requirements.	Results across programmes are calculated with different assumptions and are presented differently making dissemination difficult.	<ul style="list-style-type: none"> Involve greater diversity of sector specialists and consultants in the target and strategy formulation processes. Ensure consistency between different reporting requirements

Table 13 - Key reform options for improving use of evaluations and dissemination.

7.6. Compatibility of key reforms with current institutional configuration

Key reform options are represented in terms of their compatibility with the current institutional configuration. Reforms are represented in terms of their anticipated impact of enabling better quality evaluation processes (Figure 10). Highly compatible reform options are those which could be implemented without significantly changing embedded formal institutional rules (such as data protection laws). Options with current low compatibility with the institutional configuration that offer high potential impact should still be prioritised as necessary, whilst acknowledging that procedural rules and required political consensus may hinder how quickly these embedded rules could be reformed.

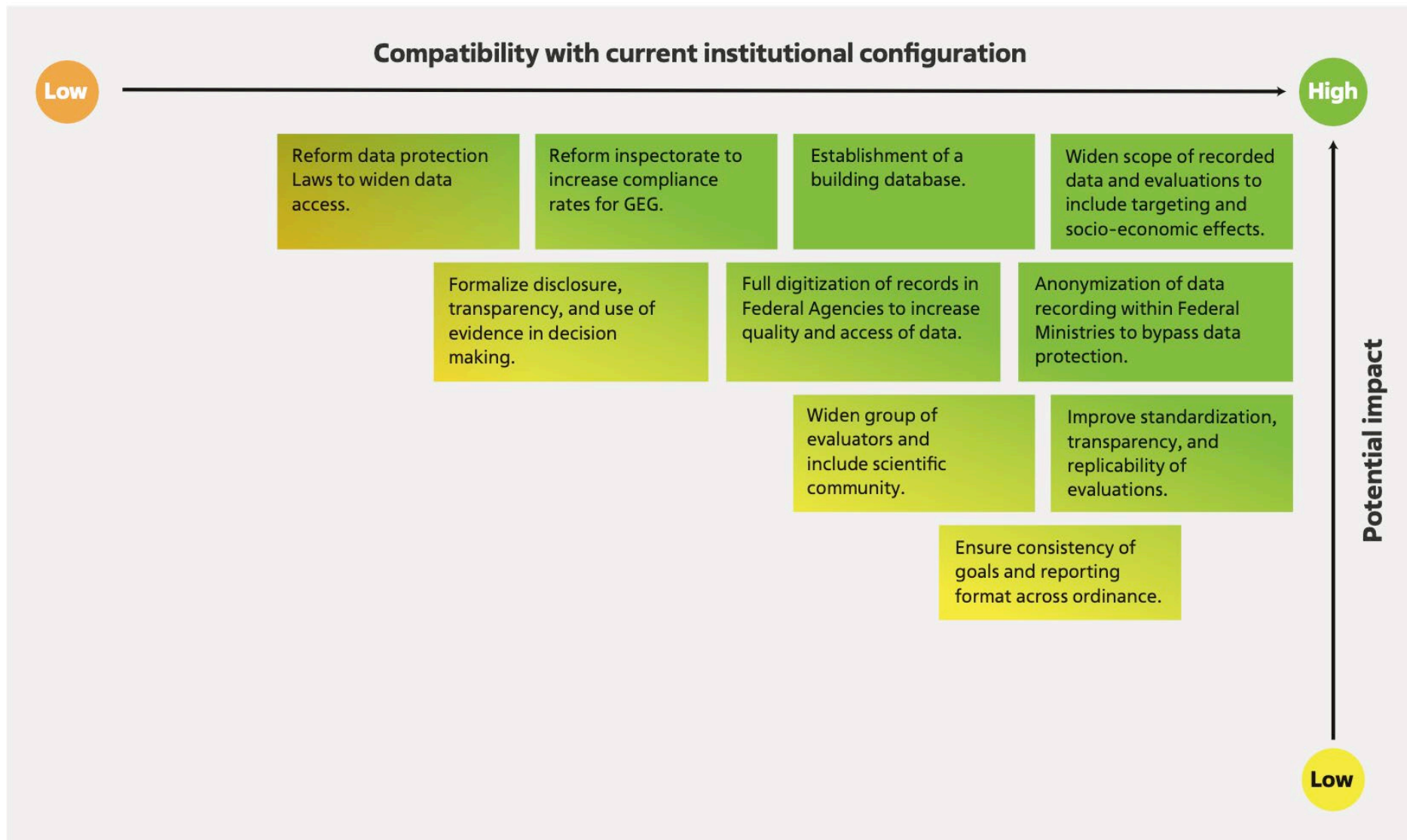


Figure 10 - Reform options ranked by impact and compatibility with the current institutional configuration.

8. Conclusions

This report utilizes an analytical approach that establishes a connection between institutions and the quality and utilization of evaluations, monitoring, and reporting within the policy process. By utilising this approach, a comprehensive understanding of the configuration, commissioning, and execution of evaluation processes is achieved. This goes beyond existing research that primarily focuses on the quality of evaluation outputs or on state capacities, without linking these two aspects. The research has generated several key considerations that warrant attention.

Inaccuracies in evaluating instrument performance may have significant repercussions on projecting the future greenhouse gas (GHG) reduction effects of policy mixes.

There is a risk that current evaluations may overestimate the performance of programmes. This arises due to two main factors. First, potential overestimation of savings from installed BEG measures due to pre-bound effects, i.e. the energy use before the installation of measures is over approximated leading to larger recorded energy savings than in reality. Second, if regulation is not effectively inspected and enforced, then a significant amount of anticipated GHG reductions from the GEG programme may not materialise due to evasion and non-compliance. Whilst it is beyond the scope of this report to estimate what the combined effects of these factors may be on GHG abatement progress, it is plausible that progress in the sector may be hindered by unrealistic expectations about the effectiveness of currently implemented instruments in ex-ante assessment models. In a worst case, the level of ambition enshrined in current mitigation strategies might not be sufficient to realise abatement targets, due to a performance gap between anticipated effectiveness and reality.

A key area of concern is the current lack of robust evaluation for the German Energy Act (GEG). The current evaluation processes for the GEG is limited to some regulatory measures included in the NAPE reporting. However, there is a lack of reliable data in many necessary criteria to thoroughly assess their impact. Additionally, the lack of enforcement and prevalent non-compliance within the building sector undermine the effectiveness of the regulations in place. Evaluation and expansion of the governance capaci-

ties to effectively administer and enforce regulatory measures is a key priority. Data limitations and difficulties in accessing regional data pose significant challenges to accountability and undermine the credibility of the inspectorate regime responsible for enforcing regulations. This lack of transparency and accountability further hinders the evaluation of energy efficiency goals set by the regulations. Addressing these areas of significant uncertainty is a key priority.

The current scope of the evaluation metrics is limited, and insufficient to generate evidence on key criteria such as distributional impacts and dynamic cost effectiveness.

Current evaluation practices need to be further developed in terms of the scope of evaluation metrics, which is currently limited. We identify several aspects which could be further developed, but most notable are in the need to further focus on distributional impacts and targeting of instruments, and expansion of a more dynamic perspective on evaluating cost effectiveness. To enhance both of these aspects not only requires methodological improvements to gain a more comprehensive understanding of the sector's performance, but also necessitates reforms must be implemented to enhance the quality and accessibility of data recording and provision. Particularly, distributional impacts should urgently be more comprehensively incorporated, and where possible integrated into ex-ante assessment and policy design. Failure to sufficiently anticipate regressive outcomes will likely lead to opposition to policy and political pressure to relax ambition. The current limited assessment of the impacts of the GEG on low-income groups, for example, is a significant oversight. Similarly, recent evaluation of the BEG indicates that it has predominantly benefitted medium-to high income groups. A reassessment of the design and targeting of current support programmes is needed, while implementation of additional programmes specifically designed at low-income housing, to reduce regressive outcomes of regulation (similar to France, or notably the US's recent Inflation Reduction Act) could also be considered.

Current application of methodologies lacks transparency in key aspects, which may favour more static perspectives on emission reduction and cost effectiveness. Notably, application of emission factors lacks transparency in analysed reports. Published reports do not clearly explain how emission factors were used to forecast energy saving effects.

Our assessment suggests static application of emission factors, not accounting for evolving cost savings over time. While energy savings are projected over the measure's lifecycle, using static emission factors may lead to underestimated GHG reduction from heat pump installation. This may favour renovation over decarbonisation, especially in the near term. A more dynamic perspective should extend beyond anticipated GHG reduction and balance cost projections of electricity decarbonisation and anticipated demand/capacity with potential savings from high standards of renovation.

We make several suggestions for reforms which would improve the quality and access of data. Establishment of a buildings database is a priority action point. The Federal Government has drafted proposals which would partly fulfil this role, and the renovation strategy as part of the revised EPBD (once implemented) will require one to be established. Prioritising reforms to data reporting and protection can be implemented would allow for current data to be stored (not as random samples) and would help improve the quality of such a database going forward. Without immediate action, there risks being a lack of historical data to draw on once reforms to establish a database are enabled.

Of note is the scarcity of data on energy use before the implementation of measures. Data access and unreliable information recording practices makes it difficult to accurately assess the impact of the measures on energy consumption and efficiency. Energy use before the installation can be overestimated prior to the installation, which then makes the measures appear to have achieved higher energy savings than in reality. This is central to the calculation of energy savings and for GHG abatement, and the aggregate effectiveness of the programme. It also means that the evaluation of cost effectiveness may be unreliable, and may present high cost renovation measures as more viable. Moreover, current adjustment effects are commonly excluded in practice due to insufficient data. To more effectively assess adoption rates and 'additionality', alternative methods need to be employed. Currently, a self-reporting bias exists due to the lack of reporting obligations for non-subsidized buildings, which hampers the collection of accurate and representative data.

Forecasting and projections are heavily reliant on the quality of analytical inputs.

While some uncertainty is unavoidable in such exercises due to unforeseeable changes

and shocks, improving the analytical capacity to strategize is essential. This requires the generation of data which is both sufficient in scope to anticipate the most potential adverse impacts of policy design, of rigorous methodological quality, while being transparent and replicable. Given the increased role of ex-ante assessment and forecasting introduced in through recent amendments to the KSG, addressing the aforementioned challenges in scope, data availability, enforcement, and accountability is essential to ensure more useful and effective evaluations practices.

A fundamental shift is required in the motivation and perceived role of conducting evaluations. Instead of primarily aiming for compliance with reporting obligations and requirements, or to justify federal spending budgets to the BRH, evaluations should be driven by the need to generate timely and reliable evidence. Better evidence is vital for making well-informed policy decisions and effectively recalibrating the existing policy mix. Identifying and addressing bottlenecks, whether in the capacity to administer, implement, enforce, or monitor, or in potential bottlenecks that may arise, can help mitigate adverse unintended consequences. By enhancing the quality of data, implementing comprehensive evaluation criteria, and bolstering enforcement measures, the governance of the German residential building sector can be significantly strengthened.

Tables

Table 1 - Policy mix design challenges and respective analytical components for evaluation.....	13
Table 2 - Websites screened for published evaluations for content analysis.....	22
Table 3 - Documents included in content analysis.....	24
Table 4 - Interview Participants.....	26
Table 5 - Formal institutions (ordinance) in the residential buildings sector.....	31
Table 6 - KfW Efficiency standards for renovations.....	41
Table 7 - Scope of ex-post evaluations.....	59
Table 8 - Funding available for KfW standards.....	64
Table 9 - Key reform options for improving institutional configuration	101
Table 10 - Key reform options for improving scope of programmes evaluated and evaluation metrics.....	103
Table 11 - Key reform options for improving data access and quality	104
Table 12 - Key reform options for improving evaluation methods	105
Table 13 - Key reform options for improving use of evaluations and dissemination	107

Figures

Figure 1 - Structural institutions for delivery and evaluation of the BEG and GEG	38
Figure 2 - Representation of bottom-up and top-down approaches.....	52
Figure 3 - Reporting of the GEG programme in most recently published NAPE report	54
Figure 4 - Metric adjustments applied to final energy savings.....	60
Figure 5 - Calculation for private investment caused and jobs created.....	66
Figure 6 - Visualisation of data quality	72
Figure 7 - Calculation for final energy savings.....	82
Figure 8 - Effect adjustments to calculate Net effects from Gross	85
Figure 9 - Use of gross/net effects in methodological guidelines.....	86
Figure 10 - Reform options and compatibility with the current institutional configuration	108

References

- Altermatt, P.P., Clausen, J., Brendel, H., Breyer, C., Gerhards, C., Kemfert, C., Weber, U., Wright, M., 2023. Replacing gas boilers with heat pumps is the fastest way to cut German gas consumption. *Commun Earth Environ* 4, 56. <https://doi.org/10.1038/s43247-023-00715-7>
- Auld, G., Mallett, A., Burlica, B., Nolan-Poupart, F., Slater, R., 2014. Evaluating the effects of policy innovations: Lessons from a systematic review of policies promoting low-carbon technology. *Global Environmental Change* 29, 444–458. <https://doi.org/10.1016/j.gloenvcha.2014.03.002>
- Baek, C., Park, S., 2012. Policy measures to overcome barriers to energy renovation of existing buildings. *Renewable and Sustainable Energy Reviews* 16, 3939–3947. <https://doi.org/10.1016/j.rser.2012.03.046>
- BDEW, 2023. Zum Entwurf eines Gesetzes für die Wärmeplanung und zur Dekarbonisierung der Wärmenetze vom 21. Juli 2023. Berlin.
- Berneiser, J., Burkhardt, A., Henger, R., Köhler, B., Meyer, R., Sommer, S., Yilmaz, Y., Kost, C., Herkel, S., 2021. Maßnahmen und Instrumente für eine ambitionierte, klimafreundliche und sozialverträgliche Wärme-wende im Gebäudesektor.
- Bovens, M., Schillemans, T., Hart, P., 2008. DOES PUBLIC ACCOUNTABILITY WORK? AN ASSESSMENT TOOL. *Public Adm* 86, 225–242. <https://doi.org/10.1111/j.1467-9299.2008.00716.x>
- Brown, R.R., Farrelly, M.A., Loorbach, D.A., 2013. Actors working the institutions in sustainability transitions: The case of Melbourne’s stormwater management. *Global Environmental Change* 23, 701–718. <https://doi.org/10.1016/j.gloenvcha.2013.02.013>
- Bugden, D., Stedman, R., 2019. A synthetic view of acceptance and engagement with smart meters in the United States. *Energy Res Soc Sci* 47, 137–145. <https://doi.org/10.1016/j.erss.2018.08.025>
- Cairney, P., Oliver, K., Wellstead, A., 2016. To Bridge the Divide between Evidence and Policy: Reduce Ambiguity as Much as Uncertainty. *Public Adm Rev* 76, 399–402. <https://doi.org/10.1111/puar.12555>
- Dubash, N.K., Pillai, A.V., Flachsland, C., Harrison, K., Hochstetler, K., Lockwood, M., MacNeil, R., Mildenerger, M., Paterson, M., Teng, F., Tyler, E., 2021. National climate institutions complement targets and policies. *Science (1979)* 374, 690–693. <https://doi.org/10.1126/science.abm1157>
- Edmondson, D., Flachsland, C., aus dem Moore, N., Koch, N., Koller, F., Brehm, J., Gruhl, H., Levi, S., 2022. Assessing Climate Policy Instrument Mix Pathways: An application to the German light duty vehicle sector. Berlin .
- Edmondson, D., Flachsland, C., Aus Dem Moore, N., Koch, N., Koller, F., Gruhl, H., Brehm, J., 2023. Anticipatory climate policy mix pathways: A framework for ex-ante construction and assessment applied to the road transport sector. Under review at *Climate Policy* .
- Edmondson, D.L., Rogge, K.S., Kern, F., 2020. Zero carbon homes in the UK? Analysing the co-evolution of policy mix and socio-technical system. *Environ Innov Soc Transit* 35, 135–161. <https://doi.org/10.1016/j.eist.2020.02.005>
- ERK, 2022. Zweijahresgutachten 2022 Gutachten zu bisherigen Entwicklungen der Treibhausgasemissionen, Trends der Jahresemissionsmengen und Wirksamkeit von Maßnahmen (gemäß § 12 Abs. 4 Bundes-Klimaschutzgesetz). Berlin.
- ESABCC, 2023. Scientific advice for the determination of an EU-wide 2040 climate target and a greenhouse gas budget for 2030-2050. <https://doi.org/10.2800/609405>
- Finnegan, J.J., 2022. Institutions, Climate Change, and the Foundations of Long-Term Policymaking. *Comp Polit Stud* 55, 1198–1235. <https://doi.org/10.1177/00104140211047416>
- Fishburn, P.C., 1988. NORMATIVE THEORIES OF DECISION MAKING UNDER RISK AND UNDER UNCERTAINTY, in: *Decision Making*. Cambridge University Press, pp. 78–98. <https://doi.org/10.1017/CBO9780511598951.006>

- Fujiwara, N., van Asselt, H., Bößner, S., Voigt, S., Spyridaki, N.-A., Flamos, A., Alberola, E., Williges, K., Türk, A., ten Donkelaar, M., 2019. The practice of climate change policy evaluations in the European Union and its member states: results from a meta-analysis. *Sustainable Earth* 2, 9. <https://doi.org/10.1186/s42055-019-0015-8>
- Galvin, R., 2023. Policy pressure to retrofit Germany's residential buildings to higher energy efficiency standards: A cost-effective way to reduce CO2 emissions? *Build Environ* 237, 110316. <https://doi.org/10.1016/j.buildenv.2023.110316>
- Galvin, R., Sunikka-Blank, M., 2016. Quantification of (p)rebound effects in retrofit policies – Why does it matter? *Energy* 95, 415–424. <https://doi.org/10.1016/j.energy.2015.12.034>
- Garmston, H., Pan, W., 2013. Non-Compliance with Building Energy Regulations: The Profile, Issues, and Implications on Practice and Policy in England and Wales Non-Compliance with Building Energy Regulations: The Profile, Issues, and Implications on Practice and Policy in England and. *Journal of Sustainable Development of Energy J. sustain. dev. energy water environ. syst* 1, 340–351. <https://doi.org/10.13044/j.sdewes.2013.01.0026>
- George, J.F., Werner, S., Preuß, S., Winkler, J., Held, A., Ragwitz, M., 2023. The landlord-tenant dilemma: Distributional effects of carbon prices, redistribution and building modernisation policies in the German heating transition. *Appl Energy* 339, 120783. <https://doi.org/10.1016/j.apenergy.2023.120783>
- Gillard, R., 2016. Unravelling the United Kingdom's climate policy consensus: The power of ideas, discourse and institutions. *Global Environmental Change* 40, 26–36. <https://doi.org/10.1016/j.gloenvcha.2016.06.012>
- Guy, J., Shears, E., Meckling, J., 2023. National models of climate governance among major emitters. *Nat Clim Chang* 13, 748–748. <https://doi.org/10.1038/s41558-023-01688-3>
- Hacker, J., Pierson, P., Thelen, K., 2015. Drift and Conversion: Hidden Faces of Institutional Change. *Advances in Comparative-Historical Analysis*.
- Haug, C., Rayner, T., Jordan, A., Hildingsson, R., Stripple, J., Monni, S., Huitema, D., Massey, E., van Asselt, H., Berkhout, F., 2010. Navigating the dilemmas of climate policy in Europe: evidence from policy evaluation studies. *Clim Change* 101, 427–445. <https://doi.org/10.1007/s10584-009-9682-3>
- Hendriks, F., Tops, P., 2003. Local public management reforms in the Netherlands: Fads, fashions and winds of change. *Public Adm* 81, 301–323. <https://doi.org/10.1111/1467-9299.00348>
- Hildén, M., 2011. The evolution of climate policies – the role of learning and evaluations. *J Clean Prod* 19, 1798–1811. <https://doi.org/10.1016/j.jclepro.2011.05.004>
- Hovi, J., Greaker, M., Hagem, C., Holtsmark, B., 2012. A credible compliance enforcement system for the climate regime. *Climate Policy* 12, 741–754. <https://doi.org/10.1080/14693062.2012.692206>
- Howlett, M., 1998. Predictable and Unpredictable Policy Windows: Institutional and Exogenous Correlates of Canadian Federal Agenda-Setting. *Canadian Journal of Political Science* 31, 495–524. <https://doi.org/10.1017/S0008423900009100>
- Howlett, M., Ramesh, M., 2003. *Studying Public Policy: Policy Cycles and Policy Subsystems*. Oxford University Press.
- Howlett, M., Ramesh, M., Stewart, J., 1996. Studying public policy: policy cycles and policy subsystems. *Aust J Polit Sci* 31, 265–266.
- Huitema, D., Jordan, A., Massey, E., Rayner, T., van Asselt, H., Haug, C., Hildingsson, R., Monni, S., Stripple, J., 2011. The evaluation of climate policy: theory and emerging practice in Europe. *Policy Sci* 44, 179–198. <https://doi.org/10.1007/s11077-011-9125-7>
- Jacob, K., Kannen, H., 2015. *Climate Policy Integration in Federal Settings: the Case of Germany's Building Policy*. Berlin.
- Jacobs, A.M., Weaver, R.K., 2015. When Policies Undo Themselves: Self-Undermining Feedback as a Source of Policy Change. *Governance* 28, 441–457. <https://doi.org/10.1111/gove.12101>

- Jagnow, K., Wolff, D., 2020. Gutachten: Energetische Gebäudesanierung. Berlin.
- Jordan, A., Huitema, D., Schoenefeld, J., van Asselt, H., Forster, J., 2018. Governing Climate Change Polycentrically, in: *Governing Climate Change*. Cambridge University Press, pp. 3–26. <https://doi.org/10.1017/9781108284646.002>
- Kaufmann, W., Hooghiemstra, R., Feeney, M.K., 2018. Formal institutions, informal institutions, and red tape: A comparative study. *Public Adm* 96, 386–403. <https://doi.org/10.1111/padm.12397>
- Kern, F., 2011. Ideas, institutions, and interests: Explaining policy divergence in fostering “system innovations” towards sustainability. *Environ Plann C Gov Policy* 29, 1116–1134. <https://doi.org/10.1068/c1142>
- Köhler, J., Geels, F.W., Kern, F., Markard, J., Onsongo, E., Wieczorek, A., Alkemade, F., Avelino, F., Bergek, A., Boons, F., Fünfschilling, L., Hess, D., Holtz, G., Hyysalo, S., Jenkins, K., Kivimaa, P., Martiskainen, M., Mcmeekin, A., Susan, M., Nykvist, B., Pel, B., Raven, R., Rohrer, H., Sandén, B., Schot, J., Sovacool, B., Turnheim, B., Welch, D., Wells, P., 2019. An agenda for sustainability transitions research : State of the art and future directions ☆. *Environ Innov Soc Transit* 1–32. <https://doi.org/10.1016/j.eist.2019.01.004>
- Krippendorff, K., 2004. Reliability in Content Analysis. *Hum Commun Res* 30, 411–433. <https://doi.org/10.1111/j.1468-2958.2004.tb00738.x>
- Kuzemko, C., 2016. Energy Depoliticisation in the UK : Depoliticisation Destroying Political Energy in the UK : <https://doi.org/10.1111/1467-856X.12068>
- Levi, S., 2021. Why hate carbon taxes? Machine learning evidence on the roles of personal responsibility, trust, revenue recycling, and other factors across 23 European countries. *Energy Res Soc Sci* 73, 101883. <https://doi.org/10.1016/j.erss.2020.101883>
- Lockwood, M., Kuzemko, C., Mitchell, C., Hoggett, R., 2017. Historical institutionalism and the politics of sustainable energy transitions: A research agenda. *Environ Plann C Gov Policy* 35, 312–333. <https://doi.org/10.1177/0263774X16660561>
- Lu, Y., Karunasena, G., Liu, C., 2022. A Systematic Literature Review of Non-Compliance with Low-Carbon Building Regulations. *Energies (Basel)* 15, 9266. <https://doi.org/10.3390/en15249266>
- Magro, E., Wilson, J.R., 2017. Governing the Evaluation of Policy Mixes in the Context of Smart Specialisation Strategies: Emerging Challenges First Draft.
- Mastenbroek, E., van Voorst, S., Meuwese, A., 2016. Closing the regulatory cycle? A meta evaluation of ex-post legislative evaluations by the European Commission. *J Eur Public Policy* 23, 1329–1348. <https://doi.org/10.1080/13501763.2015.1076874>
- Mayring, P., 2000. Qualitative Content Analysis.
- Meckling, J., Nahm, J., 2018. The power of process: State capacity and climate policy. *Governance* 31, 741–757. <https://doi.org/10.1111/gove.12338>
- Melvin, J., 2018. The split incentives energy efficiency problem: Evidence of underinvestment by landlords. *Energy Policy* 115, 342–352. <https://doi.org/10.1016/j.enpol.2017.11.069>
- Meyer, R., Berneiser, J., Burkhardt, A., Doderer, H., Eickelmann, E., Henger, R., Köhler, B., Sommer, S., Yilmaz, Y., Blesl, M., Bürger, V., Braungardt, S., 2021. Maßnahmen und Instrumente für eine ambitionierte, klimafreundliche und sozialverträgliche Wärme-wende im Gebäudesektor: Teil 2: Instrumentensteckbriefe für den Gebäudesektor.
- Michaelowa, A., 2008. German Climate Policy Between Global Leadership and Muddling Through, in: *Turning Down the Heat*. Palgrave Macmillan UK, London, pp. 144–163. https://doi.org/10.1057/9780230594678_9
- Michaelowa, A., Allen, M., Sha, F., 2018. Policy instruments for limiting global temperature rise to 1.5°C – can humanity rise to the challenge? *Climate Policy* 18, 275–286. <https://doi.org/10.1080/14693062.2018.1426977>

- Moore, T., Doyon, A., 2023. Facilitating the Sustainable Housing Transition, in: *A Transition to Sustainable Housing*. Springer Nature Singapore, Singapore, pp. 239–258. https://doi.org/10.1007/978-981-99-2760-9_8
- Morrison, T.H., 2022. Crafting more anticipatory policy pathways. *Nat Sustain* 5, 372–373. <https://doi.org/10.1038/s41893-022-00894-9>
- Mukherjee, I., Coban, M.K., Bali, A.S., 2021. Policy capacities and effective policy design: a review. *Policy Sci* 54, 243–268. <https://doi.org/10.1007/s11077-021-09420-8>
- Öberg, P., Lundin, M., Thelander, J., 2015. Political Power and Policy Design: Why Are Policy Alternatives Constrained? *Policy Studies Journal* 43, 93–114. <https://doi.org/10.1111/psj.12086>
- Patashnik, E.M., 2009. *Reforms at Risk*. Princeton University Press. <https://doi.org/10.1515/9781400828852>
- Peters, B.G., 2012. *Institutional theory in political science : the new institutionalism* / B. Guy Peters. Continuum, New York.
- Qian, Q.K., Fan, K., Chan, E.H.W., 2016. Regulatory incentives for green buildings: gross floor area concessions. *Building Research & Information* 44, 675–693. <https://doi.org/10.1080/09613218.2016.1181874>
- Rosenow, J., Galvin, R., 2013. Evaluating the evaluations: Evidence from energy efficiency programmes in Germany and the UK. *Energy Build* 62, 450–458. <https://doi.org/10.1016/j.enbuild.2013.03.021>
- Rosenow, J., Kern, F., Rogge, K., 2017. The need for comprehensive and well targeted instrument mixes to stimulate energy transitions: The case of energy efficiency policy. *Energy Res Soc Sci* 33. <https://doi.org/10.1016/j.erss.2017.09.013>
- Sabatier, P., Weible, C., 2014. *Theories of the Policy Process*. Routledge.
- Schmidt, V.A., 2008. Discursive Institutionalism: The Explanatory Power of Ideas and Discourse. *Annual Review of Political Science* 11, 303–326. <https://doi.org/10.1146/annurev.polisci.11.060606.135342>
- Schoenefeld, J., Jordan, A., 2017. Governing policy evaluation? Towards a new typology. *Evaluation* 23, 274–293. <https://doi.org/10.1177/1356389017715366>
- Singhal, P., Pahle, M., Kalkuhl, M., Levesque, A., Sommer, S., Berneiser, J., 2022. Beyond good faith: Why evidence-based policy is necessary to decarbonize buildings cost-effectively in Germany. *Energy Policy* 169, 113191. <https://doi.org/10.1016/j.enpol.2022.113191>
- Skogstad, G., 2023. Historical Institutionalism in Public Policy, in: *Encyclopedia of Public Policy*. Springer International Publishing, Cham, pp. 1–9. https://doi.org/10.1007/978-3-030-90434-0_21-1
- Steinmo, S., Thelen, K., 1992. *Structuring Politics: Historical Institutionalism in Comparative Analysis*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511528125>
- Sunikka-Blank, M., Galvin, R., 2012. Introducing the prebound effect: the gap between performance and actual energy consumption. *Building Research & Information* 40, 260–273. <https://doi.org/10.1080/09613218.2012.690952>
- Thelen, K., 1999. HISTORICAL INSTITUTIONALISM IN COMPARATIVE POLITICS 369–404.
- Tosun, J., De Francesco, F., Peters, B.G., 2019. From environmental policy concepts to practicable tools: Knowledge creation and delegation in multilevel systems. *Public Adm* 97, 399–412. <https://doi.org/10.1111/padm.12544>
- Tversky, A., Kahneman, D., 1981. The framing of decisions and the psychology of choice. *Science* (1979) 211, 453–458. <https://doi.org/10.1126/science.7455683>
- van den Bergh, J., Castro, J., Drews, S., Exadaktylos, F., Foramitti, J., Klein, F., Konc, T., Savin, I., 2021. Designing an effective climate-policy mix: accounting for instrument synergy. *Climate Policy*. <https://doi.org/10.1080/14693062.2021.1907276>
- von Lüpke, H., Leopold, L., Tosun, J., 2022. Institutional coordination arrangements as elements of policy design spaces: insights from climate policy. *Policy Sci*. <https://doi.org/10.1007/s11077-022-09484-0>

- Wang, Q., Hubacek, K., Feng, K., Wei, Y.-M., Liang, Q.-M., 2016. Distributional effects of carbon taxation. *Appl Energy* 184, 1123–1131. <https://doi.org/10.1016/j.apenergy.2016.06.083>
- Weber, K.M., Rohracher, H., 2012. Legitimizing research, technology and innovation policies for transformative change: Combining insights from innovation systems and multi-level perspective in a comprehensive “failures” framework. *Res Policy* 41, 1037–1047. <https://doi.org/10.1016/j.respol.2011.10.015>
- Weiß, J., Vogelpohl, T., 2010. Politische Instrumente zur Erhöhung der energetischen Sanierungsquote bei Eigenheimen.
- Wurzel, R.K.W., 2010. Environmental, Climate and Energy Policies: Path-Dependent Incrementalism or Quantum Leap? *Ger Polit* 19, 460–478. <https://doi.org/10.1080/09644008.2010.515838>
- Zachmann, G., Fredriksson, G., Claeys, G., 2018. THE DISTRIBUTIONAL EFFECTS OF CLIMATE POLICIES.

Annex I: Screened documents

In-cluded	Type	Title	Year	Authors	Client
No	Ex-ante evaluation	Kurzgutachten zur Überarbeitung von Anforderungssystemen und Standards im Gebäudeenergiegesetz für Neubauten sowie Bestandsgebäude einschl. der Wirtschaftlichkeitsbetrachtungen für Neubauten und Bestandsgebäude	2022	Dena, Guidehouse, iTG, Öko-Institut, Prognos, FIW München, ifeu	BMWK
No	Ex-ante evaluation	Einsparpotenziale aus der Optimierung von Heizungsanlagen in Wohngebäuden	2022	Borderstep Institut, Dena	BMWK
No	Ex-ante evaluation	Mindestvorgaben für die Gesamteffizienz von Bestandsgebäuden	2022	bbh, ifeu, Öko-Institut, Prognos	BfEE
Yes	Ex-post evaluation	Evaluation und Perspektiven des Marktanzreizprogramms zur Förderung von Maßnahmen zur Nutzung erneuerbarer Energien im Wärmemarkt im Förderzeitraum 2019 bis 2020	n.d.	ifeu, Fraunhofer, Fichtner GmbH	BMWK
Yes	Ex-post evaluation	Abschlussbericht zur Evaluation der Richtlinie über die Förderung der Heizungsoptimierung durch hocheffiziente Pumpen und hydraulischen Abgleich	2022	Arepo, Wuppertal Institut	BfEE
Yes	Ex-post evaluation	Evaluation des Förderprogramms KfW 433	2022	Prognos	BMWK
Yes	Ex-post evaluation	Evaluation der Förderprogramme EBS WG im Förderzeitraum 2018	2022	Prognos, FIW München	BMWK
Yes	Ex-post evaluation	Förderwirkungen BEG EM 2021	2023	Prognos, ifeu, FIW München, iTG	BMWK

Yes	Ex-post evaluation	Förderwirkungen BEG WG 2021	2023	Prognos, ifeu, FIW München, iTG	BMWK
No	Ex-post evaluation	Monitoring der Initiative Energieeffizienz-Netzwerke	2022	Adelphi, Fraunhofer	BMWK
Yes	Guidelines	Methodikpapier zur ex-ante Abschätzung der Energie- und THG-Minderungswirkung von energie- und Klimaschutzpolitischen Maßnahmen	2022	Prognos, Fraunhofer ISI, Öko-Institut e.V.	BMWK
Yes	Guidelines	Methodikleitfaden für Evaluationen von Energie-effizienzmaßnahmen des BMWi (Projekt Nr. 63/15 – Aufstockung)	2020	Prognos, ifeu, Stiftung Umweltenergierecht, Fraunhofer ISI	BMWK
No	National monitoring	DENA-GEBÄUDEREPORT 2023	2023	Dena	BMWK
Yes	National monitoring	Klimaschutz in Zahlen	2022	BMWK	n.a.
Yes	National monitoring	Klimaschutzbericht 2022	2022	BMWK	n.a.
Yes	National monitoring	Zweijahresgutachten 2022 Gutachten zu bisherigen Entwicklungen der Treibhausgasemissionen, Trends der Jahresemissionsmengen und Wirksamkeit von Maßnahmen	2022	Expertenrat für Klimafragen	n.a.
Yes	National monitoring	The Energy of the Future 8 th Monitoring Report on the Energy Transition – Reporting Years 2018 and 2019	2021	BMWK	n.a.
Yes	National monitoring	Energieeffizienz in Zahlen	2021	BMWK	n.a.
Yes	Other	Begleitung von BMWK-Maßnahmen zur Umsetzung einer Wärmepumpen-Offensive	2023	Dena, Guidehouse, iTG, Öko-Institut, Prognos,	BMWK

				EY, pwc, bbh, FIW München, ifeu, heimrich + hannot	
No	Other	Metastudie zur Verbesserung der Datengrundlage im Gebäudebereich	2022	bbh, dena, EY, FIW München, heimrich+hannot	BMWK
No	Quarterly reporting	Bundesförderung für effiziente Gebäude (BEG) Reporting zur BEG-Förderung im 4. Quartal 2022 (Stand: 31.12.2022)	2022	BMWK	n.a.
No	Strategy	Sofortprogramm gemäß § 8 Abs. 1 KSG für den Sektor Gebäude	2022	BMWK, BMWSB	n.a.
No	Strategy	Hintergrundpapier zur Gebäudestrategie Klimaneutralität 2045	2022	Dena, Guidehouse, iTG, Öko-Institut, Prognos, EY, adelphi, bbh, FIW München, ifeu	BMWK
No	Strategy	Klimaneutrales Deutschland 2045	2021	Prognos, Öko-Institut, Wuppertal-Institut	Stiftung Klimaneutralität, Agora Energiewende, Agora Verkehrswende
No	Strategy	Energieeffizienzstrategie 2050	2019	BMWK	n.a.
Yes	Strategy	National Action Plan on Energy Efficiency	2014	BMWK	n.a.
No	Strategy	Klimapfade 2.0 – Ein Wirtschaftsprogramm für Klima und Zukunft	2021	BCG	BDI
No	Strategy	Aufbruch Klimaneutralität	2021	EWI, FIW, ITG, (...)	Dena
No	Strategy	Langfristszenarien für die Transformation des Energiesystems in Deutschland	2022	Consentec, TU Berlin, ifeu	BMWK
No	Strategy	Deutschland auf dem Weg zur Klimaneutralität 2045	2021	PIK, MCC, PSI, RWI, IER, (...)	Ariadne - Kopernikus-Projekte
No	Yearly subsidy report	Förderreport KfW Bankengruppe	2022	KfW	n.a.

Table 14 – screened documents, n = 30

Annex II: Questionnaire for semi-structured interviews

A. Process and coordination

1. Which ex-post evaluation studies are being commissioned by ministries?
2. Which agencies and consultancies do they commission?
3. To what extent do ministries coordinate commissioning and using them?
4. Are non-commissioned studies taken into account?

B. Quality and scope

5. Are individual instruments evaluated or packages? In package assessments, are interactions among instruments analysed?
6. Do you think they are adequate in scope in terms of covering important evaluation criteria such as GHG reduction, energy efficiency improvements, costs, benefits, distributional impacts?
7. Do you think the ex-post policy evaluation are adequate in quality?
8. More broadly, which kind of ex-post evaluation studies would you find useful –in terms of evaluation criteria considered, building and technology scope, or any other relevant dimension?

C. Impact on policy process (reforms)

9. In your experience, in which ways do ex-post evaluation studies affect policymaking and reform?
10. More broadly: on which analytical input are reforms (proposals for reforms) building on? In-house or consultancies?
11. If you could make a suggestion to re-build the ex-post evaluation study framework from scratch (or even incrementally), how would you go about it?

Annex III: Additional information on Ordinance

Climate Action Programme 2020

The Climate Action Programme 2020 was first introduced in 2014. It contained 110 measures. For each measure, the amount of GHG emissions that should be saved by 2020 was determined (BRH 2022: 16). The implementation status and the expected effects of the measures were examined annually (so-called *quantification reports*). The results were published in the Federal Government's Climate Protection Report (BRH 2022: 16).

Assessment by the Federal Court of Auditors (BRH) determined that many measures in the Climate Action Programme 2020 remained ineffective (BRH 2022: 7). The German government repeatedly "*did not adopt*" the results of the quantification reports and other prognoses on GHG reductions (BRH 2022: 16). BRH argued that the calculations are only a rough estimate of the impact with a view to the target year (ex-ante) and would not replace a detailed and empirically supported evaluation of the individual measures (ex-post) (BRH 2022: 16). According to the BRH only 8 of 110 measures significantly contributed to GHG reduction, while 70 measures did not contribute to the GHG reduction. They are considered as accompanying instruments (BRH 2022: 16). This assessment is partially explained by BRH's criticism that quantifiable GHG saving are absent for many instruments.

Immediate Climate Action Programme 2022

The Immediate Climate Action Programme 2022 contained few new measures, instead increasing funding for existing programmes. The federal government intends to significantly expand the financing of important climate policy projects between 2022 and 2025 with more than 93 billion euros (BReg 2021). More than half of the additional funds in the emergency programme are earmarked to promote the energy-efficient refurbishment of buildings and the installation of energy-efficient heating systems (BReg 2021). An additional 4.5 billion euros alone are to be made available (BReg 2021). From 2023 onwards, the Federal Government no longer wants to promote heating systems that are powered exclusively by fossil fuels (BReg 2021). Assessment by the BRH that the Immediate Climate Action Programme 2022 contains few new measures and consists largely of funding increases for existing programmes or mere declarations of intent (BRH 2022: 20).

The BMU considered an ex-ante assessment of the GHG reductions of individual measures to be suitable to assess the effect of climate protection measures in a reliable and comparable

manner. However, such an assessment is not always possible (BRH 2022: 22). Continuous monitoring of the GHG reduction of individual measures beyond the annual climate protection reporting is not expedient. The BMF shares the position of BMU and additionally points out that the design and implementation of the Emergency Climate Protection Programme 2022 were omitted in view of the proximity to the Bundestag elections (BRH 2022: 22).

National Action Plan on Energy Efficiency

The NAPE (National Action Plan on Energy Efficiency) was designed as a comprehensive set of measures aimed at improving energy efficiency in Germany. NAPE was first implemented in 2014, alongside the climate action plan 2020, It outlined the strategic direction of energy efficiency policy, with three central objectives:

- Advancing energy efficiency in the building sector: The original NAPE emphasized the importance of improving energy efficiency in buildings as a key area for achieving energy savings and reducing emissions.
- Establishing energy efficiency as a yield and business model: NAPE aimed to promote energy efficiency as a viable business opportunity, encouraging investment in energy-efficient technologies and practices.
- Increasing self-responsibility for energy efficiency: NAPE sought to foster a sense of individual and collective responsibility for energy efficiency, encouraging citizens, businesses, and organizations to take proactive measures to reduce energy consumption.

NAPE 2.0, was implemented in 2019 as part of the Energy Efficiency Strategy 2050 (BMWi 2019). It incorporated policy learnings and adapts to new developments, particularly focusing on the timeframe from 2021 to 2030. The key differences between NAPE 2.0 and the original NAPE include:

- Stronger focus on the demand side of the energy system: NAPE 2.0 shifts its focus from the building sector to encompass a broader perspective on energy efficiency, with an emphasis on demand-side management and measures to reduce energy consumption throughout the economy.
- Integration of energy efficiency measures from the 2030 Climate Protection Programme: NAPE 2.0 incorporates relevant measures and targets from the 2030 Climate Protection

Programme to align energy efficiency objectives with broader climate change mitigation goals.

- Specific focus on reducing end energy consumption between 2021 and 2030: NAPE 2.0 sets explicit targets and measures to drive reductions in energy consumption during the specified time period, contributing to overall energy efficiency improvements.
- Introduction of an annual monitoring process: NAPE 2.0 establishes a monitoring mechanism to regularly assess the success and effectiveness of its measures. This ongoing monitoring allows for the identification of any necessary adjustments or modifications to enhance the plan's outcomes.

These adjustments and enhancements in NAPE 2.0 reflect the evolving understanding of energy efficiency challenges and opportunities, as well as the need to align with long-term climate and sustainability goals. NAPE 2.0 will be again updated this year in alignment with *Energieeffizienz für eine klimaneutrale Zukunft 2045*.

Data included in NAPE reporting focusses on energy savings and GHG abatement. At least the following data must be collected to fill the template used in NAPE monitoring (Schlomann et al. 2020: 50-51):

- New values added annually for the following categories:
 - Final energy savings for electricity
 - Final energy savings for fuels
 - Alternatively: aggregated values, but individual presentation preferred
 - Lifetime
 - Subsidy volume
 - Triggered investments
 - Number of cases of promotion or other activity parameter
- How the savings were determined
- Explanation of the choice of baseline
- Adjustment factors for effect adjustment

This reporting does not extend to other important criteria such as macro-economic effects, distributional impacts/socio-economic, or administrative challenges or costs (ex-post-guidelines, pp. 50-51).

Federal Funding for Efficient Buildings (BEG)

Support for the renovation of residential buildings includes the building envelope, plant technology, heat generation systems, heating optimisation, and technical planning and supervision. The BEG also includes other measures (e.g. new construction) but since these do not directly apply to residential retrofit, they are beyond the scope of this report. *Individual measures* on the building envelope contribute to increasing the building's energy efficiency, such as windows or doors and insulation of the exterior walls or roof. *Plant technology* (except heating), includes installation of systems technology in existing buildings to increase the energy efficiency of the building, such as an energy-efficient ventilation and air-conditioning system. *Heat generation systems* (heating technology) supports the installation of efficient heat generators, systems for heating support and the connection to a building, or heating network that integrates renewable energies for heat generation with a share of at least 25 percent. *Heating optimisation* supports measures to optimise the heating distribution system of a heat generation system in existing buildings (at least 2 years old) and increases the energy efficiency of the system, such as hydraulic balancing or replacing the heating pump. *Technical planning and construction supervision* provides specialist energy planning and construction supervision services in connection with the implementation of funded measures within the meaning of this funding programme.

Building Energy Act (GEG)

Retrofitting requirements. The GEG imposes so-called retrofitting requirements for certain parts (replacement of certain old boilers, insulation of certain pipelines, insulation of top floor ceilings, installation of certain control technology of heating and air-conditioning systems) independent of measures. Further retrofitting requirements are the subject of §47 (top floor ceilings), §61(1) (central control devices for central heating systems), §63 (3) (room-by-room control devices for central heating systems), §66 (control devices for room air humidity) and §§71 to 73 GEG (pipe insulation, decommissioning of old boilers). According to §46 GEG, exterior building components of a building may not be changed in a way that the energy quality of a building deteriorates. An exception concerns alterations if the area of the altered building components does not affect more than 10 % of the total area of the building component group according to Annex 7. In addition, the requirements for existing buildings do not apply if compliance conflicts with other regulations on health, noise, fire protection, occupational safety and stability.

Conditional requirements for refurbishment (§ 48 GEG). According to § 48 GEG requirements for energetic quality are attached to alterations according to Annex 7 GEG. If major measures are

required on a building component anyway (e.g. for reasons of building maintenance or duty of care) or are planned for other reasons, energy standards must also be met for the affected partial areas. The building component requirements according to § 48 GEG in conjunction with Annex 7 GEG must be complied with if certain measures are carried out. The relevant measures associated with conditional requirements are described in the second column of the table in Annex 7 GEG. In cases of renewal of exterior walls, building components in the roof area, walls against unheated rooms or the ground and ceiling areas that are separated downwards from unheated rooms, outside air or the ground, the conditional requirements do not apply to building components that were constructed or renewed in compliance with energy-saving regulations after 31 December 1983 (i.e. after the 2nd Thermal Insulation Ordinance came into force).

Overall verification can be applied to the building. Instead of compliance with the building component requirements for all relevant changes made to the building envelope, an overall verification in accordance with §50 can also be carried out for the building. In the overall verification pursuant to §50, it shall be shown that a residential building does not exceed the primary energy demand of the reference building pursuant to Annex 1 GEG by more than 40 % after renovation has been carried out and that its specific transmission heat demand is limited pursuant to Article 50(2). If, at the time of the refurbishment of a building, an extension of the existing part of the building is carried out at the same time and the overall verification pursuant to § 50 GEG is chosen, the calculations must be carried out for the entire building. Based on § 79 (2) GEG, in this case an energy performance certificate shall be issued for the entire building and marked on page 1 as occasion modernization. Compliance with the requirements for the new part of the building must be demonstrated independently and separately.

Annex IV: Recommended procedures

Box 1: Ex- post evaluations – recommended procedure

Step:

1. Characterisation of the funding programme
 - General overview of the measure, including its target group/sectors, budget, funding/implementing bodies, legal basis, related policies and funding process
 - Development and analysis of the impact model of the measure
2. Assumptions made regarding key framework data for the impact assessment
 - Determination of the framework assumptions of the data provided that are relevant for the impact assessment
3. Identify the objectives of the programme
 - Description of the requirements and expectations for the evaluation
 - Analysis of the top-down objectives for the funding programme based on public documents, guidelines and laws
 - Analysis of the bottom-up objectives of the funding programme based on ex-ante estimates and funding guidelines
 - Definition of the main areas of interest for the evaluation
4. Definition of indicators to measure the achievement of objectives
 - Selection of indicators that reflect progress in the relevant areas of evaluation (goal achievement, impact and efficiency monitoring)
 - Operationalization of indicators: Choice between qualitative/quantitative type, description and delimitation, calculation model, type of result, units (quantitative) or scales including interpretation rules (qualitative).
5. Data collection
 - Creation of a data collection concept based on the selection and structure of the indicators
 - Implementation of the data collection
6. Methodological procedure for determining the (gross) impact of the measure
 - Determination of the calculation methodology for determining the gross impact
 - Calculation of the gross effect
7. Methodological procedure for effect adjustment (net effect of the measure)
 - Determination of the methodology for effect adjustment
 - Carrying out the effect adjustment: Adjustment of the gross effect for deadweight, pull-forward, spill-over, lag, structural and rebound effects.

Box 2: Ex-ante evaluation recommended steps:

Step:


1. Evaluation criteria considered and indicators to be shown
 - Determination of the evaluation criteria to be taken into account: in addition to the energy and GHG savings effect, depending on the evaluation objective, consideration of other criteria such as economic effects (effect on energy costs, investments, employment, value creation), social effects or acceptance issues.
 - Definition of quantitative or qualitative indicators for the concrete evaluation of a measure: for the criterion of savings effects, these are the final, primary and GHG savings, whereby different calculation modes (new / added annual savings, added savings over a period) must be taken into account.
2. Assumptions made regarding key framework data for the impact assessment
 - Determination of the framework assumptions on energy prices, lifetimes and emission and primary energy factors relevant for the impact assessment.
 - Establish a reference (baseline) to ensure additionality of impact compared to the status quo.
3. Selected approach (static / dynamic)
 - Static view: Consideration only of the current decision-making situation for a measure.
 - Dynamic view: Consideration of the planning status of a measure (e.g. by updating financial resources or similar).
4. Methodological procedure for determining the (gross) impact of the measure
 - Determination of the impact model.
 - Determination of the calculation methodology for determining the gross impact (in each case depending on the specifics of a measure).
5. Methodological procedure for effect adjustment (net effect of the measure)
 - Effect adjustment at the level of an individual measure (net 1): Adjustment of the gross effect for deadweight, pull-forward, spill-over, lag, structural and rebound effects.
 - Effect adjustment at the level of a bundle of measures (net 2): Adjustment of the net 1 effect for interactions between the individual measures.
6. Dealing with uncertainties
 - Reduce methodological uncertainties by applying a consistent and comprehensive assessment methodology.
 - Transparent documentation of remaining uncertainties (e.g. due to regulatory uncertainties, price uncertainties, lack of acceptance, etc.).



Ariadne's thread through the energy transition: The Kopernikus project Ariadne leads the way in a joint learning process with representatives from politics, business and society, exploring options for shaping the energy transition and providing scientific guidance to policy makers along the pathway towards a climate-neutral Germany.

Follow Ariadne's thread:

 @AriadneProjekt

 Kopernikus-Projekt Ariadne

 Ariadneprojekt.de

More about the Kopernikus projects at kopernikus-projekte.de/en/

Who is Ariadne? In Greek mythology, Ariadne's thread enabled the legendary hero Theseus to safely navigate the labyrinth of the Minotaur. This is the guiding principle of the Ariadne energy transition project, in which a consortium of over 25 partners is providing guidance and orientation for shaping the energy transition through excellent research as a joint learning process between science, politics, business and society.

We are Ariadne:

adelphi | Brandenburg University of Technology Cottbus-Senftenberg (BTU) | German Institute for Economic Research (DIW) | German Aerospace Center (DLR) | Ecologic Institute | Research Institute for Sustainability Helmholtz Centre Potsdam (RIFS) | Fraunhofer Cluster of Excellence Integrated Energy Systems (CINES) | Friedrich-Alexander-Universität Erlangen-Nürnberg | Hertie School | ifo Institute – Leibniz Institute for Economic Research at the University of Munich | German Economic Institute | Julius-Maximilians-Universität Würzburg | Mercator Research Institute on Global Commons and Climate Change (MCC) | Öko-Institut | Potsdam Institute for Climate Impact Research (PIK) | Leibniz Institute for Economic Research | Stiftung Umweltenergierecht | German Institute for International and Security Affairs | Technische Universität Berlin | Technical University of Darmstadt | Technical University of Munich | University of Duisburg-Essen | University of Greifswald | Universität Hamburg | University of Potsdam | University of Stuttgart – Institute of Energy Economics and Rational Energy Use (IER) | ZEW – Leibniz Centre for European Economic Research