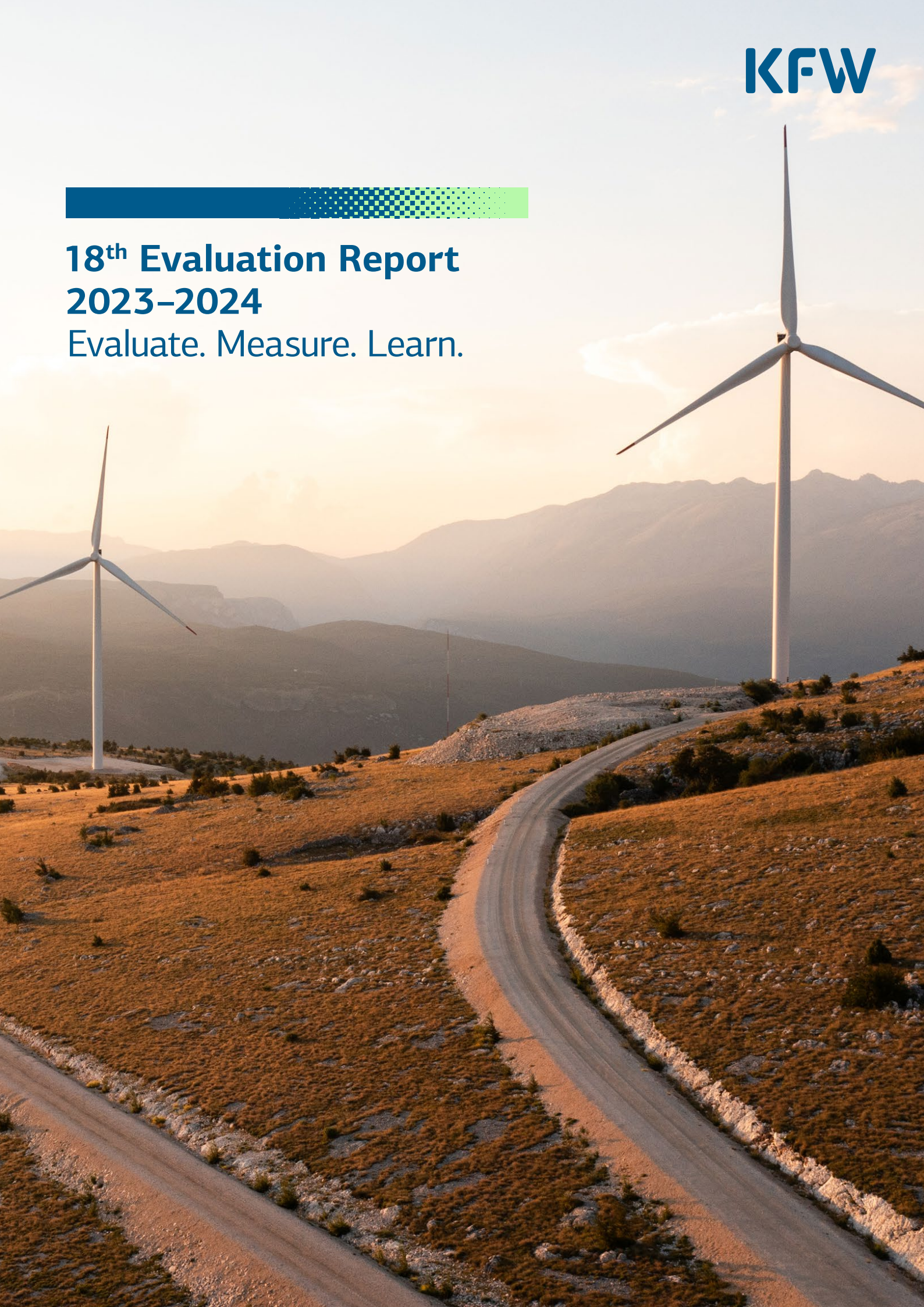
18th Evaluation Report 2023–2024

Evaluate. Measure. Learn.



Foreword

Christiane Laibach

KfW Executive Board Member, responsible for development cooperation and international financing.

Dear readers,

In current times, characterised by volatility and geopolitical uncertainties, we are facing major challenges both in Germany and in our partner countries. The federal elections and the recently concluded coalition negotiations have shown that public awareness of development cooperation is greater than ever and its benefits for Germany are being scrutinised. This is a trend that we can observe in our European partner countries and not least in the USA.

We want to bring facts and good arguments to bear on the increasingly critical public debate on the effectiveness of development cooperation. The biennial evaluation report has been published for more than three decades. In this time, the measurement of our impact and its transparent presentation in the public debate has become ever more important. In the current global situation, evaluations are an important key to strengthening confidence in our measures and underscore the significance of development cooperation.

“Evaluate. Measure. Learn.” is the motto of the evaluation report. The "Evaluate" chapter presents the work of the KfW Development Impact Lab, which was founded in 2023 - an instrument that we apply to focus even more strongly on the impact of our financing. In an overview of ongoing impact evaluations worldwide, many colleagues from KfW Development Bank have their say and share their experiences. There are also exciting insights into impact evaluations of projects from Vietnam, Tanzania and Madagascar. As a board member, I am delighted to see how much has been achieved here in such a short space of time.

Looking back at a long history: KfW Development Bank has been publishing its evaluation results since 1991.

In the "Measure" section, you will find a world map with the voices of operational colleagues who have conducted ex-post evaluations, as well as an overview of the most important ex-post evaluation results from the last two years. Our tool for public accountability and transparency. Particularly noteworthy are the insightful lessons learned from evaluations of projects in the financial, energy and education sector.

We want to keep improving. That is why the last chapter - "Learning" - shows us what we can learn from these results. Here, the focus is on an evaluation of the cooperation between technical and financial development cooperation. The results not only illustrate how well we are jointly advancing the projects in our partner countries, but also where there is still room for improvement. Finally, there is a particularly timely contribution: Jochen Kluve, Director of the Evaluation Department, and Estelle Raimondo from the World Bank discuss how artificial intelligence can support the evaluation of development cooperation - and the risks and opportunities associated with it.

I invite you to explore this report and discover the insights we have gained from our evaluations. Use the results about our impacts to inform discussions in your professional and private lives. I am convinced that each and every one of us can help shape the current debate.

Sincerely,

2025

Contents

Evaluation report 2023-2024: Evaluate

There from the beginning

The start of the project also marks the start of the impact evaluation: The project's impacts are analysed together with the partner.

Impact Evaluation
Champions worldwide

The number of impact evaluations of Financial Cooperation (FC) projects is growing steadily. These evaluations are usually carried out by academic partners in cooperation with KfW's evaluation department – and occasionally even completely independently by FC Evaluation. This ensures that the results are relevant not only for current and future KfW projects but well beyond. The involvement of our project partners, KfW employees abroad and KfW project managers is crucial for the success of the evaluations and often requires venturing into uncharted territory. Here, a few of KfW's "Impact Champions" share their experiences and motivations.

"The impact evaluation conducted by the KfW evaluation department, in collaboration with researchers from the University of Groningen, supported the project team from early on in the project cycle in developing impact hypotheses and intervention strategies. Additionally, the team helped develop a systematic and transparent strategy for selecting project participants. In Honduras, participants are now randomly selected from the official lists of returnee migrants and then invited to participate in the KfW project via phone by the Ministry of Social Affairs. We are still at the beginning of the evaluation, but the data-oriented support by the evaluation team has already provided exciting insights."

Elisabeth Hoffmann and Liliana Monteiro,
Occupational and social integration in Honduras, Guatemala, and El Salvador
with the University of Groningen

"It is exciting to learn more about the actual impacts of the project through the accompanying impact evaluation. Being able to assess the influence of the project on the health and education/employment of women and children is invaluable and further motivates us as a project team. The evaluation is also viewed very positively by our partners. The project benefits from household survey results during the implementation phase, allowing us to potentially use these findings for adaptive project management."

Freya von Negenborn,
Drinking water supply in Tanzania,
with KfW evaluation department

- Multisectoral UN-Projects / Fragility
- Education
- Finance
- Energy
- Agriculture & Natural Resource Protection
- Health
- Water


"The tracer and impact study assessed the short- to medium-term effects of our skills project on participants, particularly regarding their employment, employability, lifestyle, and psychosocial well-being. The study consisted of three rounds of data collection between 2022 and 2023, involving both participants and non-participants, utilizing a mix of quantitative and qualitative research methods. This was complemented by individual interviews, group discussions, and employer surveys.
The American Institutes for Research (AIR), UNICEF, KfW evaluation department, and the KfW project team closely and effectively collaborated in carrying out the study. The results were shared with Germany's Federal Ministry for Economic Cooperation and Development (BMZ) and disseminated on various platforms, marking an important milestone in impact evaluations within the context of non-formal vocational education."

Anne Wessendorf and Leonard Dlubatz,
Vocational Education in Lebanon,
with AIR

"I am excited that Igloo, as a rapidly growing insurance technology company, and BlueOrchard, as an impact fund manager, have, with great enthusiasm, committed themselves to a multi-year scientific collaboration. In the area of climate risk insurance, there is currently very little scientific evidence on development impacts. The study is likely to help us further improve approaches to climate risk financing, particularly under the Global Shield against Climate Risks."

Stefan Hirche,
Sustainable Economic Development in Vietnam,
with RWI Essen

"Can improved access to drinking water enhance social cohesion between refugees and the local population? We often state this in our Theories of Change without clear evidence. My colleague from FC Evaluation and I are testing this hypothesis in a data-driven impact evaluation in collaboration with Professor Anselm Hager from Humboldt University Berlin. This is no easy task in a fragile context, especially if academic standards are not to be comprised: there are conflicts, floods, and intelligence agencies worried about data collection. To achieve robust results despite these challenges, we got creative and have now successfully completed the first two baseline data collections. However, we will only have an answer to our evaluation question once the final data collection is completed in about four years."

Daniela Henrike Klau-Panhans,
Drinking water supply in Ethiopia, Somalia, and Sudan
with HU Berlin

"The impact evaluation enabled us to gain a more comprehensive understanding of the interactions between nature conservation and the socioeconomic conditions of neighboring communities, and even to quantify them. These insights are crucial for our work in project management, as the success of our measures largely depends on whether the local population supports conservation efforts. In all likelihood, the close collaboration with our local partners and the inclusion of all perspectives in the evaluation design have significantly enhanced the future relevance of the results."

Sebastian Manthey,
Natural Resource Protection in Madagascar,
with IRD

KfW Development Impact Lab
Measuring and understanding impact

The KfW Development Impact Lab as part of FC Evaluation has been managing the impact evaluations (IEs) of Financial Cooperation (FC) projects since 2023. The main task of the Lab is to answer important operational questions about project impacts using modern quantitative methods and data. The impact evaluations are often implemented in cooperation with the scientific community, where the Impact Lab acts as a bridge between the scientific and practical objectives of an IE.

Impact evaluations complement ex-post evaluations (EPEs) of KfW projects. EPEs assess the success of a project several years after the end of the project according to the six OECD-DAC criteria of relevance, coherence, efficiency, effectiveness, impact and sustainability (see section Measure on p. 18). Impact evaluations, in contrast, measure causal project effects (attribution) and ideally accompany an FC project from the beginning and throughout its entire duration. Due to the greater effort involved, IEs are specifically used where the demand for insights is particularly high (see world map of impact evaluation champions on p. 8).

The aim of every impact evaluation is to measure the impacts of a project causally. This means clearly attributing impacts to an FC project and ruling out the possibility that impacts would have occurred even without the FC project - for example due to economic growth or changes in climatic conditions. Methodologically, IEs compare the impacts actually achieved in the treatment group - i.e., the people, companies or regions targeted by the FC project - with those of a control group, i.e., similar people, companies or regions that are not part of the project ("no action" scenario).

The methodology is diverse: Household surveys (see example Vietnam on p. 16) as well as satellite or other secondary data (see example Madagascar on p. 14) are analyzed. The analyses not only provide quantitative answers as to the size of the causal effects but also why they have (not) occurred and which section of the population has benefited the most. To this end, the KfW Development Impact Lab develops the best possible methodological design for each IE of an FC project, which is adapted to the context of the project ("form follows function") and can support the project throughout its lifecycle. This ensures that findings from the IE can be useful for both current and future projects.

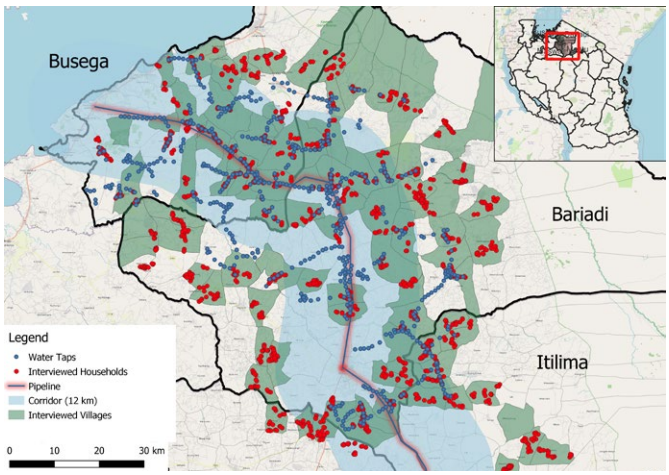
Impact evaluations utilise digital solutions for data collection - in the field with tablets or in the sky with satellites.

KfW Development Impact Lab: 2023-2024 in figures

Impact Evaluation
Initial results from the Simiyu Climate Resilience Project in Tanzania

The Simiyu Climate Resilience Project (SCRP) is jointly financed by the German government via KfW, the Green Climate Fund and the Tanzanian government. The project aims to strengthen the resilience of 495,000 people living in the rural, arid Simiyu region in northern Tanzania to the impacts of climate change. Activities include improving water supply and sanitation facilities, and promoting climate-smart agriculture.

At the heart of the project is the construction of a 100-kilometre-long water pipeline that will treat and transport water from Lake Victoria to remote areas and make it available to the local population - clean, safe and all year round. Expectations for the SCRP are high. Water must be delivered reliably to the people, who must then be able to use it and pay for it. The water pipeline is expected to reduce diseases and the burden on women and children, who currently have to walk long distances to fetch water, which is usually of poor quality.



The map shows the Simiyu project region, the planned pipeline and the public water pumps (blue dots), the villages randomly selected as part of the survey (green areas) and the locations of the households interviewed (red dots). Around two thirds of the villages are located within the supply corridor of the project (area shaded light blue). The remaining third of villages lies outside the corridor.

The impact evaluation is being carried out by the KfW Development Impact Lab of the KfW Development Bank, which has been part of the project since its inception. In 2024, before construction work began, a team of interviewers from the Tanzanian research partner ideas in Action (iAct) traveled to 90 randomly selected villages in the Simiyu region and interviewed 90 village leaders and surveyed around 1,800 households. The survey focused on individual water consumption and willingness to pay for clean water. In addition, water analysts used portable laboratories to take samples from 270 water sources to determine their chemical and physical parameters.

The results support the implementation of the project with information on consumption and the willingness to pay of future consumers, which is particularly relevant for setting an appropriate tariff for water, and thereby increasing revenues of water services to finance operation and maintenance of the water facilties.. These results allow the assessment of whether the project's forecasts are in line with the realities of the people. They show, for example, that households consume on average less than 20 liters per person per day for drinking, cooking and washing. Project plans often assume significantly higher values. The detailed data collection determines and visualizes decision-relevant parameters such as willingness to pay, disease incidence and water quality at village level. This allows, for example, the identification of locations in the project area that can be prioritized for connection to the water system in order to further increase the effectiveness of the project.

The results also confirm the relevance and potential impact of the project: The majority of village leaders surveyed cited inadequate water supply as a major obstacle to development. Water is extremely scarce in the region, especially during the eight month long dry period. As a result, local people spend a considerable amount of time fetching water: an average of three hours a day during the dry season. This is at the expense of schooling, since in more than a fifth of the households surveyed, children and young people - especially girls - are responsible for fetching water. One in four families report that one or more of their children regularly miss school because they have to fetch water.

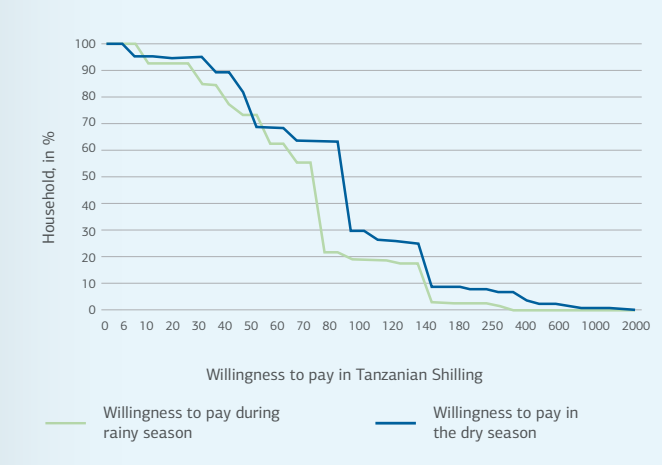
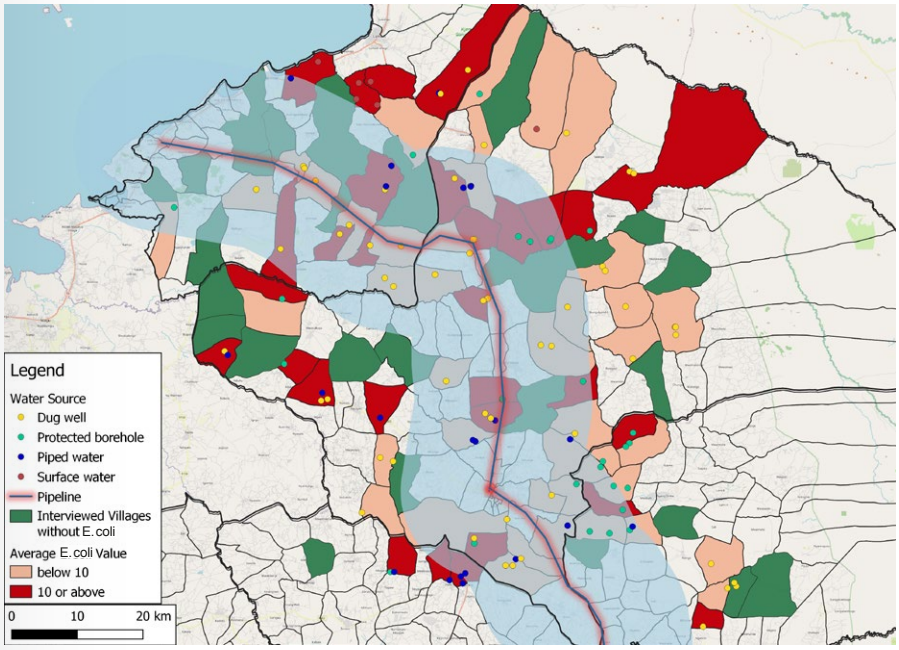


Illustration of willingness to pay for one 20-liter bucket of clean water in the Simiyu region.

Willingness to pay for clean water. One focus of the study was to determine the willingness to pay for clean water of the inhabitants of the Simiyu region. The results in the figure show that more than 90% of the respondents would be willing to pay for clean water. The graph shows the demand curves of households for a 20-liter bucket of safe water in the dry and rainy seasons. The downward trend of the curves shows that the proportion of households that would pay a certain price decreases as water prices increase. People are also prepared to pay more for water in the dry season than in the rainy season. This underlines the importance of water: Households are prepared to spend an average of 12% of total household expenditure on their need for clean water.

This 17-year-old girl in Kigoma, Tanzania, sets off to collect water: The walk to the lake takes 20 minutes there and 20 minutes back, and the water tastes salty.



Overview of contaminated water sources.

E. coli in excess of permitted levels: Water samples were taken from a total of 272 water sources regularly used by the people in the villages using mobile water laboratories. The figure shows those water sources in which the E. coli concentration in the sample is significantly above the national threshold value. 54% of the water sources show poor values. E. coli is a type of bacteria found in the intestines of humans and animals. Some strains of E. coli are harmless, while other pathogenic strains can cause disease. E. coli is often used as an indicator for water quality, as the presence of these bacteria can indicate possible fecal contamination. These results show that a piped water supply with safe, clean water will reduce diseases in the project region.

Impact Evaluation

Between biodiversity conservation and economic development in Madagascar

Madagascar is a global biodiversity hotspot. Yet, the country is one of the poorest countries in the world and ranked 177 out of 193 in 2024 according to the UNDP's Human Development Index; in addition the WHO forecasts a population growth of 70% by 2050. With high poverty levels and a growing population, it is not surprising that the demand for agricultural land is increasing, and exerting enormous pressure on the forests. Ecological protection and economic development are in a rivalry with each other here.

The impact evaluation examines tensions and synergies between environmental protection and the economic needs of local communities. The project-related research project BETSAKA (Biodiversity-Economic Tradeoff and Synergy Assessments for Conservation Areas) looks at the development of conservation forests in Madagascar from 2000 to 2024, which were supported by Financial Cooperation (FC) and the French development agency Agence Française de Développement (AFD). BETSAKA provides data-based answers to four hitherto little-studied questions that are of central importance to KfW as one of the world's largest financiers of nature conservation:

- Does KfW protected area financing reduce deforestation in Madagascar?
- Does the funding limit fires in and around these areas?
- Does it improve the living conditions of residents?
- To what extent are these effects influenced by financing, governance systems and management practices of protected areas?

These questions are being examined in a joint impact evaluation with several international partners: the evaluation department of KfW Development Bank, the evaluation department of AFD and the French National Research Institute for Sustainable Development (IRD). The University of Antananarivo, the Madagascar National Parks (MNP) and the Foundation for Protected Areas and Biodiversity (FAPBM) provide their expertise.

The impact evaluation uses complementing methodological approaches to holistically analyze the effects of protected areas on forest loss, forest fires and socio-economic conditions of neighboring communities. This includes the use of geo-referenced satellite data of protected areas to compare their actual forest area development with developments in identical, non-protected areas ("no action" scenario). In addition, socio-economic household surveys will be conducted to determine the impact of the protected area projects on the local population.

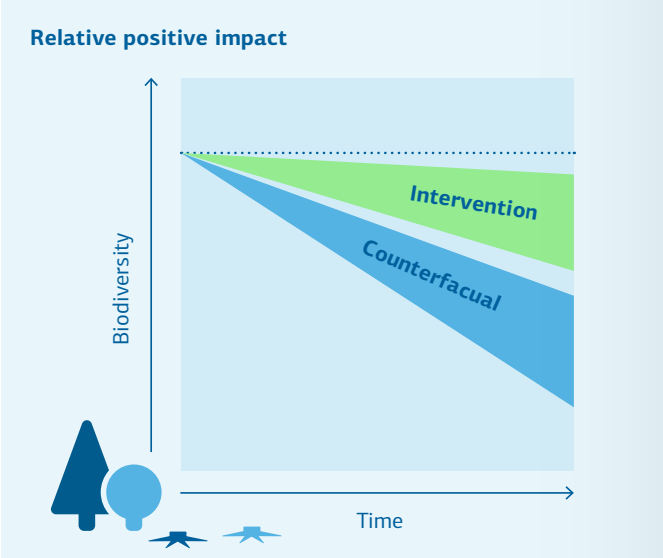


Illustration of the typical effect of protected area promotion on forest areas: significantly higher forest area/biodiversity compared to the "no action" scenario.

The two quantitative approaches are complemented by expert interviews to better assess the governance systems of the protected areas and their impacts in qualitative terms.

In addition to the impact measurement itself, **the results of the impact evaluation strengthen project implementation** by providing data and methodological tools for monitoring and evaluating KfW's involvement in Madagascar. For example, the BETSAKA team has obtained an official update of the protected area boundaries from the World Database on Protected Areas (WDPA), thus granting scientists and practitioners access to updated data. This will facilitate the planning of further projects for the Malagasy government and German development cooperation. Furthermore, the final results on the effectiveness and impacts of nature conservation policies feed not only into the planning of future projects in Madagascar with KfW, but also help to improve similar projects in other countries or with other donors.

Participants of a BETSAKA-workshop at the University of Antananarivo.

Ony Rakotoarisoa, Director General of Madagascar National Parcs (MNP):

"The BETSAKA project offers great added value in terms of the management effectiveness and efficiency of the protected areas managed by the MNP. It allows us to identify the key parameters that we need to pay attention to in our management so that the ecosystem services of our parks and reserves have a positive impact on the socio-economic conditions of local residents and contribute effectively to the conservation of biodiversity in Madagascar."

Antje Göllner-Scholz, Economic Cooperation (EC) Officer in Madagascar (BMZ):

"Madagascar's extraordinary flora and fauna are facing serious challenges due to the loss of forest areas and biodiversity. These losses not only have a local but also a global impact on the climate and jeopardize the resilience of the forests. German development cooperation is supporting Madagascar's national parks and their neighbors in mastering the balancing act between the protection of biodiversity and economic development and exploiting synergies. The BETSAKA evaluation plays a central role in verifying our successes, closing knowledge gaps and gaining valuable insights for the future."

Ingrid Dallmann, Impact Evaluation Manager, Agence Française de Développement (AFD):

"The cooperation in the BETSAKA project is based on the idea of mutual learning: it combines the expertise of biodiversity experts (MNP, FAPBM) with that of the academic world (IRD, University of Antananarivo, KfW, AFD). This teamwork enables us to carry out our assessment comprehensively, to strengthen the capacities of our partners - e.g., through exchanges on geo-spatial impact assessment or support for dissertations. It also allows us to speak with a common and therefore more influential voice. Working together in such an evaluation project is a unique experience that I can highly recommend to my colleagues at other institutions."

Impact Evaluation
A weather insurance scheme for farmers in Vietnam

Extreme weather events will occur more frequently and with greater intensity as a result of climate change. In Vietnam, small farms are increasingly exposed to heavy rainfall, tropical storms and droughts. Without effective adaptation measures, these can destroy livelihoods and increase poverty risk. Weather insurance is a promising tool for protection against weather risks.

One such insurance product was developed by Singaporean FinTech start-up Igloo and was funded by the German Federal Ministry for Economic Cooperation and Development (BMZ) via KfW Development Bank and BlueOrchard Finance, an impact investment firm. By subscribing to the insurance policy, small farms insure themselves against specific weather events. A subscription for 30 days costs 50,000 Vietnamese Dong (around 2 euros) per hectare of land. The insured rice farmers receive a payout within five days if heavy rainfall occurs. As there is no need for a time-consuming individual assessment of the damage, such index-based weather insurance can be offered at a significantly lower cost than conventional agricultural insurance products.

Despite the well designed adaptation instrument, the purchase of the weather insurance by smallholder farmers has been much lower than expected.

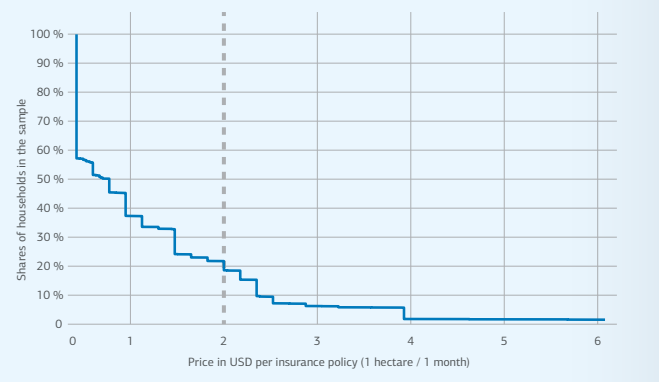
The impact evaluation accompanies the project and answers the following questions:

- How much are small agricultural farms willing to pay for index-based weather insurance?
- What factors explain the low demand so far?
- Are premium discounts and training effective channels for increasing demand for insurance?

The evaluation is being carried out by researchers from the RWI - Leibniz Institute for Economic Research and supported by DEval with funds from the Federal Ministry for economic cooperation and development (BMZ). An initial representative household survey of around 3,000 rice farmers in Vietnam's Mekong Delta determined, among other things, the willingness-to-pay of farmers for an index-based insurance against heavy rainfall.

Initial results of the study show that the willingness to pay is higher than expected. A significant proportion of the small businesses surveyed are interested in Igloo insurance: 56.8% stated a positive willingness to pay and 17.8% even stated a willingnesss to pay above the market price for the insurance - a surprising result.

Based on these survey results, it is possible to test two different activities - premium discounts and training - to increase the demand in randomly selected households. For this the households are split in two groups and the first group receives a premium discount: a voucher with which small farms can purchase Igloo's insurance against heavy rain at a reduced price. The level of discount is determined at random and is between 10% and 90% off the regular market price. The second group is offered training on climate resilience and financial literacy.



Willingness to pay for Igloo's insurance against heavy drought (per hectar, for 30 days), in Vietnamese Dong

How do the results of the impact evaluation strengthen project implementation? Insights into willingness to pay for weather insurance help project partners to identify barriers in demand and develop strategies to address these barriers. The evaluation tests the concrete costs and effectiveness of two potential strategies - premium discounts and training - directly with potential subscribers to the insurance in order to support the operational implementation of the FC project by Igloo. At the same time, the results can be applied to many weather insurance companies worldwide and can help support the implementation of weather insurances in other contexts.

Digital data collection in the field in Vietnam.

Vietnam: A farmer pulls out notes during an interview.

"The impact evaluation of the KfW project could potentially be highly valuable to Igloo. It enables Igloo to showcase the value of weather index insurance for farmers, which in turn could help Igloo attract support from the Government for this program and to scale up to reach millions of farmers."

Mahesh Joshi, BlueOrchard Finance Ltd, Head of Asia Private Equity Investment

The evaluation team

Prof. Dr. Kati Krähnert, RWI and Ruhr-University Bochum

"Our evaluation project is a good example of how rigorous scientific methods can provide policy makers with a more informed basis for decision-making."

Dr. Lukas Mogge, RWI

"Our research shows which factors motivate small farms to implement climate adaptation measures. Through this, we can help to overcome obstacles to climate adaptation."

Dr. Emmanuel Nshakira Rukundo, RWI

"It is exciting that our analyses contribute something new to the state of scientific research. So far, the willingness to pay for weather insurance has not been well understood."

Evaluation report 2023-2024: Measure

Independent review

FC Evaluation scrutinises around half of all projects following their completion.

Ex-post evaluation
Champions worldwide

The evaluated Financial Cooperation (FC) projects are spread all over the world. KfW Development Bank employees carry out evaluations of projects on behalf of the independent evaluation department - of course only if they have not previously been involved in that project. The aim is to think outside the box and learn from evaluations for current and future projects. Some of them share their experiences here.

"During the evaluation, I gained fascinating insights into the complex political and societal challenges in Bosnia and Herzegovina. I was particularly impressed by the high level of professionalism demonstrated by the local partners in addressing the aftermath of the catastrophic floods of 2014. The flood relief project was implemented swiftly and effectively in both the Federation of Bosnia and Herzegovina and the Serb Republic, without political interference. I have carried out several evaluations and have always learned a great deal from the process. Therefore, I highly recommend that all colleagues in Financial Cooperation consider conducting an evaluation."

Dr. Matthias Schlund,
Recovery from Flood Damage in Bosnia and Herzegovina

"I was particularly impressed by the gender-transformative design of the project, which emphasized the sexual self-determination of women and the responsibility of male partners in its communication campaigns. The Project-Executing Agency also prioritized continuous impact measurement and learning. Thanks to the multitude of accompanying studies conducted as part of the project, we were able to base the ex-post evaluation on a solid foundation of high-quality data."

Hannah Linnemann,
Family Planning and HIV Prevention in Côte d'Ivoire

"During the evaluation of the biodiversity project, I was visiting a country where the overall conditions had unfortunately deteriorated since the project's implementation (the economy, the impact of COVID-19, and drug-related crime). Against this backdrop, I was pleasantly surprised by the effectiveness with which the structures established at that time continue to be utilized and the way they still contribute to the project's intended conservation goals, despite the limited resources available."

Ingo Baum,
Natural Resource Protection in Ecuador

- Multisectoral UN-Projects / Fragility
- Education
- Finance
- Energy
- Agriculture and Natural Resource Protection
- Health
- Water
- Decentralization and Governance
- Other

“It was encouraging to see that even after the project had been completed, the partner still remained committed to the project's core concept and intended to expand it further. The project faced significant delays during implementation due to factors such as flooding, cyberattacks, and issues with customs. We were able to learn a great deal for future projects through these experiences. I found the conflict surrounding Serbia's energy supply particularly interesting: despite the expansion of renewable energy sources, coal is likely to remain a significant energy source for a long time. Therefore, it is essential to make the use of coal as efficient as possible.”

Aaron Müller,
Regulation of Coal Energy in Serbia

“I was particularly impressed by the deep emotions and gratitude expressed in conversations with the beneficiaries. Investments in agroforestry measures have enabled indigenous Indian families in rural areas to reshape their lives and to offer new prospects to future generations. This evaluation experience has strengthened my conviction that we can achieve great things with our projects.”

Theresa Schneider,
Promotion of Agriculture in India

“We could confirm the flagship character of the project, even from our critical perspective. The vocational training institutes are excellently equipped, and the dedicated teaching and management staff have successfully integrated practical content into the curriculum. In addition, close cooperation with the private sector has been established, in-house training for teaching staff promoted and trainees given easier access to local companies through internships. The high employment rate of over 80% of graduates speaks for itself. The joint evaluation gave us great joy and facilitated an interesting professional exchange. In our view, our diverse experiences and assessments significantly enhanced the quality of the evaluation.”

Annelyse Umunyana and Nicole Turad,
Technical and Vocatinal Education in Vietnam

“The repeated expressions of appreciation from various beneficiaries highlighted the significant relevance of this food security initiative. This particularly highlighted how crucial access to basic supplies is in times of crisis. By issuing food vouchers, it was possible to make a quick, uncomplicated and effective contribution to food security - a decisive success factor in the exceptional situation caused by the pandemic. Thanks to the tandem evaluation approach, we were able to discuss our impressions and perspectives time and again, allowing us to identify aspects together that we might not have noticed individually.”

Maren Linda Verspohl and Marc-André Hensel,
COVID-19 Emergency Assistance for Food Security in South Africa

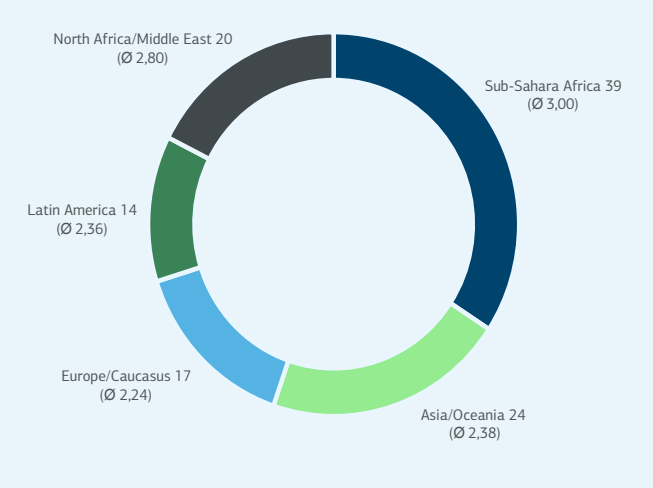
Ex-post evaluations
A stroll through sectors and regions

How was the group of evaluated projects in 2023 and 2024 composed, and what are the key evaluation findings? Below is an overview of the evaluated projects by sector and regions, their ratings and the Lessons Learned that could be derived from the evaluations. The individual evaluation reports are available in the *IDEaL* app.

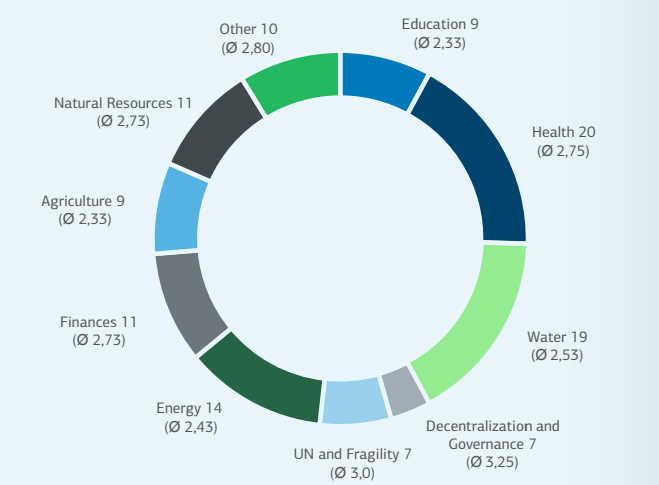
In the period 2023-2024, 114 Financial Cooperation (FC) projects were evaluated, selected as a random sample from all completed FC projects and representative of the FC portfolio. A further 26 FC projects were included in these 114 ex-post evaluation reports, as they were closely linked to the projects in the sample (see chapter Methodology on p. 44).

The largest share of the 114 evaluated projects was in Sub-Sahara Africa (39), followed by Asia/Oceania (24) and North Africa/Middle East (20). The Europe/Caucasus (17) and Latin America (14) regions accounted for slightly fewer projects. Overall, the regional distribution of evaluations shows continuity over time: The African continent and the Middle East already accounted for just over half of the projects in previous years. From a sectoral perspective, the health, water and energy sectors are the sectors with the largest number of evaluated projects.

On the standardized OECD DAC rating scale of 1 (very successful) to 6 (highly unsuccessful) used in ex-post evaluations, projects from Europe/Caucasus have the best average rating (Ø 2.24), closely followed by Latin America (Ø 2.36) and Asia/Oceania (Ø 2.38). The Sub-Sahara Africa region has the lowest average rating in a regional comparison, but is still successful overall with a rating of 3.0.



Number of evaluated projects and average ratings by region.



Number of evaluated projects and average ratings by sector.

From a sectoral perspective, the education, agriculture (Ø 2.33 each) and energy (Ø 2.43) sectors performed particularly well, while the lowest average scores were achieved in the decentralization and governance (Ø 3.25) as well as UN and fragility (Ø 3.0) sectors.

Projects in the education sector achieve above-average results. Here: Pupils at the technical vocational school in Porto Novo, Benin, are trained in land surveying and use laboratory and workshop equipment funded by FC.

Natural resource protection often finds creative solutions, but also faces challenges

Natural resource conservation projects received an average rating of 2.73. The projects focused on the promotion of protected areas: Nine of the eleven projects were intended to strengthen selected protected areas in countries in Africa, Latin America and Asia. Common measures included the creation of protected area management plans, the construction of control posts, administrative buildings or tourism infrastructure as well as the financing of training and further education and the promotion of the population living in or around the borders of protected areas. Some projects provided national nature conversation funds with capital.

The evaluated projects were highly relevant, as the selected target regions were often biodiversity hotspots and the planned measures were well suited to addressing the prevailing core problems.

Many of the supported protected areas are located in low-income countries and are chronically underfunded. In these contexts, the commitment of FC and other donors is essential in order to achieve effective operation and thereby the desired protective effects. When, in the course of the COVID pandemic, subsidies from public budget funds were cut and other income sources, such as tourism, completely collapsed, FC ensured ongoing operations through its support and thereby made a significant contribution to stabilization. The evaluation results also show that the problem of inadequate basic financing can be significantly and sustainably reduced by the capital income from nature conservation funds. The fund assets are invested on the capital market and the income generated can be used for the operating costs of the protected areas. In three evaluations (Madagascar, *Ecuador* and *Côte d'Ivoire*), this approach proved to be a key success factor.

Nature conservation concepts can only be sustainably anchored if they are implemented in cooperation with the local population, and take their interests into account. Many of the evaluated projects promoted greater integration of the local population into park management (*Brazil*, *Ecuador* and *Namibia*) and financed measures to support the local population. However, the evaluation results show that such measures often only account for a small proportion of the overall budget (*Côte d'Ivoire* and *CCAD*) and that necessary budget reductions during the implementation phase can be at the expense of support for local residents (*Madagascar* and *Ecuador*).

Forest reserve near Khovaling, Tajikistan: this farmer and his wife have grown over 1,000 trees and around 5,000 seedlings with the support of a FC development program. They used the yields to plant wheat and set up a tree nursery.

One project in the *South Caucasus* stands out particularly positively. In this project, nature conservation agreements were concluded with selected municipalities located in important eco-corridors of natural areas worth of protection. The agreements stipulate land and resource conservation measures, for the implementation of which the affected communities receive annual compensation payments from a fund. The payments have led to a noticeable improvement in income and living conditions, while at the same time increasing wildlife populations. The project is a prime example of a holistic and highly participatory approach.

One challenge for the implementation and evaluation of the projects was the lack of data on the status of the protected areas, their biodiversity and species diversity. The diversity, dynamics and complexity of the various ecosystems often make it difficult to develop standardized indicators and measurement methods: However, thanks to the use of modern satellite data, it is now increasingly possible to record and evaluate longer impact periods (*Madagascar*, *Ecuador*, *Côte d'Ivoire*, *Brazil* and *CCAD*).

Market women sell agricultural produce at a town market in Zimba, Zambia.

The top sector: Reasons for the above-average performance of the agricultural sector

The above-average performance of the agricultural sector is due in particular to three projects, each of which achieved the highest rating of 1 "very successful" in the evaluation (two projects in *India* and one in *Ecuador*). All three projects generated high increases in agricultural productivity and income. This in turn reduced the previously forced migration of labor to nearby cities due to poverty and hardship. In India, the projects were so successful that they were replicated nationwide.

The key success factors of the projects included the close alignment of the projects with the needs and capabilities of the family farms, a high level of motivation of the participants and a holistic approach. The projects not only supported irrigation infrastructure and its maintenance, but also the production and marketing of agricultural products. In *Ecuador*, the proximity of the project to regional economic centers played a favorable role, as the agricultural products can be marketed there. However, external factors can occasionally have a negative influence on evaluation results: A project to modernize irrigation perimeters in *Tunisia* did not have much impact due to the drastic deterioration in water availability and was rated 4 ("moderately unsuccessful").

In general, the findings of the evaluations point to two central areas of tension in the sector. On the one hand, conflicts between efficiency and sustainability can arise if the introduction of agricultural resource-saving technologies is associated with maintenance costs that cannot be met by the target group without long-term technical support. On the other hand, the consequences of climate change are ambivalent: irrigation approaches and integrated water resource conservation can demonstrably help with adaptation to the consequences of climate change (*Peru*, *Ecuador*, *India* and *Niger*). However, the increase in extreme weather events can also impair the effectiveness and sustainability of such projects, as in *Tunisia*.

Challenges in conflict-affected contexts: Results of the UN and fragility sector

Evaluation results of UN projects that are implemented in highly fragile contexts point to the structural challenges in this sector and at the same time underline how relevant the implementation of projects in difficult contexts via UN organizations is. For example, two thirds of the evaluated projects are still rated as successful overall, despite the difficult conditions.

A key structural challenge for these projects is the need to create long-term impact, even though they are primarily focused on addressing the immediate needs of affected populations. In extremely conflict-affected areas, where living conditions are fragile and volatile, there is an urgent demand for rapid-response measures that are often not designed for long-term impact from the outset (e.g., *Ethiopia*, *Uganda*, *Syria* and *South Sudan*). Examples of such measures include cash-for-work programs, which provide temporary employment in local infrastructure projects, allowing individuals to earn a short-term income. While the broader impacts may be less pronounced, evaluation results clearly indicate that UN projects are effective in achieving significant short-term outcomes (e.g., in *Syria*).

One particularly encouraging finding from the evaluations is that the implementation of the projects through UN organizations was an important factor for efficient and effective implementation, even in difficult intervention contexts (e.g., *Ethiopia*, *Uganda*, *Syria*, *South Sudan*, *Sudan* and *Jordan*).

The young social activists inform new mothers about immunization, the importance of breast milk and nutrition in a shelter for internally displaced persons in Dollow, Somalia.

The success rate 2023-2024
85 Percent

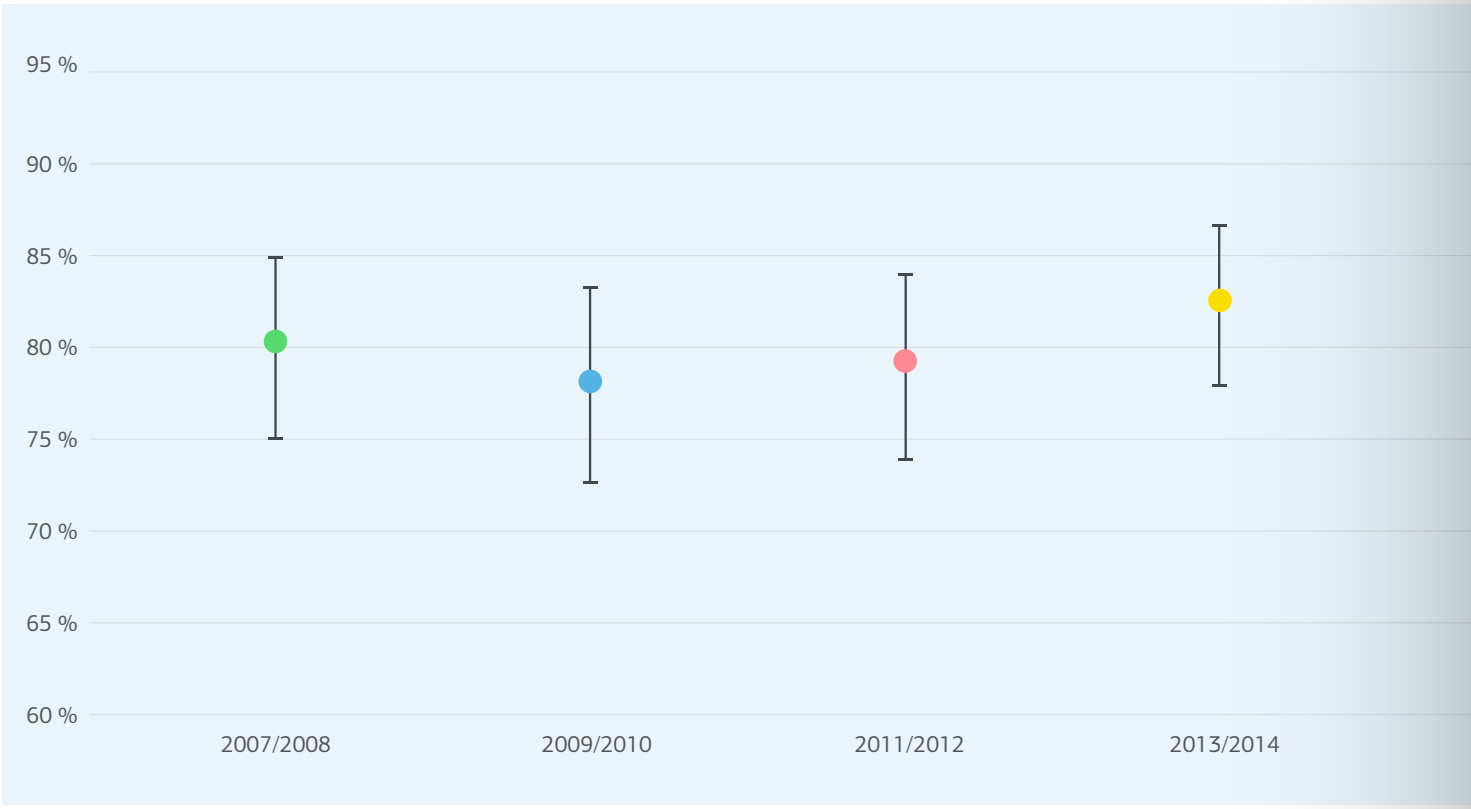
The success rate of Financial Cooperation projects is a key figure, as it provides at a glance the share of successfull projects over a two-year evaluation period. Notably, this success rate applies - within a certain margin of estimation accuracy - to all projects implemented by FC during that timeframe (see chapter Methodology, p. 44).

The average rating of the 65 FC projects evaluated from the 2023/24 sample is 2.71. The distribution of the ratings shows that just about half (46%) of the 65 projects achieved a rating of 2 (successful) and more than a third (38%) achieved an overall rating of 3 (moderately successful). A total of nine projects were rated 4 (moderately unsuccessful) and one project was rated 5 (unsuccessful).

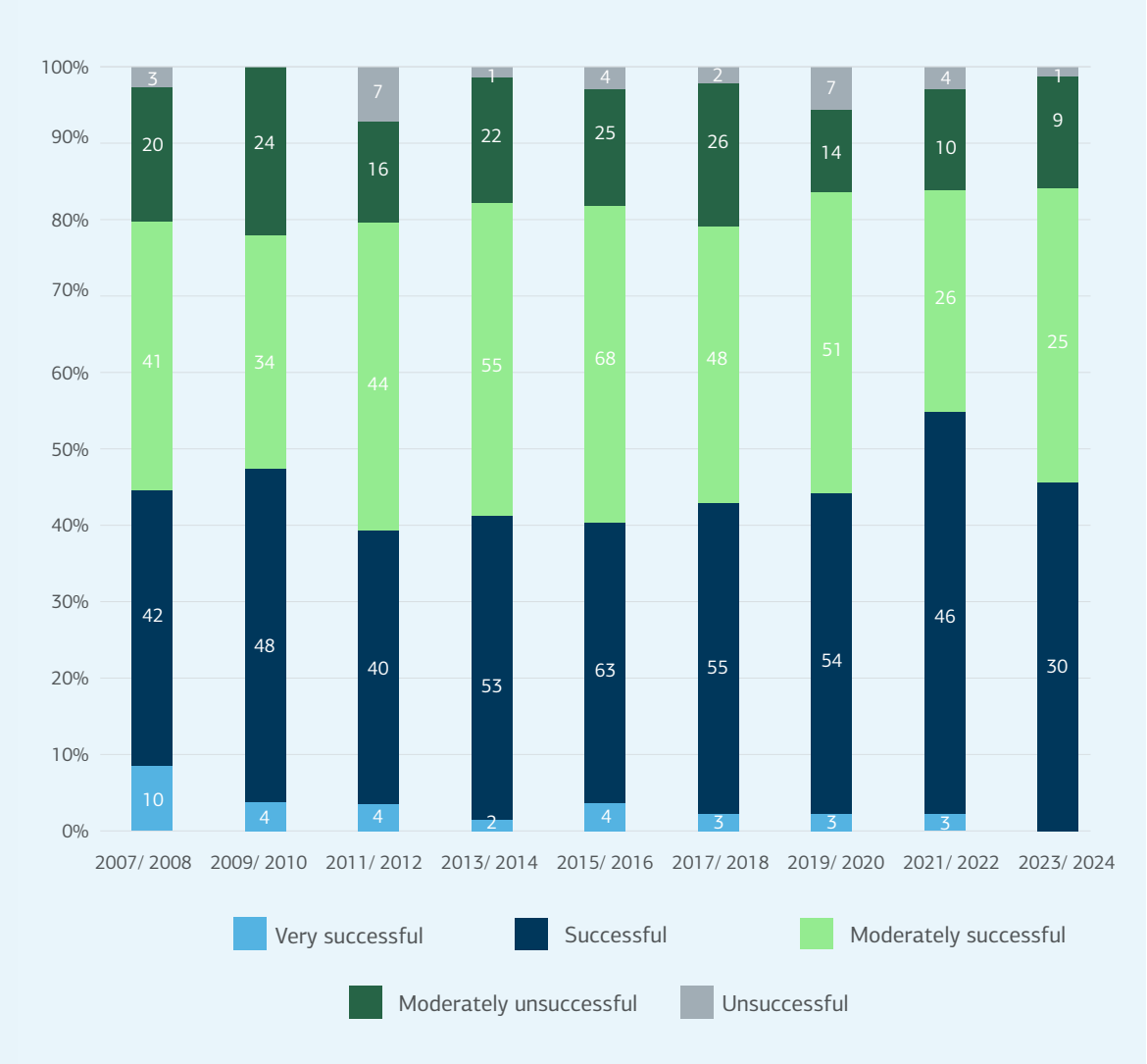
The success rate, i.e., the proportion of projects rated at least 3 (moderately successful), is 85%. The current success rate is therefore above the long-term average success rate of 81% from 2007-2024 and is in line with the slightly positive trend over the past 18 years.

Thanks to the representative sampling, this success rate of 85% can also be applied to non-evaluated projects and thus to all 268 FC projects from the evaluation period. However, there is a certain statistical uncertainty here, which is expressed by the confidence interval: The success rate of all (evaluated and non-evaluated) 268 projects in the 2023-2024 reporting period is between 77% and 92% with 95% certainty.

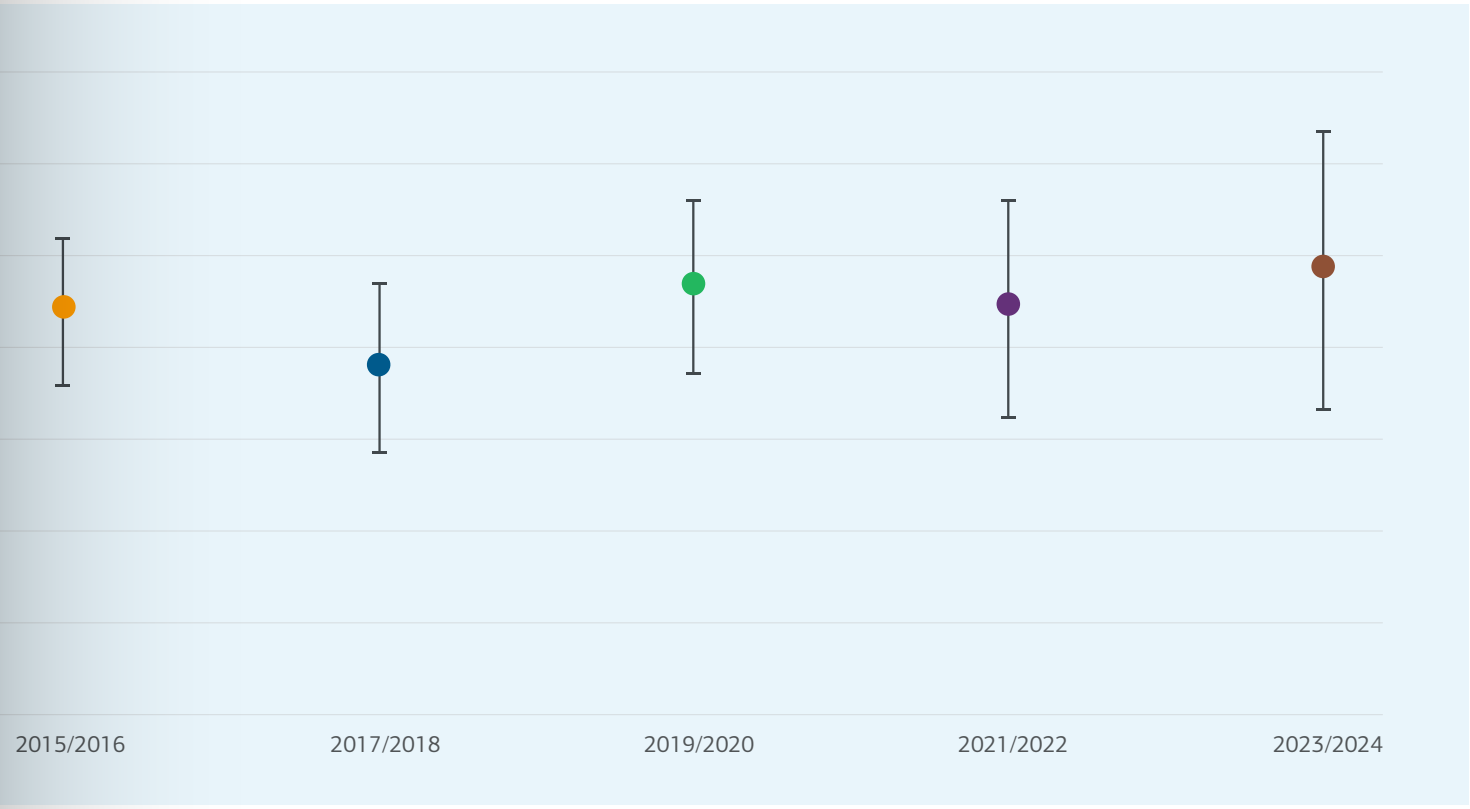
The confidence interval for 2023/2024 is still comparatively broad, as only 50% of the projects from the 2023-2024 sample have been evaluated to date (see chapter Methodology, p. 44). The estimation accuracy will therefore increase with each additional evaluation carried out until the entire sample has been evaluated. In this reporting period, 36 evaluations from the previous sample were completed for the 2021-2022 population, which narrowed the confidence interval by 2.5 percentage points (i.e., the estimation accuracy increased). The success rate for the 2021/22 reporting period is therefore 82% (84% in the 2021/22 evaluation report).



Success rate of FC projects, i.e., proportion of projects rated at least 3 overall, and 95% confidence interval, for two-year samples from 2007/08 - 2023/24.



Distribution of project ratings in two-year samples from 2007/08 - 2023/24.



Three sectors in the spotlight
Lessons Learned from years of ex-post evaluations

Lessons Learned from the education sector

A look at the Lessons Learned from ex-post evaluations over the last five years shows that the focus of educational projects has shifted from basic education to vocational training. Nevertheless, there are still comprehensive programmes to ensure primary education in fragile contexts, often in cooperation with UN organizations.

In *Vietnam*, an up-and-coming developing country, the importance of close cooperation with industry in vocational training has once again been demonstrated. The project, which was rated as successful, not only supported the transition of graduates into employment, but also ensured that the training and equipment of the educational institutions met the needs of industry. Investments in modern hardware and in the development of curricula and teacher training served this purpose.

Vocational training approaches face particular challenges in fragile contexts. When the prospects on the labour market are poor due to low labour demand and low wages, the value of education is not always immediately apparent. This can lead to frustration among young people. Against this backdrop, the evaluations have shown that positive effects on resilience and employability can nevertheless be achieved.

Commitment at school level, ensuring the financing of ongoing school costs and the institutionalization of maintenance of school infrastructure should be a central component of any project. At the same time, it is important to adapt the standard designs of educational projects to local conditions. In remote areas without electricity and water connections, for example, electrical installations and water closets cannot be used. In *Senegal*, the construction of walls around schools has been shown to reduce vandalism and protect against use by unauthorized third parties.

In general, it is evident that the educational landscape can be sustainably improved with commitment and coherent cooperation. The commitment of school management, teachers and parents remains a key success factor here. In *Tajikistan*, the municipalities were successfully involved in the school organization, which led to better support in the operation and maintenance of school infrastructure despite limited financial resources.

214 pupils attend a new FC-financed school (grades 1-11) in Danghara District, Tajikistan.

Lessons learned from the financial sector

Ex-post evaluations of financial sector projects in recent years show two important trends: Firstly, there is a shift from projects with an environmental focus, such as environmental credit lines, to projects that concentrate on improving energy efficiency (e.g., energy efficiency credit lines). Secondly, ever more projects are structured as funds.

The evaluations show: A decisive success factor for funds is an experienced fund manager with extensive market knowledge and a good network of qualified experts in the region. It is also important that the fund manager has an internal impact and sustainability team and sufficient time to actively support clients. This was demonstrated, for example, in the evaluation of the Microfinance Initiative for *Sub-Sahara Africa*, where the fund manager's good expertise and network were decisive for the high quality of the fund portfolio.

Projects in very poor or possibly conflict-ridden areas face special challenges. On the other hand, projects in which the financing was adapted to the individual needs of the borrowers were especially successful. For example, in a project borrowers were allowed various types of collateral (such as cars, gold or group collateral from women's cooperatives) and flexible repayment periods. The latter is important for agricultural loans, for example, where repayment ability is based on harvest times.

The evaluations moreover show that in emergency situations, e.g., after a natural disaster, preexistent structures should be used to implement projects. Apart from other advantages, this can be critical to scaling support quickly. A successful example of this is a project in the aftermath of a flood in *Bosnia and Herzegovina*, where more than 2,000 loans were disbursed to private households and small and medium-sized enterprises in just a few weeks thanks to the use of preexisting structures.

Documents and savings book of a savings group in Afar, Ethiopia. The women's group uses the FC-supported loans, which were granted via the Bank of Ethiopia, to breed cows and goats and to produce meat.

Lessons Learned from the energy sector

In terms of financing volume, the energy sector is the largest sector at KfW Development Bank. It consists of the following areas: Credit lines, renewable energies, electricity transmission and energy efficiency. Both fundamental and sub-sector-specific Lessons Learned can be derived from ex-post evaluations of the last 10 years.

To enable partner countries to invest in renewable energy, FC has in recent years increasingly provided credit lines. Credit lines are an important instrument for increasing the use of innovative energy technologies while contributing to the development of national financial markets. As national capital markets often only provide loans with short terms and high interest rates, more attractive (low-interest, long-term) loans can successfully increase investment in renewable energies. The basic prerequisite for a successful implementation: countries and financial intermediaries that have considerable experience with the implementation of financial products in the energy sector.

Another Lesson Learned: It is crucial for the success of the project to diversify the mix of energy sources, i.e., to use different technologies to prepare for the challenges of climate change. For example, there is a high degree of complementarity between the energy sources wind, sun and water and it therefore makes sense in the long term to promote them in parallel within a country.

The evaluation results around renewable energies show that projects with private sector involvement implemented in countries without state subsidies or guaranteed feed-in tariffs generally only come to fruition if larger financial volumes are in place. Otherwise, it may not be worthwhile for private companies to participate in a construction project, for example. Guaranteed feed-in tariffs can help reduce investment risks. Evaluations in the field of electricity transmission also show that an expansion of electricity grids is only effective if the generation and transmission capacities are expanded simultaneously, and vice versa.

We also learn that the involvement of homeowners and tenants is crucial for the success of renovation projects to increase the energy efficiency of buildings. If people are sufficiently involved, the CO_2 savings achieved thanks to the project deviate less from the (previously modelled) planned CO_2 savings. The expected magnitude of effects hence only occurs because behavioural assumptions in planning match with actual behaviour of energy users. Furthermore, ex-post evaluations show that residents use energy more efficiently if their energy bills are calculated based on actual consumption. For the housing market, the evaluations show that the use of clear energy efficiency criteria and national building standards increases the attractiveness of buildings for buyers.

The engineer gives a tour of the power plant in Puerto Villamil, Galapagos Islands, Ecuador. Around half of the plant (800KW) is powered by solar energy and the other half by diesel generators.

Substation in Gaita, Tanzania. This farmer also works part-time as a security guard for the plant. She is the mother of two children. The electricity system supplies three regions: from the Kidatu and Mtera power plants via the Buly substation towards Nyakanazi.

Evaluation report 2023-2024: Learn

Learning from experience

Promoting learning from completed projects is the core of our work.

Artificial Intelligence
Potentials and perils for the generation of evidence

Generative artificial intelligence (AI) keeps making headlines and has experienced breakneck advancements since the public launch of ChatGPT 3.5 in November 2022. The remarkable surge in AI capabilities and its usage has attracted significant investments in AI technologies, further enhancing the sophistication and applicability of AI across sectors.

With the promise of boosting productivity and efficiency, AI is transforming the workplace by reshaping job roles and redefining skill requirements. This transformation, of course, also extends to the field of evaluation. AI can potentially enable evaluators to deliver quicker, better or more impactful project evaluations, but also reshape the role and tasks of evaluators. When exploring the potentials and perils of AI for the generation and communication of evidence from evaluations, several critical questions arise.

Meet the experts

Estelle Raimondo is the Head of Methods in the World Bank's Independent Evaluation Group. At the World Bank, she advises and trains teams on a wide range of evaluation methodologies, drives innovation, and leads the integration of data science and artificial intelligence into evaluation practice. With over 15 years of experience in evaluation and evaluation research, her work has been published in international peer-reviewed journals and blogs. She is the author of the book *Dealing with Complexity in Evaluation.*

Jochen Kluve is Director of Evaluation at KfW Development Bank and Professor of Economics at Humboldt University Berlin. With 25 years of experience in impact evaluation and economic policy analysis, he has advised and collaborated with a multitude of international development organizations, governments, and the EU. At KfW, he leads a team of evaluators that has brought about the *KfW Development Impact Lab* and that focuses on useful and useable knowledge products, such as IDEaL – Interactive Database for Evaluation and Learning.

Estelle Raimondo and Jochen Kluve in conversation in Washington DC, USA.

Estelle, as the Head of Methods at the Independent Evaluation Group (IEG) of the World Bank Group, you possess a comprehensive understanding of the methodologies applied in evaluations. Over the past 5 to 10 years, what significant trends have you observed?

"Before distilling bigger trends, I would like to emphasize that my vantage point is that of someone working in evaluations in the institutional realm of international organizations. Overall, I identify three such big trends:

The first trend is the ability **to broaden the range of methodologies and approaches** that we have been using in evaluation to capture impact. In 2012 there was a publication by Elliot Stern and others, which was really critical in helping us reach different disciplines, different fields, different paradigms and tap into that to expand our toolbox quite significantly. For example, qualitative comparative analysis, process tracing, and contribution analysis have seen a kind of rebirth. I think we are now very much open to a lot of different approaches and trying to really understand which one is fit for purpose depending on the questions and context and depending on how much complexity we want to bring into the fold. Tom Aston and Marina Apgar have coined the term Bricolage to capture this moment, which is the idea of being practical and pragmatic in using and combining the tools as needed to approach that complexity.

Second, I do feel like we are much more able to **deal with complexity from an empirical point of view.** We have done a lot of work in bringing complexity theory and systems thinking in evaluation and also trying to have the right empirical toolbox, but where I see an issue is that there is not much appetite on the user end for that.

We are facing a new paradox, where for a while there was the intuition that we were not able to sufficiently deal with the complexity of development with our conventional methods. Now we are much better equipped to do so, but on the receiving end there is no appetite, so we see a bit of a return or the pendulum, **going back to simple tools** like scorecards, results-based management approaches, indicator frameworks and RCTs that tend to oversimplify a complex reality.

Third, of course the **data revolution**. In the 2010s it was called big data. Now we are thinking more in terms of Artificial intelligence, but the common thread is really about being able to harness data that we couldn't tap into before in the evaluation space. At the Independent Evaluation Group, we have moved full speed on two sets of data, image as data and text as data. On the image side, the evaluation field is finally starting to catch-up on its use of geospatial analysis. On the text side, we have now a few years of experience with machine learning and making progress in experimenting with generative AI. In the end, evaluation is at its core an interdisciplinary field and harnessing methods and approaches from various disciplines is what makes us progress."

Jochen, as the head of the evaluation department at KfW Development Bank, do you find this perspective aligns with your experiences?

"I think it resonates very well with our thinking and experiences. Our work fully aligns with the Bricolage-Approach Estelle mentioned: over the past five years we have been actively using the impact evaluation toolbox with a pragmatic and practical mindset; in our team we like to say 'form follows function'.

To harvest the fruits of the data revolution mentioned by Estelle, we founded the KfW Development Impact Lab in 2023 that hosts and streamlines, besides impact evaluations, data science activities, leveraging a variety of data sources, including satellite imagery and high-frequency traffic data. The limited appetite for complex evaluation results is also something we are confronted with. In the last 5 to 10 years, we have therefore broadened or, better said, **recalibrated our focus from 'evaluation for accountability' to providing more practical and usable tools for our users.** I am curious to see how AI will further evolve as a tool to help us make that appetite larger."

Estelle, AI can be a tool to simplify the analysis of evidence. Would you agree?

"Absolutely. Within-document summarization is a clear comparative advantage of genAI and large language models and we are testing its performance on cross-document summarization (or synthesis). In addition, these types of models can be very helpful as a 'writing coach'—enabling evaluators to write for multiple audiences at lower costs—especially in the anglophone space. But let me add one thought.

There is some incipient evidence on what AI can do to our critical thinking, especially if we fall into the overreliance on AI. This can also abate our critical thinking and our ability to deal with, absorb and make sense of complexity. Hence, I think there is potential, but there is also a peril there."

Jochen, before we talk in more detail about AI, let's get on the same page: What is AI and what is the difference of the "new" AI tools compared to the past?

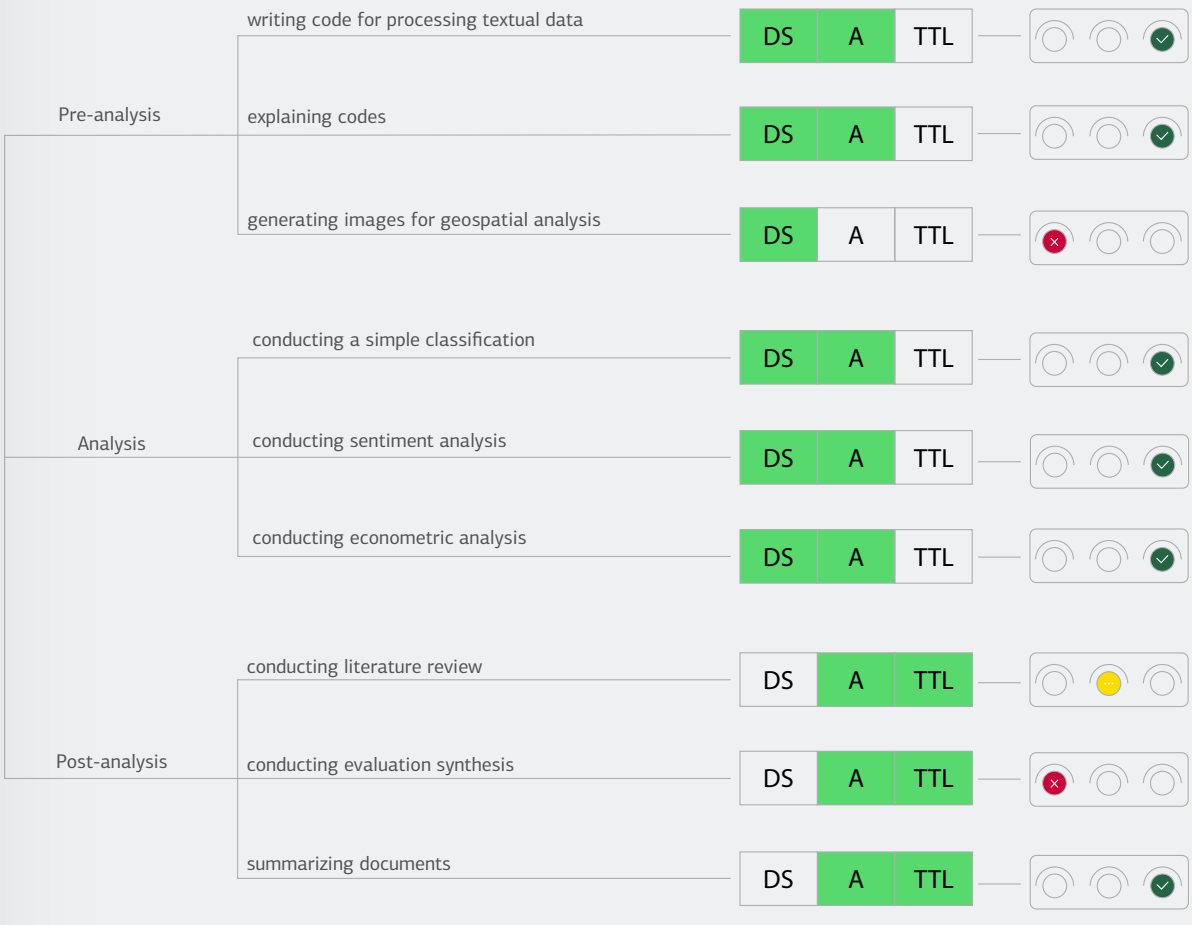
"Well, if you look up AI in a dictionary it is defined as a computer system that performs or imitates tasks that typically require human intelligence. This distinction between performance and imitation of tasks is exactly the difference between discriminative and generative AI: Before 2022, only discriminative AI was available, namely machine learning or supervised learning models that were able to reproduce decisions on which it had been trained. With the launch of ChatGPT we saw a shift towards generative AI, meaning the creation of new content."

Estelle, some of our team members listened to a very interesting talk of yours at the European Evaluation Society Conference. You mentioned a set of experiments your team has been running on AI uses for various evaluation tasks. Can you tell our readers about it?

"Sure. **We conducted our first batch of experiments in spring 2023, testing both the World Bank's enterprise version of ChatGPT and other open-source models.** Our goal was to perform clear experiments on ten routine tasks that our data scientists, analysts, and evaluators conduct in IEG.

We tested various applications, from writing and explaining code to producing econometric analyses, summarizing documents, and conducting literature reviews. We then compared work outputs generated by conventional AI or human intelligence with those produced by generative AI to assess accuracy, while also checking for errors and hallucinations. At the time we captured our main recommendations in a simple graph (see below).

Since then, a lot has changed. There are many more genAI models out there and their performance on most tasks has increased tremendously. We have also learned a great deal from this first set of experiments and improved how we use genAI for these tasks. For instance, we know that we really need to parse out the complex tasks for literature review and syntheses, much more than we did in that first batch of experiments to get better output. We also know how to prompt better to avoid hallucination, etc. we've learned from these experiments and improved our models and usage strategies. We are now embedding routine experimentation in our evaluation workflow. We plan to continue this approach for the foreseeable future while always ensuring that we apply good practices of human-in-the-loop, and—I'd even say, human-in-the-driving-seat."



Recommended Uses of GPT for Evaluation Practice.

Source: Intendent Evaluation Group, World Bank. Note: A = Analyst, DS= data scientist, TTL = task team leader.

Your experiments indicate that many of the tasks you tested can be performed by AI. There is often speculation that AI will lead to job loss and significant changes in work tasks. Based on your experimental results, do you believe this will also apply to the field of evaluations?

"It's important to clarify that just because we've run some successful AI experiments, we haven't replaced all our analysts! However, when we find a promising use of AI, we try to replicate the experiments in a handful of evaluation contexts and then share our findings with our colleagues to encourage analysts and evaluators to consider leveraging AI in similar tasks. In response to your question, I **believe there will be a shift in skill sets, but it will be a phased approach**. Currently, **the focus is on developing minimal AI literacy for everyone, as we need educated users rather than blind ones.** For those performing analytical tasks—working with data and code writing—there will be changes in the division of labor. For instance, we recruited five analysts this year, and their job descriptions now emphasizes data science skills a lot more than those we hired five years ago. Regarding job losses, I prefer not to speculate. Brilliant colleagues at the World Bank have worked on what generative AI might mean for the future of jobs. You should take a look if interested. With regards to evaluators, my hope is that we will be able to enhance our capabilities, use AI for more mechanistic and routine tasks (e.g., classification, summarization, document reviews, etc.) and dedicate more time, expertise and human skillset to analysis, engagements with stakeholders, theory building and complex tasks.

That said, **we very much need to learn how to use and govern AI effectively while mitigating its risks, including its environmental impact. There will still be plenty of jobs for humans; we just need to rise to the occasion and better understand and define the role of AI.** Personally, I believe we have room for improvement in that regard."

Estelle Raimondo is working at the World Bank to integrate data science and artificial intelligence into evaluation practice.

Jochen Kluve and Estelle Raimondo in a lively exchange about the role of artificial intelligence in evaluation practice.

Jochen, we've gained valuable insights into the potentials, perils, and practical applications of AI in evaluations. What can you share about the role of AI in evaluation at the KfW Development Bank? Are there any specific use cases or observations you can share?

"First, I want to emphasize that the systematic experimentation done by Estelle and her team is extremely useful for us, as it helps guide our work with AI. We're advancing nicely in making use of genAI in our everyday work, e.g., leveraging KfW's own ChatKfW for evaluation needs while adhering to institutional data protection obligations. Besides daily uses, we are increasingly incorporating structured AI applications. **One idea, for instance, is to use AI for transcribing and translating interviews conducted in local languages, which allows us to hear voices we may not have been able to hear beforehand.** Another example is the endeavor to **set-up an AI-based platform for standardized ex-post evaluations:** we want to leverage AI from front-to-end, i.e from formulating indicators and theory of change models at the beginning up to the user-friendly communication of evaluation results.

Besides those positive developments, my main concern is that while we are using AI tools to enhance the efficiency of our everyday evaluation work, there is a potential **risk of overreliance on AI that may endanger the trust in our evaluation results.** In case information provided by AI is incorrect and human oversight insufficient, users of the evaluation may begin to doubt the validity of our results. This could **undermine the foundation of our role as evaluators,** which is to convey truth data and maintain 100% accountability."

May erroneous, AI-generated content fundamentally erode the trust that underpins our evaluation work? What do you think, Estelle?

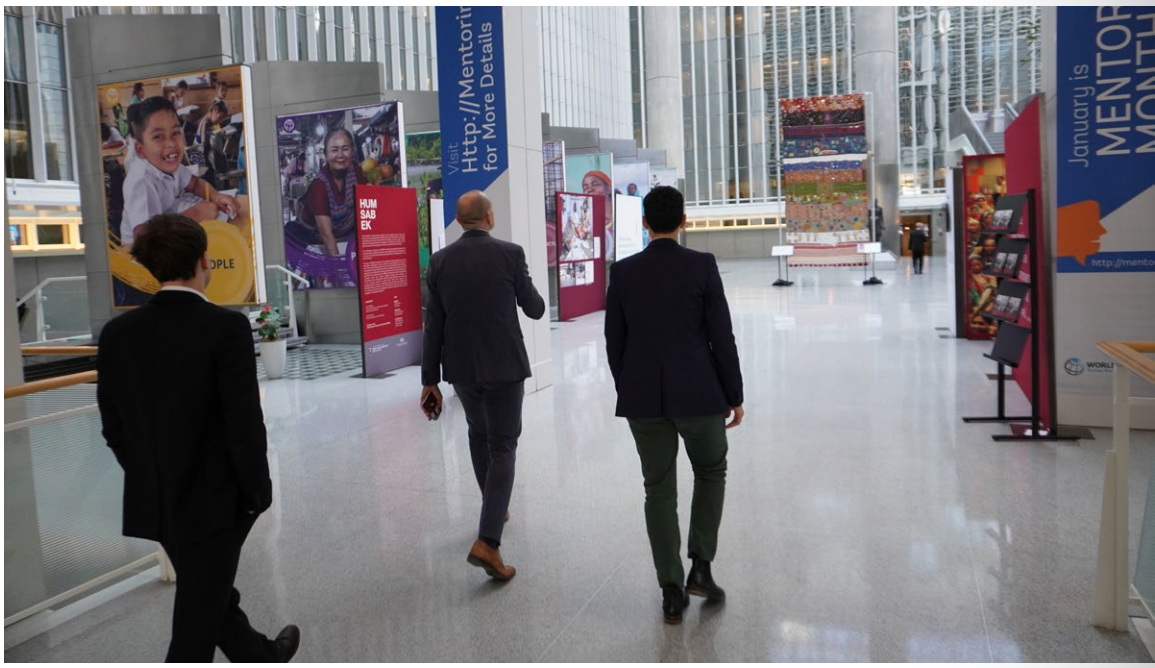
"I understand your concerns, but my thinking has evolved a bit, and I now view this issue in two ways:

First, there is the **assumption that human judgment is always correct.** For certain tasks, such as project evaluations involving a single case, human intelligence is indeed essential. However, in my field, some tasks require months of manual coding and classification from multiple analysts who follow a systematic template and apply relatively simple and repetitive decision rules. In these cases, there is a risk of human error due to the repetitive nature of the work. And in some of our experiments, we have found that machines can outperform humans on these types of tasks.

Second, we need to consider our **tolerance for error**. For some tasks we require 100% accuracy because these findings will be picked up by major publications in the public. In such cases, human oversight is crucial. Conversely, for tasks earlier in evaluation production process, we can accept some inaccuracies. For instance, when conducting literature reviews, we may encounter errors of inclusion or exclusion in our search, but **if the overall results are comparable to what a human would produce, we might prioritize efficiency over absolute accuracy.**

Ultimately, I see two areas where I am willing to not be 100% accurate if I don't have to: **First, we should leverage AI to accomplish tasks that were previously unfeasible,** allowing us to take calculated risks. Whether it's leveraging images as data, or whether it's dealing with huge segments of text. We could not do that without AI, so there I have a higher risk appetite. Second, there are certain tasks that are less fulfilling but necessary. Examples could be manual coding, data classification etc. In these cases, I prefer to gain efficiency so that analysts, evaluators, and data scientists can **redirect their time to more meaningful work.** This could mean redirecting the work to conduct more interviews, dedicate more time and energy to meaningful engagements with a broader range of stakeholders, etc., which will allow us to perform deeper analyses."

Thank you so much, Estelle and Jochen, for sharing your thoughts and experiences.



Estelle Raimondo and Jochen Kluve at an exhibition at the World Bank in Washington, DC.

A personal take on AI

What is the biggest change you have in your daily life due to AI?

"My own role and responsibilities have evolved significantly because of AI. I have the privilege to work with wonderful data scientists and lead our Data and Tech initiatives."

"The existence of genAI forces us to re-think the way we work, how we can integrate and use it effectively, how we can avoid traps, how we want to map out the future. That's inspiring."

Which AI-based app do you use the most?

"Copilot – I use it a lot to assist me in daily tasks."

"ChatKfW."

What is your biggest fear and hope related to AI?

"Fear: that we are not able to govern AI. Hope: That we will succeed in mastering AI for a greater good."

"Fear: That we and our kids will soon not be able to distinguish what is AI-made and what is not. Hope: That AI and human intelligence will interact in a beneficial way!"

Collaboration within German development cooperation
A joint evaluation by GIZ and KfW

The majority of German development cooperation projects are initiated in the framework of governmental negotiations between the Federal Ministry for Economic Cooperation and Development (BMZ) and the government of the partner country. **The actual implementation is then the responsibility of the partner country together with one of the implementing organisations (DOs):** In addition to KfW Development Bank, these include the German Society for International Cooperation (GIZ), the National Metrology Institute (PTB) and the Federal Institute for Geosciences and Natural Resources (BGR).

The DOs therefore often meet - equipped with different competences - on site in the partner countries and sometimes work in similar sectors or with the same partners to implement the objectives of German development cooperation. In the spirit of "alone we are strong, together we are stronger", **collaboration with a focus on their respective expertise is therefore essential in order to enable different projects to work together, to exploit synergies and to act with "one voice" abroad.** Collaboration between GIZ projects, known as "technical cooperation (TC)", and KfW Development Bank projects, known as "financial cooperation" (FC), is already a reality in many partner countries and is sometimes set up strategically by the BMZ.

Together with GIZ's Evaluation Unit, KfW Development Bank's Evaluation Department examined this collaboration with the aim of promoting joint institutional learning. The independent evaluation units of both organisations jointly addressed the following questions: What forms of collaboration exist in practice in the partner countries? Which factors are relevant for successful collaboration? How does collaboration influence the developmental impact of the projects, e.g., on areas such as poverty reduction, environmental protection, education, health or infrastructure promotion? Over 1,000 GIZ and KfW Development Bank employees took part in the interviews, focus group discussions and an online survey.

The results of the evaluation show that collaboration between the two DOs is part of everyday working life for many employees: Around two thirds of GIZ respondents and 90% of KfW respondents reported experiences of collaboration. Due to the diversity of contexts and constellations, there is not only a lot of collaboration, but also a wide variety of means of collaboration: these range from fixed exchange formats, joint planning workshops and progress reviews to the co-production of studies and joint public relations work.

Development cooperation on the ground in Cotonou, Benin.

Collaboration does pay off - but not equally in all project constellations: It adds the greatest value where there is potential for synergies, mutual risk reduction or the resolution of contradictions and competition. Synergies arise in the vocational training sector, for example, when GIZ supports curriculum development and training, while KfW Development Bank finances the premises and technical equipment required for training. In another example, the capacity strengthening of small business borrowers by GIZ led to higher repayment rates of credit lines financed by KfW Development Bank. This, in turn, reduced the financial risks for FC.

Overall, the evaluation shows that close collaboration can have a positive impact on German development cooperation projects: Especially on the effectiveness, impact and sustainability of projects. At the same time, collaboration contributes to the coherence of German development cooperation.

The evaluation also identifies success factors and good conditions for successful collaboration. These include, for example, knowledge about the working methods of the other organization, active consideration of collaboration when projects are handed over from one project manager to the next, and the active promotion of collaboration at the managerial level.

The evaluation offers valuable starting points for the design of future collaboration, which have already been taken up by the organizations. For example, new employees now get the chance to get to know the other organization right at the start of their employment and to network. There is also an exchange on specific good practices in cooperation and there are initiatives to expand joint public relations work in the partner countries. The successful collaboration between the two evaluation units for this particular evaluation is also an example of the goal-oriented and value added collaboration between GIZ and KfW.

German development cooperation: GIZ and KfW Development Bank in their joint office in Yaoundé, Cameroon.

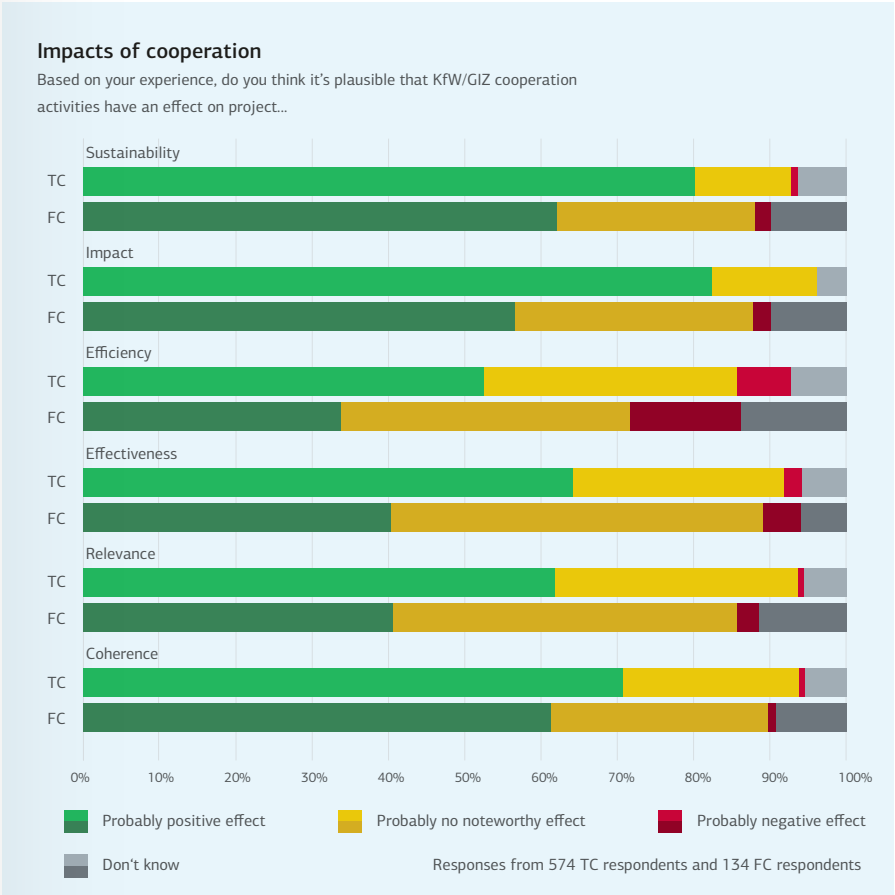


Figure: The impact of collaboration on development cooperation projects.

Evaluation report 2023-2024: Methodology

From projects to success rate

Every year, FC Evaluation draws and evaluates a representative stratified sample from all FC projects. This sample then forms the basis for calculating the success rate.

Ex-post evaluations
From all projects to success rates

The evaluation department of KfW Development Bank draws an annual representative random sample, stratified by sector, from the population of all completed projects of financial cooperation (FC). An ex-post evaluation must be carried out for all projects that were drawn into the sample. This procedure ensures the statistical representativeness and transferability of the evaluation results to the full FC portfolio and ensures independence in the selection of the projects to be evaluated. In the two-year period 2023-2024, 129 projects were drawn into the evaluation sample from a total of 268 completed projects. This sample size roughly corresponded to the long-term average.

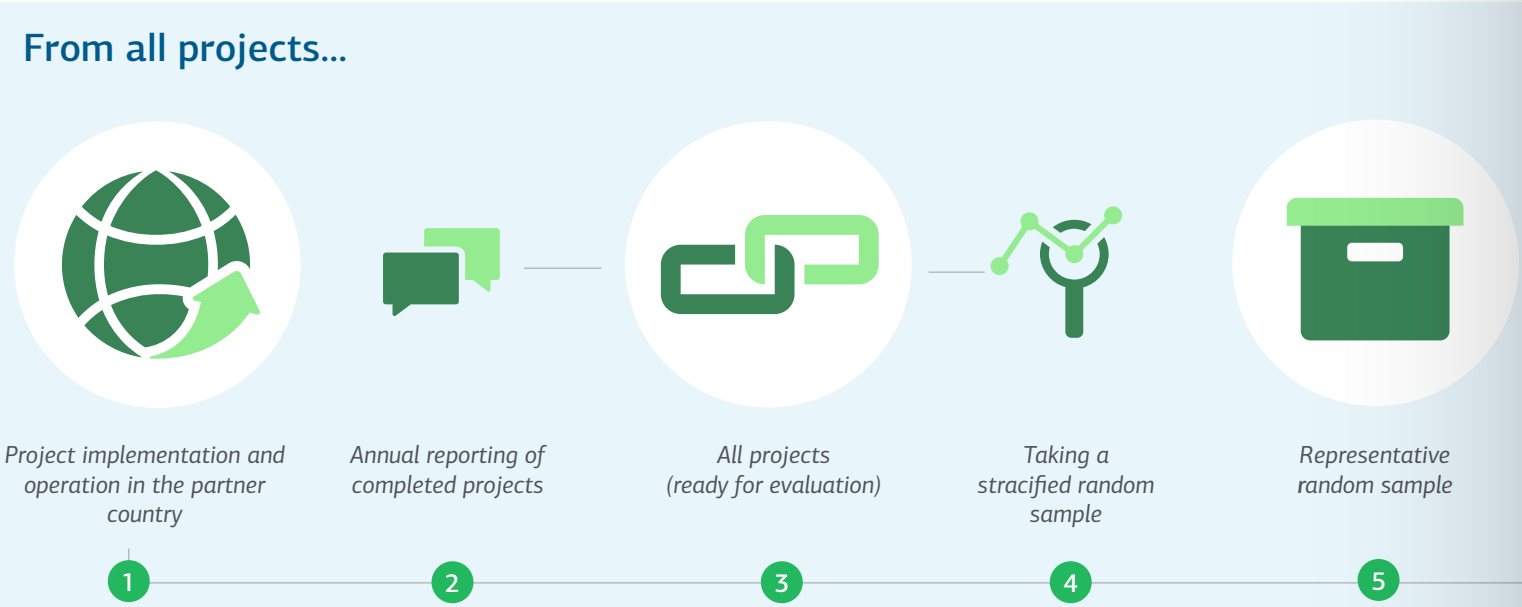
Not all evaluations can be carried out directly in the year of sampling: 65 projects from the 2023-2024 sample have already been evaluated. In addition, 49 projects from the samples of previous years were evaluated.

Furthermore, FC Evaluation also evaluates projects if they can be combined with a to-be-evaluated project even if they are not themselves part of a sample. Criteria for this would be, for example, spatial and temporal proximity to projects from the sample. In the 2023-2024 evaluation period, 26 such "bundled" projects were evaluated.

In summary: A total of 140 financial cooperation projects were evaluated in the years 2023-2024. Of these, 65 came from the 2023-2024 sample and 49 from previous samples, while 26 were bundled.

From the 2023-2024 population of 268 projects, 65 have already been evaluated (blue bar), 64 are still being evaluated (green bar) and 139 were not selected for evaluation (gray bar).

The projects to be evaluated are assessed by FC Evaluation based on the six OECD-DAC criteria: relevance, coherence, effectiveness, efficiency, impact and sustainability. The assessment of these criteria is based on the evaluation guidelines of the German Federal Ministry for Economic Cooperation and Development (BMZ). In concrete terms, this means that each criterion is considered separately and rated in the range from one (very successful) to six (highly unsuccessful). Based on this, the overall assessment of the project is determined, which is often the arithmetic mean of the six sub-ratings. However, this is not always the case: Especially serious deficiencies in individual sub-criteria should not be offset by positive results in other sub-criteria. For this reason, ratings lower than 3 (moderately successful) in the effectiveness, impact and sustainability criteria mean that the project cannot be rated as successful overall. The basic idea: Projects that do not have a (sustainable) impact cannot be "successful".

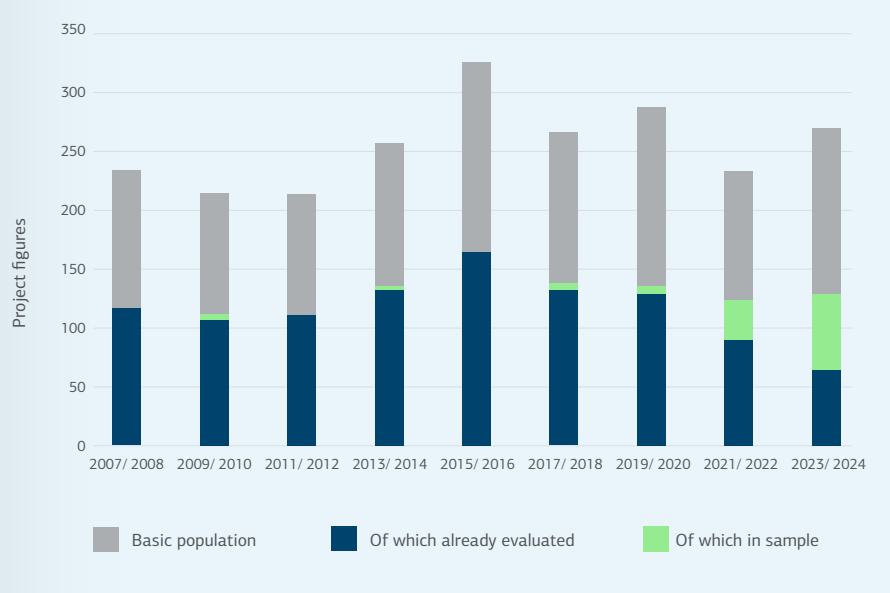


From the full project population to the success rate - A process in nine steps

The overall success rate is then estimated based on all individual evaluations: the rate is the statistical proportion of successful FC projects - i.e., the proportion of projects that received a rating of 3 or better as overall assessment. Only projects that were drawn from the sample for the particular time period, in this case 2023-2024, are used to estimate the success rate (unbiased estimate). For the period 2023-2024, therefore, only 65 project evaluations, that have already been finalized, are included.

The success rate for the period 2023-2024 is 85%, which is slightly above the average success rate for 2007-2024 (81%), but may still change because some evaluations from the sample are yet to be carried out.

Important for interpreting success rates: they are estimates and are therefore subject to statistical uncertainty. The confidence interval provides information on the accuracy of the estimate. The success rates of the past two-year evaluation periods since 2007/2008 differ slightly from each other (81% and 85%). Yet, this difference is not statistically significant, i.e., the values lie within the same confidence interval and are therefore not reliably different from each other, as the small difference between them may be due to chance.



Number of all completed FC projects (full population), projects already evaluated and yet to be evaluated, by two-year sample periods from 2007/08 to 2023/24.

...to success rates

Independet ex-post evaluation — 6

Project database with results — 7

Estimated success rates — 8

success rate with 95% confidence interval — 9

FC Evaluation: Who we are

We are an independent evaluation department that uses recognised and modern methods to support the three pillars of our evaluation work: ex-post evaluations, the KfW Development Impact Lab, and institutional learning.

Vision

We improve the effectiveness of Financial Cooperation (FC) in line with the 2030 Agenda and its Sustainable Development Goals (SDGs) by supporting FC projects with our evaluation expertise by critically assessing them, and by identifying lessons learned for future projects.

Mission Statement

We combine internal and external evaluation expertise to systematically assess the effects of completed FC projects and to support ongoing FC projects by means of impact evaluations. We enhance institutional learning by facilitating evaluation results in products that are tailored to the target group and can be used efficiently. We are committed to professional excellence, continuous skills enhancement and innovation.

Team

As an agile team of sector experts with scientific and/or operational experience, we see ourselves as a "knowledge hub" for evaluation and impact measurement within KfW Development Bank. We cooperate with the Federal Ministry for Economic Cooperation and Development (BMZ) and network with other national and international institutions dedicated to evaluating development cooperation. The involvement of local partners and their interests and learning objectives are important to us. Our work incorporates relevant standards of the evaluation community (DeGEval, OECD-DAC) and of empirical research, in particular with respect to the protection of personal rights and the anonymity of personal data as well as an ethical code of conduct.

The three pillars of our evaluation work

KfW Development Impact Lab

- We support ongoing FC projects by developing and implementing customised impact evaluation designs.
- We answer questions that are of particular relevance to FC.
- We use modern methods of rigorous impact evaluation (RIE) and use and adapt a design from the RIE toolbox that is appropriate for the respective project (form follows function).
- We seek to collect and use data in innovative ways, especially satellite data.
- We cooperate with suitable external partners on a case-by-case basis, especially research institutions and other development banks.

Ex-post evaluations

- Every year, we draw a representative random sample of completed FC projects, stratified by sector, for ex-post evaluation (EPE).
- We conduct ex-post evaluations ourselves and work together with external experts and secondees from the operational departments of KfW Development Bank.
- Ex-post evaluations have been carried out systematically by the FC Evaluation since 2000. They promote accountability and provide the key knowledge base for institutional learning.
- Ex-post evaluations are carried out systematically on the basis of the OECD-DAC criteria.
- Ex-post evaluations are the main pillar of our evaluation work.

Institutional Learning

- We are guided by the interests and learning objectives of our target groups and provide up-to-date information in modern formats.
- We offer topical learning formats on various themes, organise training on evaluation methods and contribute our knowledge to peer discussions and sector retreats.
- Our colleagues in operations actively carry out ex-post evaluations together with us (secondment).
- We report on new results via newsletters and publication series.
- We provide the content of all ex-post evaluations since 2007 in digital and interactive form within our IDEaL-App (Interactive Database Evaluation and Learning).

KfW Group
KfW Development Bank
Palmengartenstraße 5 – 9
60325 Frankfurt am Main

Tel. +49 69 7431-0
Fax. +49 69 7431-1234
FZ-Evaluierung@kfw.de
www.kfw.de

Design and layout by
new office GmbH, Frankfurt am Main

Published: May 2025

Photos

Cover: Jonas Wresch; p. 1 Jonas Wresch; p. 2 Thomas Meyer; p. 2 + 3 new office; p. 4 + 5 Jonas Wresch; p. 6 + 7 Jonas Wresch; p. 8 + 9 new office, Svenja Klingspor; p. 9 Jens Drillisch, Sebastian Manthey; p. 10 Yahia Maraka; p. 12 + 13 Nicole Karnaus, Jonas Wresch; p. 14 new office; p. 15 Jean Joseph Manampisoa RATOVO-SON, Ony Rakotoarisoa, Ingrid Dallmann; p. 16 Kati Krähnert; p. 17 Nguyen Minh Phuong, Mahesh Joshi, Sven Lorenz; pp. 18 + 19 Jonas Wresch; p. 20 + 21 new office, Svenja Klingspor, André Hensel; p. 22 + 23 Svenja Klingspor, Jonas Wresch; p. 24 + 25 Jonas Wresch, Mark Naftalin; pp. 26 + 27 Svenja Klingspor, Noten, die FZ; p. 28 + 29 Jonas Wresch; p. 30 + 31 Jonas Wresch; pp. 32 + 33 Jonas Wresch; p. 34 Estelle Raimondo, Jochen Kluve; p. 35 + 38 + 39 + 40 Melvin Wong; p. 42 Jonas Wresch; p. 43 Franziska Grimm; p. 44 + 45 Jonas Wresch; p. 46 + 47 new office; p. 50 Jonas Wresch