

MID-TERM EVALUATION OF THE GLOBAL TAX PROGRAM (GTP)

MID-TERM EVALUATION

FINAL EVALUATION REPORT

seo • amsterdam economics

AUTHORS

NIENKE OOMES, THIERRY BELT, CLAUDIO CORTE COI, MAARTJE DE GROOT (SEO)
& HILDE VAN DIJKHORST (MDF)

COMMISSIONED BY

THE WORLD BANK GROUP (WB)

JANUARY 5, 2023

Executive Summary

The World Bank (WB) commissioned SEO Amsterdam Economics to conduct a mid-term evaluation of the Global Tax Program (GTP). The overall finding is that the GTP is a well-managed program and has thus far performed well against the OECD-DAC evaluation criteria, with opportunities for further improvements.

RELEVANCE

Various sources confirm that GTP-supported projects are well aligned with the needs and priorities of recipient country authorities. The GTP has also responded well to emerging issues such as COVID-19, climate change and gender inequality.

- Stakeholders widely agreed that GTP-supported TA projects are demand driven, relevant to the recipient country, despite a few concerns about absorption capacity.
- Case study projects scored well on relevance in all cases, since they were generally found to be (1) demand driven (i.e., responding to country requests), 2) addressing important gaps (e.g., revenue gaps or capacity gaps), and (3) in line with national priorities.
- One key advantage of the GTP is that it can leverage the World Bank's strong local presence through its network of country offices, which can facilitate the alignment of projects with the needs and priorities of TA recipients.
- The GTP has thus far responded well to emerging issues, by addressing topics such as climate change, gender inequality and health taxation. The WB's general strength in such thematic areas enables the GTP to be especially relevant in these fields.
- The GTP generally adapted well to the COVID-19 pandemic, including by replacing in-person activities with online activities where possible. However, travel budget savings could have been reallocated more quickly.

COHERENCE

The GTP actively monitors related activities by other WB teams and development partners. Internal and external coordination is taking place, and overlap and conflicting policy advice with other development partners is usually avoided. However, there is room to further increase complementarities and strengthen synergies, particularly with other development partners.

Internal Coherence

- GTP project teams have an eye for complementarities and synergies with other Bank operations in the tax as well as the fiscal planning and budgeting domain, mostly other WB concessional lending operations or Advisory Services and Analytics (ASAs).
- Stakeholders generally saw GTP TA as coherent and synergetic with other WB operations (especially concessional lending operations), although some indicated that the GTP may be constrained somewhat due to the fact that not all WB teams see taxation as a key development tool.
- Case studies confirmed that internal coherence was generally largely or fully achieved, but that the linkages between TA and concessional lending operations could be further streamlined.

External Coherence

- The GTP actively monitors complementarities with other development partners, such as the IMF or the Organisation for Economic Cooperation and Development (OECD).

- Multiple information sources (document review, stakeholder survey, case studies) confirmed that the GTP has generally coordinated at an acceptable level with other development partners.
- Overlap and conflicting policy advice is usually avoided, but there is room to increase complementarities and further strengthen synergies with other development partners. This applies in particular to the IMF and the OECD, including through the Platform for Collaboration on Tax (PCT).

EFFECTIVENESS

Despite COVID-19, most GTP-supported projects were effective in achieving their targets for outputs and short-term outcomes. However, longer-term outcomes were difficult to assess, since most projects were still ongoing and the GTP results framework does not yet include longer-term indicators.¹

- Both the GTP-wide portfolio analysis and the in-depth case studies confirmed that the targeted outputs and short-term outcomes were largely achieved, and that most projects were on track to meeting their objectives, despite some delays and constraints caused by COVID-19.
- The GTP results framework is well developed and appreciated by donor partners. The underlying logical framework shows an operationalized intervention logic, but the overall “Theory of Change” can be improved.
- Most indicators are formulated at the output level (e.g., “recommendations proposed”) or the short-term outcome level (e.g., “recommendations endorsed/adopted”). The indicators are well formulated, although the terms “endorsed” and “adopted” could be better defined.
- One shortcoming is that the results framework does not (yet) include longer-term indicators related to the implementation of reform recommendations, since they were seen as “aspirational” and beyond the GTP’s control.

SUSTAINABILITY

The WB’s strong country presence and its ability to provide additional funding through concessional lending operations bode well for ensuring the sustainability of GTP results. However, generic risks to sustainability include high turnover of local officials, limited absorption capacity, political instability, and the fact that the results framework does not (yet) include long-term outcomes.

- Stakeholders were generally positive about the likelihood that GTP results will be sustainable over time.
- The strong local presence through WB country offices increases the GTP’s potential to achieve sustainable results.
- In our case studies, sustainability was typically seen as only partially achieved. This is partly due to the fact that it was too early to judge, but also because sustainability is an endemic issue in development, especially in fragile states.
- Risks to sustainability identified during case studies included the high turnover of local officials, the sometimes limited absorption capacity of recipients, and political instability.
- Risks to sustainability at the strategic level included the short-term focus of the results framework and a possible loss of focus if the GTP would cover too many countries and areas.
- Factors that were seen as promoting the sustainability of results in case studies included high ownership and relevance, a change in the legal framework, and the combination of various types of support.

¹ While the GTP committed to track these so-called “aspirational” indicators, the ToR denotes that “most of them are out the projects’ control and the associated results cannot be solely attributed to GTP-funded project activities.”

EFFICIENCY

The GTP generally manages its projects efficiently and in a timely manner. The evaluation team did not find evidence of any major issues regarding cost-effectiveness. However, operational efficiency could be further improved.

- The fact that the GTP Secretariat monitors the “output efficiency” of its projects is a strong aspect of GTP project management, but the definition of output efficiency could be further improved.
- The evaluation team assessed all case study projects as moderately efficient and cost-effective.
- Stakeholders were generally satisfied with procurement processes and the organizational efficiency of the GTP, apart from some minor concerns about timeliness, exceptions to procurement procedures, and some issues related to the way in which the GTP fits inside the WB’s structure.
- Most project teams delivered the results in a timely manner and gave realistic explanations for any delays (e.g., due to COVID-19).

Table S.1 Aggregate GTP ratings by OECD-DAC criteria

Criteria	GTP-wide assessment	Project-level assessment	Overall score
Relevance	4	4	4
Internal Coherence	3	3.5	3.25
External Coherence	3	2.5	2.75
Effectiveness	3	3	3
Sustainability	3	2.5	2.75
Efficiency	3	3	3

1 = Not achieved; 2 = Partially achieved; 3 = Largely achieved; 4 = Fully achieved. The overall score was determined by using equal weights for the assessments at both levels.

Table S.2 Case Study ratings by OECD-DAC criteria

Criteria	CS 1	CS 2	CS 3	CS 4	CS 5	CS 6	CS 7	CS 8	CS 9	CS 10	Overall rating
Relevance	4	3	4	4	3	4	4	4	4	4	4
Internal Coherence	3	3	3	4	2	3	4	4	4	4	3.25
External Coherence	3	3	3	2	2	3	3	2	4	4	2.75
Effectiveness	3	3	3	3	3	3	3	3	3	3	3
Sustainability	2	3	2	3	2	2	4	3	2	2	2.75
Efficiency	3	3	3	3	3	3	3	3	2	2	3

1 = Not achieved; 2 = Partially achieved; 3 = Largely achieved; 4 = Fully achieved. The overall score was determined by using equal weights for the assessments at all levels and rounding the resulting figure to the nearest quarter point.

List of Abbreviations

ASA	Advisory Services and Analytics
ATI	Addis Tax Initiative
CD	Capacity Development
DP	Development Partner
DPF	Development Policy Financing
DRM	Domestic Resource Mobilization
FP	Fiscal Policy
GTP	Global Tax Program
IDA	International Development Association
IEG	Independent Evaluation Group
IMF	International Monetary Fund
M&E	Monitoring and Evaluation
MTE	Mid-Term Evaluation
MTRS	Medium-Term Revenue Strategy
OECD	Organisation for Economic Cooperation and Development
OECD-DAC	OECD - Development Assistance Committee
PCT	Platform for Collaboration on Tax
SC	Steering Committee
TA	Technical Assistance
TADAT	Tax Administration Diagnostic Assessment Tool
ToC	Theory of Change
ToR	Terms of Reference
TTL	Task Team Leader
WB	World Bank

Table of Contents

Executive Summary	i
List of Abbreviations	iv
Table of Contents	v
1 Introduction	1
1.2 Background of the GTP	1
1.3 Purpose and scope of the mid-term evaluation	1
2 Analysis and findings	3
2.1 Relevance	3
2.2 Coherence	7
2.3 Effectiveness	11
2.4 Sustainability	22
2.5 Efficiency	25
3 Conclusions and recommendations	30
3.1 Conclusions	30
3.2 Recommendations and lessons learned	33
Appendix A Evaluation questions	37
Appendix B Case study selection	39
Appendix C GTP Theory of Change	43

1 Introduction

1.2 Background of the GTP

The Global Tax Program (GTP) Umbrella is a multi-donor trust fund program to provide an integrated framework for tax support aimed at strengthening tax systems in developing countries. Established by the World Bank (WB) in June 2018 as a follow-up to the Addis Tax Initiative (ATI), the GTP aims to support developing countries with strengthening their tax institutions and improving domestic resource mobilization (DRM). This may include advancing such objectives as improving taxpayer morale, compliance, and tax certainty; reducing the compliance burden; simplifying administrative procedures; reducing wasteful tax expenditures; and ensuring greater participation of developing countries in the global tax agenda. As of end-2021, the GTP supported activities in 76 countries. The GTP Umbrella is financed with contributions from ten donor partners, valued at US\$ 84 million.²

The GTP Umbrella has two main pillars: the DRM pillar and the Fiscal Policy (FP) pillar. In 2020, the GTP Steering Committee (SC) agreed to create an additional temporary pillar to accommodate fiscal policy responses to address the economic and fiscal impact of the COVID-19 crisis through advice on effective, efficient and balanced fiscal policy responses, including strengthened accountability and monitoring. This Mid-Term Evaluation is focused on the DRM pillar and does not include the FP pillar.³

The SC guides the GTP's strategic direction and implementation. It is composed of representatives from donor partners and Bank management. Meeting twice a year, SC members discuss the Program's strategic direction, review progress against its work plan, and discuss and endorse work plans and any changes. The SC also reviews and provides feedback on new project proposals received on a rolling basis and through Calls for Proposals. The work of the GTP is underpinned by a Program Document, annual work plan, and results framework.

The selection process of GTP-funded projects is led by a Technical Committee, while the GTP Secretariat takes care of the general coordination of this process. The Technical Committee is comprised of representatives of technical and management staff across the WB. The GTP Secretariat also reviews all project proposals to ensure that they meet the basic quality standards agreed with donor partners. Moreover, the proposals are checked with respect to their strategic fit (country and program level), client demand, and complementarity with ongoing projects (internal and external). Lastly, the Secretariat reviews the financial planning of the proposals and verifies their alignment with the GTP Umbrella Program Results Framework.⁴

1.3 Purpose and scope of the mid-term evaluation

The GTP Program Document stipulates that a Mid-Term Evaluation (MTE) be undertaken in 2021-2022 to assess the GTP's program strategy, management, and activities to date. The main aim is to distill lessons learned from the first few years of implementation and provide recommendations for a "course correction" where needed.

² The ten donor partners are the governments of Australia, Denmark, France, Japan, Luxembourg, the Netherlands, Norway, Switzerland and the United Kingdom, as well as Bloomberg Philanthropies.

³ The contributions from donor partners valued at US\$ 84 million are split into US\$ 78.5 million for the DRM pillar, and US\$ 5.6 million for the FP pillar. Moreover, the FP pillar will close at the end of FY23.

⁴ GTP Program Document, p.79

Following a competitive selection procedure in 2021, SEO Amsterdam Economics was commissioned in October 2021 to carry out the MTE.

Compared to the original Terms of Reference (ToR) for this MTE, the evaluation team expanded the scope of the MTE to also cover all Organisation for Economic Cooperation and Development - Development Assistance Committee (OECD-DAC) criteria at the GTP-wide level. The ToR originally included 28 evaluation questions, of which 20 project-level questions and 8 GTP management and strategy questions. At the suggestion of SEO, it was decided during the inception phase to broaden the scope somewhat and split the evaluation into two parts: (1) a GTP-wide assessment of all OECD-DAC criteria (including the evaluation questions related to program strategy and management), and (2) a project-level assessment of the same OECD-DAC criteria (still guided by the project-level questions in the ToR).

The evaluation used various information sources.

- For the GTP-wide assessment, the evaluation team conducted a full portfolio analysis based on available portfolio data for all GTP projects, complemented with an assessment of OECD-DAC criteria at the GTP-wide level, based on desk research, interviews with GTP-wide stakeholders, and a stakeholder survey.
- For the project-level assessments, the evaluation team reviewed detailed project data and project-specific documents and interviewed a wide range of stakeholders for each case study, including representatives from the WB such as Task Team Leaders (TTLs) and project team members, TA recipients, donor partners and other external stakeholders (including other development partners active in case study countries).

The case studies used for this evaluation are a sample of ten Technical Assistance (TA) projects selected from the overall GTP list of TA projects, based on a list of selection criteria to maximize representativeness. The specific criteria considered for the selection of case studies were discussed and agreed with the GTP Secretariat during the inception phase. They are presented in Annex B, along with the specific selection of ten GTP projects proposed for case study evaluation.

This report presents the aggregate findings of the GTP-wide assessment and the project-level assessments. The report is organized by OECD-DAC criterion, i.e., for each criterion we first discuss the GTP-wide findings and then the project-level findings. As requested in the ToR, the evaluation also provides a general assessment of the quality of the logframes (included under “Effectiveness”).

The COVID-19 pandemic did not affect the modality of the evaluation, but the way in which it affected GTP projects was taken into account during project-level assessments. The ToR already envisaged that all interviews would take place through virtual means, while document reviews and surveys also did not require in-person field work. However, as also requested in the ToR, the team evaluated how projects had responded to the change in circumstances related to the pandemic, for example by checking whether country priorities had changed or whether projects were able to effectively implement new modalities for capacity development (CD) delivery (e.g., remote missions).

2 Analysis and findings

2.1 Relevance

Various sources confirm that GTP-supported projects are well aligned with the needs and priorities of recipient country authorities. The GTP has also responded well to emerging issues such as COVID-19, climate change and gender inequality.

2.1.1 Relevance: GTP-wide assessment

The GTP has various processes in place to ensure the relevance of GTP project proposals. The GTP Secretariat coordinates the call for proposal. After a proposal is drafted by the project team, the Task Team Leader (TTL), in close collaboration with the Global Tax Program Manager and Monitoring and Evaluation (M&E) specialist from the GTP Secretariat, aims to ensure alignment of the proposal with 1) the development objectives of the GTP and 2) the GTP Theory of Change (ToC). The Secretariat conducts a review to ensure that proposals meet the quality criteria agreed with the donors, while the WB Technical Committee (composed of representatives of WB technical and management staff) leads the selection process and ensures a high-quality project design and implementation. The Program Manager maintains the overview of the overall portfolio of GTP projects and can advise on whether a proposal is well aligned with the broader objectives.⁵ In addition to the GTP Secretariat and the WB Technical Committee, the proposal is also reviewed by a practice manager and the donor partners.⁶

The GTP results framework could play a stronger role in guiding the design and development of new projects. The alignment of country projects with the GTP results framework plays a role, but interviews suggest that this is not the leading document for the TTLs that develop/propose the project. Proposals and projects are based on the country partnership framework and country core diagnostics.⁷

Stakeholders largely agreed that GTP-supported TA projects are relevant to the recipient country, but had a few concerns about absorption capacity. In the survey, 98 percent of TA recipients agreed that the WB responded adequately to TA requests; more than 80 percent of TA providers and recipients agreed that projects were designed in close consultation with the recipient country; and more than 90 percent that GTP-supported TA projects reflected the priorities of the recipient country (see Figure 2.1 and Survey Annex). Moreover, over 80 percent of provider-respondents also agreed that the design of GTP TA took absorption capacity and ownership into account. However, some respondents questioned whether the TA design took the local context and available human capacity at the country level sufficiently into account.

⁵ Based on interviews, we understand that the GTP defines relevance through three criteria:

1. Focus on countries that are IDA-eligible, including fragile and conflict-affected states (FCS);
2. Alignment with the Bank's overall DRM approach and strategy;
3. Within countries, alignment with the Country Partnership Strategy and Systematic Country Diagnostic (SCD) tool developed by the WB.

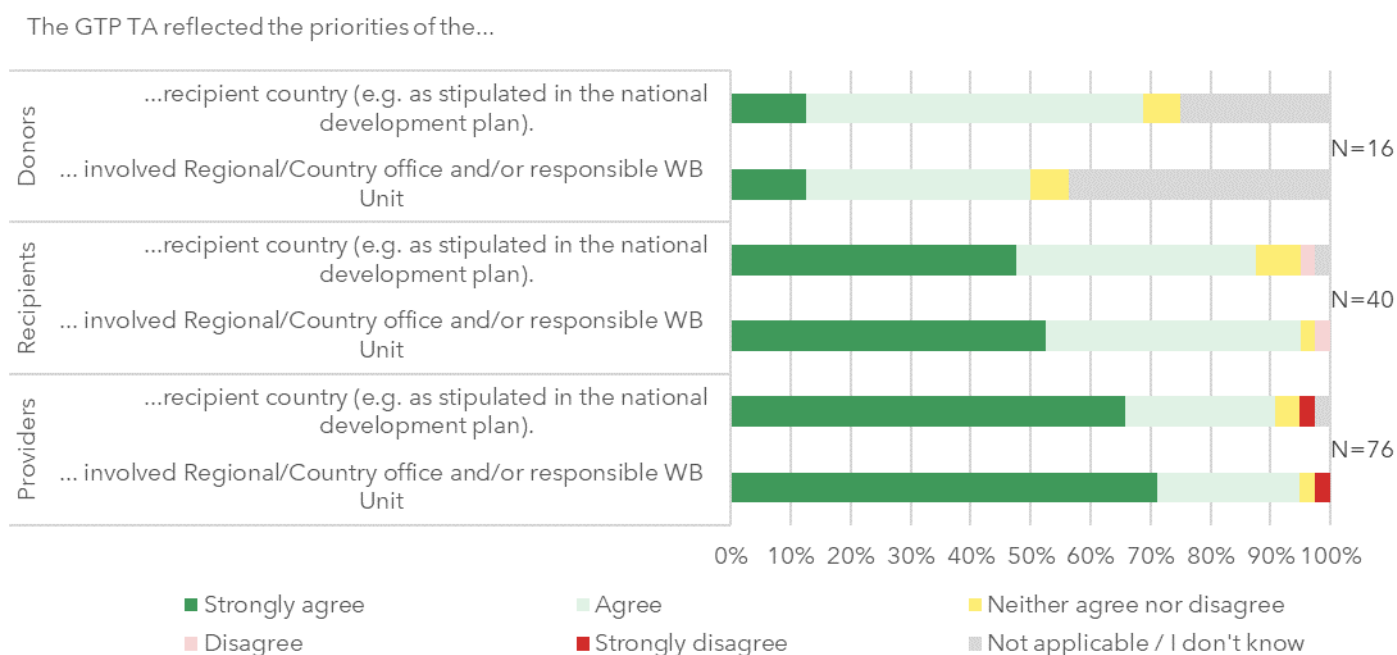
⁶ Under the new assessment model, donor partners do not need to review the proposal. They can also not disapprove a proposal.

⁷ <https://www.worldbank.org/en/projects-operations/country-strategies>

Despite the positive stakeholder views on relevance, there is room for further improvement. While it is generally accepted that projects should be “demand-driven” and that client needs should be central to developing a project, the system currently includes few formal controls to ensure that this is the case. Respondents indicated the importance of proposals and projects being based on recipient country demand before they are developed. In theory, the WB representative/TTL first discusses the project idea with recipient country representatives and internal clients (e.g., country unit, project leaders). If they give their approval, the project is assumed to fit the country and tax strategy. However, we learned from our interviews that, in practice, a project can also get approved by a manager and director at the WB without any formal checks within the system to see whether (a) it is based on or in line with client needs, or (b) it focuses on those areas with the highest revenue gaps and/or highest potential to reduce these revenue gaps.

For donor partners, the balance of the project portfolio or focus on particular topics (e.g., gender, climate) also played an important role in assessing the relevance of the project portfolio. In their view, a “balanced portfolio” takes into account regional balance, inclusion of fragile states, and focus on countries with high revenue gaps.

Figure 2.1 Survey respondents largely agreed that GTP TA reflected the priorities of the main stakeholder groups.



Source: SEO Amsterdam Economics, based on a survey of 132 stakeholders.

Multiple stakeholders noted that the WB’s strong country presence facilitates better alignment with client and country needs. Several stakeholders considered the GTP as more demand-driven than other trust funds, and specifically mentioned the Bank’s strong local presence through its large country offices as a comparative advantage (which distinguishes the WB from many other development partners). This strong local presence was seen as facilitating the identification of and alignment with recipient needs, and therefore improving the relevance of GTP-supported projects for the recipient. As one stakeholder noted: “There’s a lot of people in the Bank (country teams) who will watch whether you are delivering what the client wants”.

Only a few risks for relevance were identified. First, it is precisely the WB's strong country presence that could lead to a risk that the GTP may "source" more projects in countries with strong WB country teams, since these teams can have an important influence on identifying "demand" and co-designing strong GTP project proposals. This, however, is only a problem if the countries with the strongest country teams are not the countries with the highest needs. Second, some survey respondents raised concerns about whether the TA is always aimed at countries that need it most. This does not necessarily mean that the GTP is seen as operating in irrelevant countries, but rather that the GTP may need to better communicate how country needs are assessed, e.g., during the project proposal stage. Having said this, it is also important to recognize that country needs are in flux, and in line with sustainability considerations it would not be advisable to constantly change the country portfolio accordingly. Moreover, as a GTP management team member noted rightly: "Demand and supply are not perfectly elastic—it may take time to mobilize the resources for a country with an emerging need. But the flexibility in the way resources are planned and allocated does provide for this to an extent."

2.1.2 Relevance of GTP program strategy

The WB's strengths in thematic areas such as gender and inequality enable the GTP to be especially relevant in these fields. The GTP's strategy and portfolio has been aligned with thematic commitments and contextual choices linked to the yearly update of the International Development Association (IDA). In 2020, the main IDA20 themes were climate change; fragility, conflict and violence; gender; and jobs and economic transformation. The design of the program and allocation of resources are also guided by IDA's focus on fragile contexts as facing the biggest challenges and therefore best served by the program.⁸ Furthermore, being situated within the WB, the GTP is well placed to connect the tax domain with broader socioeconomic issues such as inequality and poverty, both thematically and within the project portfolio. These strategic priorities are not reflected in GTP's ToC and only partly in the logframe, which includes more of these broader indicators. Considering the chosen prioritization of fragile contexts, a more refined assumptions and risk analysis would be expected, beyond the mention of political economy as a risk.

The program has thus far responded well to emerging issues by addressing issues such as climate change, gender equality and health taxation ("sin taxes"), both thematically and through the project portfolio. With regard to gender equality and taxation, for example, the GTP started a new project in 2021 that will run until 2024 with the objective to "improve the ability of low and middle income countries to use fiscal policies and practices to contribute toward achievement of gender equality goals".⁹ The importance of gender equality in taxation is in line with IDA 20, but has also been specifically promoted by FCDO, which is the single donor of this project, as well as through underlying country projects on gender equality in Ethiopia and India. For the health taxation work stream, Bloomberg Philanthropies and the Norwegian Agency for Development Cooperation (NORAD) have been instrumental.

The GTP adapted its strategy and TA delivery to the COVID-19 pandemic. The GTP's response to the pandemic included the development of a Fiscal Policy Program to mitigate the impact of COVID-19 through tailored advice to governments on fiscal policy responses, and to facilitate monitoring of resource use for the COVID-19 response.¹⁰ Around 80 percent of the survey respondents indicated that the GTP (fully, or at least sufficiently) adapted its TA delivery to the COVID-19 pandemic. Some recipients noted that they "actually received more support due to the flexibility of virtual meetings". However, respondents also noted that virtual workshops tend to have different

⁸ GTP (2020) Strategy Note for the Steering Committee meeting January 2020

⁹ GTP (2021) Integrating Gender Equity into Tax Reform, concept note, p. 5.

¹⁰ GTP Umbrella Program Document, 2020, p. 82

dynamics than face-to-face workshops, with the latter being more effective. Also, one interviewee indicated that the GTP could have more flexibly adapted its TA delivery to COVID-19. A few stakeholders indicated that the GTP could have better directed its efforts to one-on-one meetings rather than virtual ones.

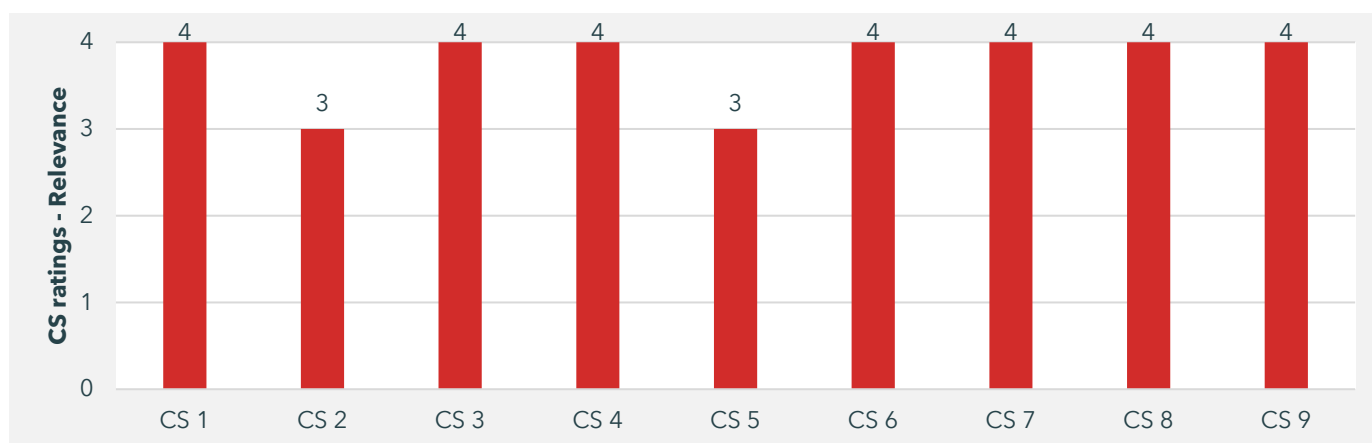
The GTP's response to COVID-19 had an unintended positive result on promoting specific strategies. Other adaptations that were made, as shared in the interviews, were to alter the sequence of activities to ensure that training components would follow later, and also that the GTP shifted its focus to more short-term issues. An unintended positive consequence of the response to the COVID-19 pandemic was that it created a push for more "fair, equitable and inclusive growth"¹¹ and "green recovery" activities through, for instance, the environmental taxation and CPAT projects in the GTP portfolio.¹²

2.1.3 Relevance: project-level assessment

The relevance of the case studies scored well in all cases (Figure 2.2). The ten selected projects were generally found to be 1) demand driven (i.e., responding to country requests), 2) addressing important country gaps (e.g., the tax-to-GDP ratios), and 3) in line with national priorities. Per definition, projects focusing on improving DRM in lower-income countries have high potential to be relevant. All individual country projects within the set of case studies were formulated in (close) cooperation with the local governments and/or were informed by assessments using diagnostic tools such as the Tax Administration Diagnostic Assessment Tool (TADAT). Two case studies were not fully demand driven in the sense that the GTP team took an active role in initiating or formulating the request. However, they were still relevant to the countries in question. Moreover, virtually all case study projects managed to shift focus and modality based on country priorities in response to the COVID-19 pandemic, although some stakeholders mentioned that this process was not always straightforward.

Having said this, political economy risks and the technical capacities of recipient authorities are two points of attention. For one country, the political economy risks that later impeded project progress could have been better anticipated and dealt with during project design. For another country, there were some concerns that the project's design was not sufficiently sensitive to the technical capacity gaps within the authorities.

Figure 2.2 Most case studies were rated as having fully achieved relevance



Source: SEO Amsterdam Economics. 1 = not achieved; 2 = partially achieved; 3 = largely achieved; 4 = fully achieved.

¹¹ Meeting notes of the GTP Steering Committee - September 21-22, 2020, p.1

¹² GTP (2022) Annual Report FY 2021, p. 21

2.2 Coherence

The GTP actively monitors related activities by other WB teams and development partners. Internal and external coordination is taking place, and overlap and conflicting policy advice with other development partners is usually avoided. However, there is room to further increase complementarities and strengthen synergies, particularly with other development partners.

2.2.1 Coherence: GTP-wide assessment

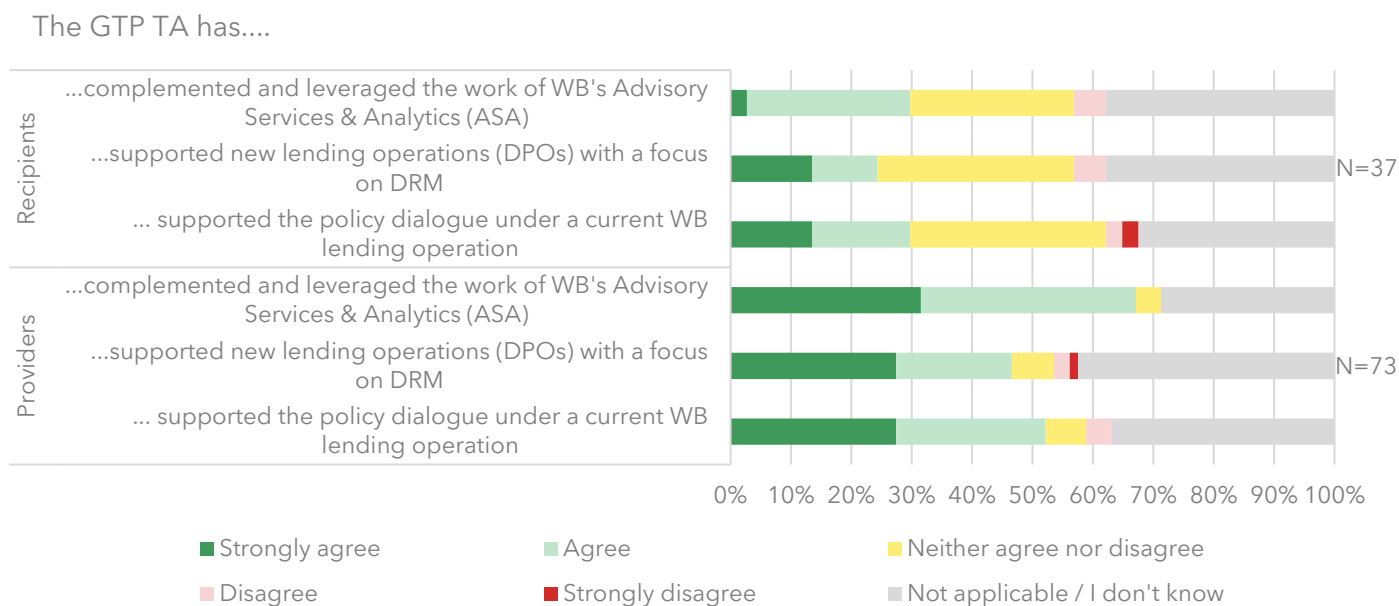
Internal Coherence

GTP project teams and the GTP Secretariat all have an eye for complementarities and synergies with other Bank operations in the tax domain as well as the fiscal planning and budgeting domain. Program documents and annual reports provide evidence that the GTP Secretariat actively maps complementarities with other tax-related activities within the WB on an annual basis, along with complementarities with other development partners (DPs). In addition, interviews suggest that project teams often look for synergies with concessional lending operations. This is helped by the fact that in some cases the TTL of a GTP project is the same as the TTL for the concessional lending operations (e.g., Kyrgyz Republic).

Many survey respondents were unable to comment on internal coherence, but those who did generally saw GTP TA as coherent and even synergetic with other WB interventions (Figure 2.3). Neither in the survey nor in the case studies did we see signs of insufficient coordination with WB concessional lending operations or Advisory Services and Analytics (ASAs). On the contrary, nearly half of provider-respondents noted that GTP TA had supported new concessional lending operations with a focus on DRM, and more than 40 percent of the surveyed TA recipients indicated to also be interested in a WB concessional lending operation (or budget support) with a focus on DRM. A recent study by the WB's Independent Evaluation Group (IEG) found evidence that Technical Assistance increases the success of Development Policy Financing (DPF) operations, supporting the notion that TA can strengthen concessional lending operations. Similarly, our case studies revealed several examples where TA gave rise to concessional lending operations and vice versa.¹³

¹³ However, a risk highlighted by one internal WB stakeholder is that if GTP TA focuses too much on actively seeking out opportunities for WB concessional lending operations, GTP carries out work that should inherently be part of the Bank's core business and therefore should be carried out using the Bank's budget. In this person's view, activities focused on looking for concessional lending opportunities should be part of the concessional lending operations itself as a WB core business, rather than GTP TA. However, other WB stakeholders firmly agreed with this statement.

Figure 2.3 TA providers saw more synergies with other WB interventions than TA recipients



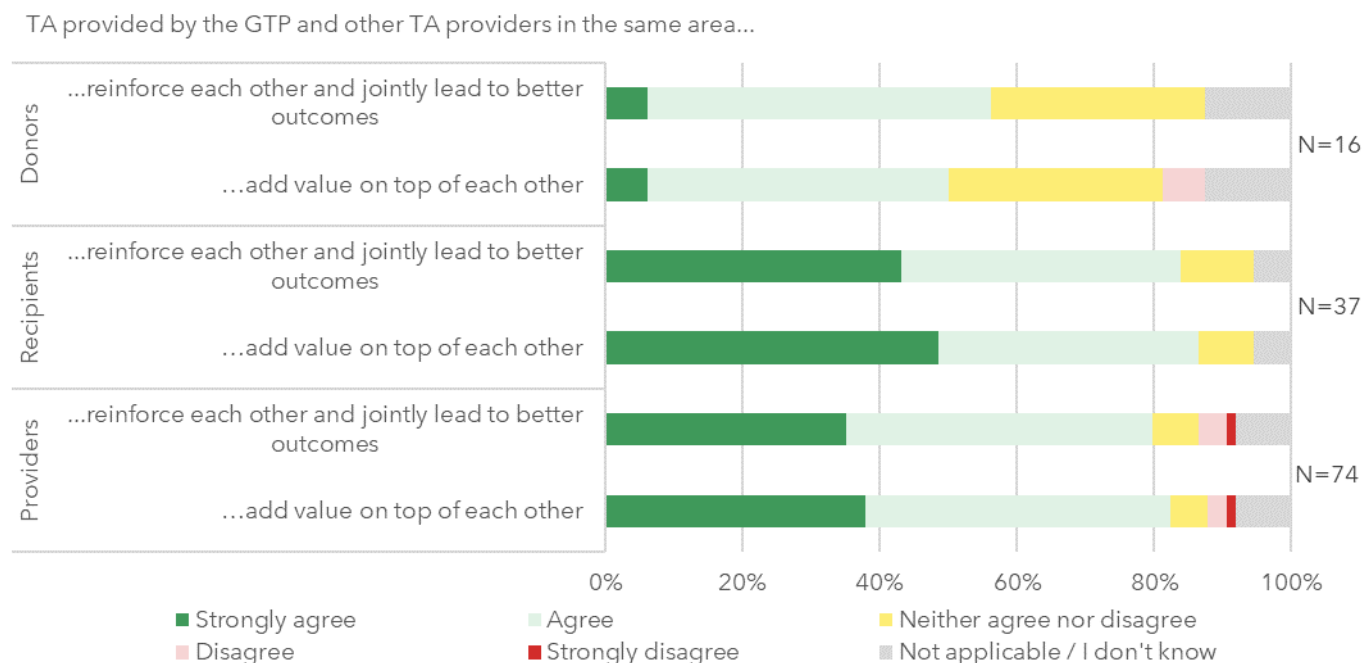
Source: SEO Amsterdam Economics, based on a survey of 110 stakeholders.

Several interviews with current or former WB representatives indicated that the GTP may be constrained somewhat due to the fact that not all WB teams see taxation as a key development tool. This is perhaps logical because taxation only constitutes a small part of all WB activities, whereas for the International Monetary Fund (IMF) it is a core part of its mandate (macro-fiscal stability). Another explanation offered for this is that many WB staff, including the board and senior management, may be more conscious of poorly designed and implemented taxation, which they tend to see as a development constraint, than of the positive impact that well-designed tax reforms can have. This could result in uneven internal support from senior WB managers. This may exacerbate the challenge for the WB to attract and retain highly skilled tax experts (which not only affects internal coherence, but also poses risks for effectiveness and efficiency). However, the recent adoption of taxation targets in IDA is a positive sign. According to the World Bank itself, "taxation has become prominently integrated into the broader macro-fiscal and debt management at the Bank," and the GTP supported the scale-up of ASAs with a focus on DRM.

External Coherence

Just as the GTP has an eye for complementarities with other WB activities, it also actively monitors complementarities with other developing partners, such as the IMF or the OECD. In the project proposal phase, the GTP asks TTLs to "describe coordination/consultation arrangements with other key Partners who are active in the area of DRM for that country". In addition, the GTP Secretariat performs an annual mapping exercise to identify tax-related projects, also based on the Platform for Collaboration on Tax (PCT) partners' Online Integrated Platform database and presents the results in its annual reports. According to the reported results of this mapping exercise, the identified external complementarities mainly entail links between GTP projects and technical assistance missions of the IMF. The annual reports provide an extensive overview of collaborations with development partners. Nonetheless, in some cases it is not clear from document reviews and interviews what concrete actions projects take to achieve synergies with DPs during project development or implementation, after complementarities have been identified.

Figure 2.4 In the survey, donor partners were least positive about the GTP's external coherence.



Source: SEO Amsterdam Economics, based on a survey of 127 stakeholders.

Based on a desk review, interviews and survey results, we conclude that overlap and conflicting policy advice is usually avoided, but that there is room to increase complementarities and synergies. Multiple stakeholders noted in interviews that GTP project teams typically establish good informal coordination channels with other providers at the country level, “since they have an incentive to do so”. However, it was also noted that this type of coordination is still too dependent on individuals and is not yet fully embedded in the overall strategy (despite the effort of the mapping exercise). Interviewees stated that external coordination has improved a lot compared to the past, especially with the IMF (particularly on a Medium-Term Revenue Strategy [MTRS] and at the global management level). This is confirmed by the survey, in which more than 80 percent of recipients and providers (strongly) agreed with the statement that the TA provided by the GTP reinforces and/or adds value on top of the TA provided by other TA providers in the same area (Figure 2.4).

2.2.2 Coherence: project-level assessment

Internal coherence

Case studies confirmed that internal coherence was generally largely or fully achieved (Figure 2.5), but that the linkages between TA and concessional lending operations could be further streamlined. In some case studies the GTP project and the WB concessional lending operations are clearly complementary, aided by the fact that the TTL of the project is also leading the concessional lending operations. In one case study, the GTP TA support focused on the reform of tax functional areas, tax law reform, taxpayers assistance/service and tax administration capacity building, while the tax component of the concessional lending operation focused on the development of the necessary IT systems to support those reforms. In another case study, the recipient government requested additional technical support to design and support the implementation of the longer-term DRM reform agenda

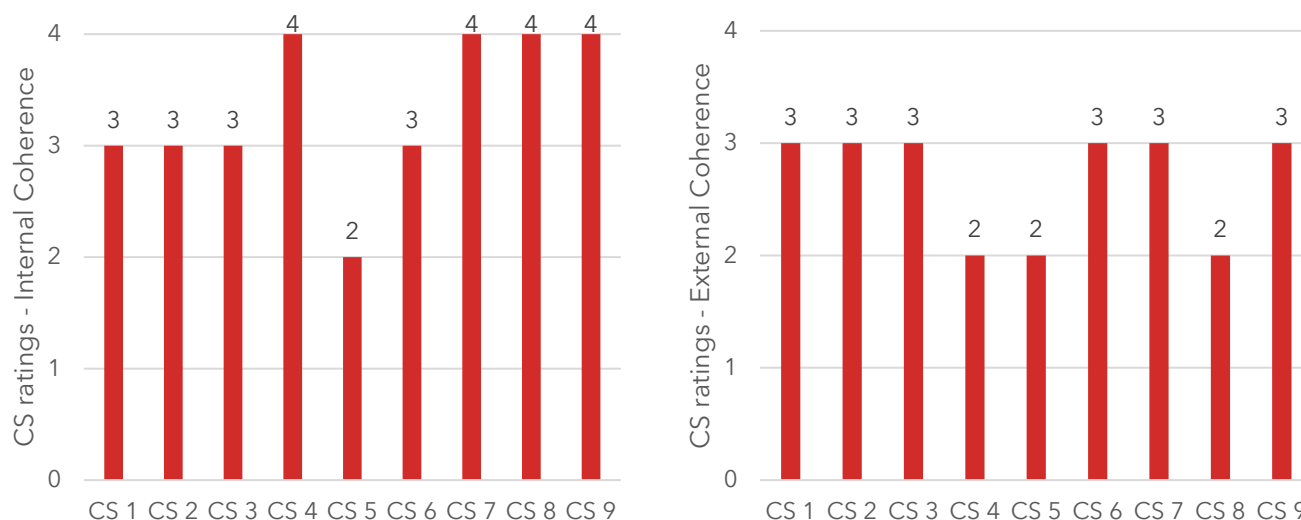
supported by the Bank's investment project. As these examples show, some country teams have actively tried to link GTP TA with concessional lending operations. The lack of information on coordination in some of the reviewed project proposals indicates that there is, in fact, no formal requirement in this regard. Efforts to coordinate with other WB activities therefore seemed to remain quite project and people-dependent, possibly reflecting the Bank's country-driven model.

External coherence

Based on multiple information sources (desk review, survey results and case study interviews), it seems clear that the GTP has generally coordinated at an acceptable level with other development partners and has largely avoided overlap with them in the case study countries. Most case studies rated external coherence as largely achieved, whereas a few case studies rated it as partly achieved (Figure 2.5). For example, there was an apparent agreement with the IMF in one case study country that only the WB would provide assistance on tax policy reforms, while both WB and IMF would contribute to advising on tax administration. The collaboration included IMF attendance at GTP-supported workshops, mutual briefings during country missions, as well as acting as "observers" during each other's missions. In another project, the various DPs produced a report to ensure that their CD activities did not overlap and helped each other through technical and legal support on their technical assistance programs. Project proposals generally elaborate on past and intended future collaboration with other development partners, indicating that the GTP has an eye for synergies with other DPs.

However, synergies and complementarities with other DPs could still be further improved. Both in case studies and in GTP-wide interviews, the evaluation team came across DP interviewees who were not well informed about the GTP and its various activities. This suggests that the GTP Secretariat and WB staff could still do more to provide information on the GTP and its activities to other DPs. Although the evaluation team acknowledges that external coherence should also be a responsibility (and is also dependent on the willingness) of the recipient authorities, in multiple cases a more systematic approach on DP coordination could have helped to strengthen synergies, for example through better sharing of documents, coordinating policy advice, and co-designing of TA activities. While GTP projects cards already identify many (potential) complementarities with other DPs, owing once again to the efforts of the GTP Secretariat, transforming these complementarities into synergies could be further supported by concrete actions by WB teams on the ground.

Figure 2.5 In the case studies, internal and external coherence were fairly good on average, with areas for improvement mostly in external coherence.



Source: SEO Amsterdam Economics. Note: 1 = not achieved; 2 = partially achieved; 3 = largely achieved; 4 = fully achieved.

2.3 Effectiveness

Despite COVID-19, most GTP-supported projects were effective in that they achieved their targets for outputs and short-term outcomes. However, longer-term outcomes were difficult to assess, since most projects were still ongoing and the GTP results framework does not yet include longer-term indicators.

2.3.1 Effectiveness: assessment of the results framework

The ToC of the GTP is large and complex, but has tailored micro-indicators for project-specific objectives. The GTP's ToC contains ten outputs (sphere of control) that are expected to jointly result in improved DRM and nine 'sub-impacts' (sphere of interest), acting via intermediate and long-term outcomes (sphere of influence) (see Appendix C). As per the ToC, tax reforms, policy amendments and increased organizational and technical capacity are instrumental to improving DRM. Other impact levels that the GTP has identified relate to issues such as transparency, tax avoidance and evasion, compliance, government expenditures, but also to the broader development areas such as equality/fairness and climate change. A selection of aggregated standardized indicators is used for monitoring at the Window and Program level.¹⁴

The GTP results framework is well developed and monitors project indicators, mostly defined as proposed, endorsed and adopted reform recommendations. The standard output and outcome indicators used for the achievement ratings in the results framework (see Table 2.2 and Table 2.3) mostly focus on proposed (rather than endorsed) tax reform recommendations. The results framework does not attempt to attribute long-term outcomes to GTP activities. It is said to be very difficult to do so in relation to tax, since outcomes can also be affected by many

¹⁴ Global Tax Program Umbrella Program Document, 2020, p. 21

other (external) factors. As stated earlier, results at the outcome level and especially the impact level can hardly be fully attributed to a single intervention, but a program should strive to accomplish outcomes and impacts nevertheless and make their contribution to a change explicit.

The logical framework (hereafter: logframe) that is elaborated upon in the program document shows an operationalized intervention logic. It translates outcomes to Key Performance Indicators that can be measured and monitored across the different windows.¹⁵ This is in contrast to the ToC, which is not very helpful in understanding the program's overall intervention logic since it does not distinguish between result areas/windows and distinct pathways that lead to certain outcomes and impacts. When drafting project proposals, the TTL, in collaboration with the GTP Secretariat, decides on the set targets and the outcome and output indicators that are most relevant to the project's objective. Impacts and most long-term outcomes are not part of the results framework since the GTP considers them "outside their sphere of control". However, project teams have committed to monitoring relevant micro-level indicators in addition to outputs and intermediate outcomes in the project results frameworks to provide context to our operations and inform management decisions.¹⁶ Table 2.1 shows the standard indicators that the GTP measures across projects.

Table 2.1 The GTP measures 3 key outcome and 8 key output standard indicators

Standard indicators	Description
Outcomes	
1.0.1	Number of reform recommendations (laws/regulations) endorsed
1.0.2	Number of recommendations to improve procedures/practices/standards adopted
1.0.3	Number of strategy/action plans adopted
Outputs	
1.0.4	Number of reform recommendations (laws/regulations) proposed
1.0.5	Number of recommendations to improve procedures/practices/standards proposed
1.0.6	Number of workshops/policy dialogues/consultations conducted
1.0.7	Number of capacity building events conducted
1.0.8	Number of strategy/action plans developed
1.0.9	Number of documents/reports/studies/notes prepared
1.0.10	Number of assessments completed
1.0.11	Number of databases developed

Source: GTP M&E Masterfile, Aggregated Results Framework

Donor partners seemed generally content with the results framework. Multiple donor partners indicated that, compared with other trust funds, the GTP's reporting is good, transparent and user friendly (e.g., can be used by donor partners for their own reporting, for instance to their own government). One donor partner noted that they particularly valued being able to compare indicators over time and compare and aggregate them between projects.

¹⁵ The GTP Umbrella Program Document (2020, p. 66) states that "Window 2 will often be scaled up through WB concessional lending operations. This will enable the GTP to achieve results that are not included in this logframe." For more on the synergies between GTP-supported projects and concessional lending operations, see the previous chapter on (internal) coherence.

¹⁶ Global Tax Program Umbrella Program Document, 2020, p. 64

However, the definition of “endorsed” and “adopted” in the results framework is not sufficiently clear. For example, it is not clear whether written endorsement or formal adoption by the government/parliament is required, nor whether adoption implies that implementation will definitely take place or will still be dependent on uncertain processes at the government level. Overall, the number of “proposed” indicators exceeds the number of “endorsed” indicators, as Table 2.4 shows. The evaluation team assessed progress of projects based on the available indicators in the results framework (see Table 2.1). These mostly deal with recommendations on tax policy reforms and not so much with the actual implementation.

Indicators on the implementation of reform recommendations are not yet included, since they are seen as “aspirational” and beyond the GTP’s control. The GTP Program Document further notes that “most expected long-term outcomes and especially impacts included in this results chain are outside the Program’s control, and therefore, will not be part of the projects’ results framework.” Since longer-term outcomes (e.g., related to the implementation of reforms) are seen as “aspirational”, they are not included in the frameworks that projects report on, and do not feature in the assessment of projects results. Considering the effects of COVID-19 and the fact that many projects have not yet been completed, it is not surprising that results on aspirational outcomes were not yet available during this mid-term evaluation. However, during project-level assessments the evaluation team did find preliminary signs of achievement of these aspirational outcomes, for instance regarding the implementation of a policy reform. Also, the program does actively monitor if recommendations are implemented during the projects’ lifetime.¹⁷

The results framework could be further improved by including longer-term outcomes related to reform implementation. The evaluation team acknowledges that there are often delays in getting reforms implemented, and that external “push and pull” factors can have a great effect. However, including outcomes in the results framework that are outside of the GTP’s sphere of control is not a problem, as long as the GTP shows how it has contributed to changes and outcomes, acknowledging the effect of external factors and influences. Given that the extent to which reform recommendations are implemented is highly informative and greatly affects the GTP’s potential to have a long-term impact on strengthening tax systems, we would strongly recommend monitoring the extent to which reform recommendations are implemented (long-term impact monitoring (post-completion) is currently not foreseen by the GTP). Considering the WB’s strong local presence, it is actually well positioned to monitor the implementation of reforms also beyond the scope of the projects.¹⁸ It would be useful if the GTP Secretariat could explore whether a certain share of the GTP budget could be allocated to such follow-up monitoring activities.¹⁹

The focus on numbers of outputs (activities, recommendations) has its pros and cons. On the one hand, such indicators are appealing because they can be aggregated within and across projects, which is something that donor partners have indicated to appreciate. On the other hand, this focus on numbers can be misleading, because: 1) aggregating numbers of outputs or outcomes implicitly treats each output or outcome as having equal weight (equal importance). However, some recommendations may be much more relevant or effective than others in terms of their potential effects on revenue collection; and 2) focusing on the number of outputs (e.g. documents or

¹⁷ For example, the recommendation to rewrite the tax code in Tajikistan was endorsed in FY20, and in FY22 the tax code was approved. Outcomes are claimed when recommendations are endorsed. If adoption or implementation occurs at a later stage, the WB does not claim the same outcomes again.

¹⁸ The GTP was originally scheduled to end in 2024, but has been extended until FY28. This can be an opportunity to monitor implementation of reforms beyond the lifetime of the projects that end in FY24.

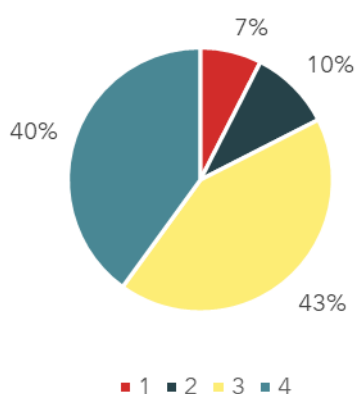
¹⁹ The evaluation team was told that there was a budget line for this type of follow-up monitoring activities in the previous program under the umbrella of the IFC, which suggests that allocating budget for follow-up activities may be possible within the World Bank Group.

recommendations produced) could provide incentives to divide some outputs into multiple outputs (although the GTP Secretariat makes an effort to count related outputs as one output). Using hybrid indicators that allow for quantitative and qualitative assessment and reporting would allow the narrative and success factors behind adopted and endorsed recommendations to be revealed, which would support learning and program management.

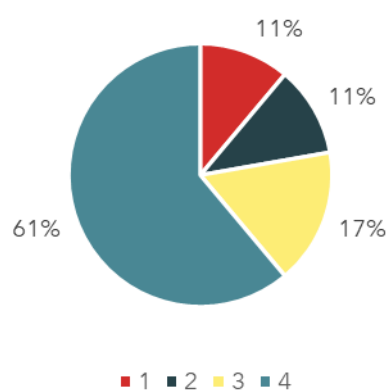
While many outcomes and outputs could not yet be rated because of project timelines in the near future, the GTP results framework illustrates that a majority of the rated outputs and outcomes have largely or fully satisfactory achievement scores. While 11 percent of the GTP project outputs was not rated, 83 percent of the remaining outputs was rated (largely or fully) satisfactory, as shown in Figure 2.6. Some indicators had no inception-to-date targets by June 2021, so for them it was not possible to assess progress. With regard to GTP project outcomes, almost 60 percent was not rated. Three-quarters of the rated outcomes had a fully satisfactory score as of June 2021. Overall, three-quarters of the rated outputs and outcomes were rated (largely or fully) satisfactory as of June 2021.

Figure 2.6 Three-quarters of the rated outputs and outcomes were rated (largely or fully) satisfactory as of June 2021.

a.) Distribution of output ratings across GTP projects



b.) Distribution of outcome ratings across GTP projects



Source: SEO Amsterdam Economics based on GTP M&E Masterfile (June 2021). Note: For aggregation purposes, average ratings were rounded up to the next whole number. Not all outputs and outcomes were rated, since many projects are ongoing and their targets are to be reached in the upcoming phases only.

Table 2.2 Overview of targets, results and achievement rates, FY18-FY21.

Standard indicators		FY 2018			FY 2019			FY 2020			FY 2021		
		Target	Result	Achievement rate %	Target	Result	Achievement rate %	Target	Result	Achievement rate %	Target	Result	Achievement rate %
WINDOW 1													
OUTPUTS													
1.1.1	Number of new modules/tools developed	0	0	N/A	4	3	75%	10	10	100%	7	12	171%
1.1.2	Number of documents/reports /studies/notes/guidelines/training materials prepared	0	0	N/A	2	3	150%	4	1	25%	9	6	67%
1.1.3	Number of workshops/policy dialogues/consultations conducted	0	0	N/A	0	0	N/A	1	1	100%	4	2	50%
WINDOW 2													
OUTCOMES													
2.2	Number of reform recommendations (laws/regulations) endorsed	0	2	100%	15	17	113%	19	14	74%	9	17	189%
2.3	Number of recommendations to improve procedures/practices/standards adopted	0	0	N/A	0	1	100%	4	11	275%	17	11	65%
OUTPUTS													
2.2.1	Number of reform recommendations (laws/regulations) proposed	0	11	100%	36	42	117%	35	33	94%	48	43	90%
2.3.1	Number of recommendations to improve procedures/practices/standards proposed	0	3	100%	12	14	117%	26	21	81%	29	34	117%
2.8.1	Number of capacity building events conducted	0	14	100%	19	23	121%	43	36	84%	74	40	54%
WINDOW 3													

OUTPUTS													
3.1.1	Number of KM documents developed	0	0	N/A	0	0	N/A	5	5	100%	6	14	233%
3.1.2	Number tools/modules/training materials	0	0	N/A	1	2	200%	5	2	40%	0	1	100%
3.1.3	Number of documents/reports/studies/notes prepared	0	0	N/A	0	0	N/A	0	0	N/A	0	2	100%

Source: GTP M&E Masterfile, Aggregated Results Framework.²⁰

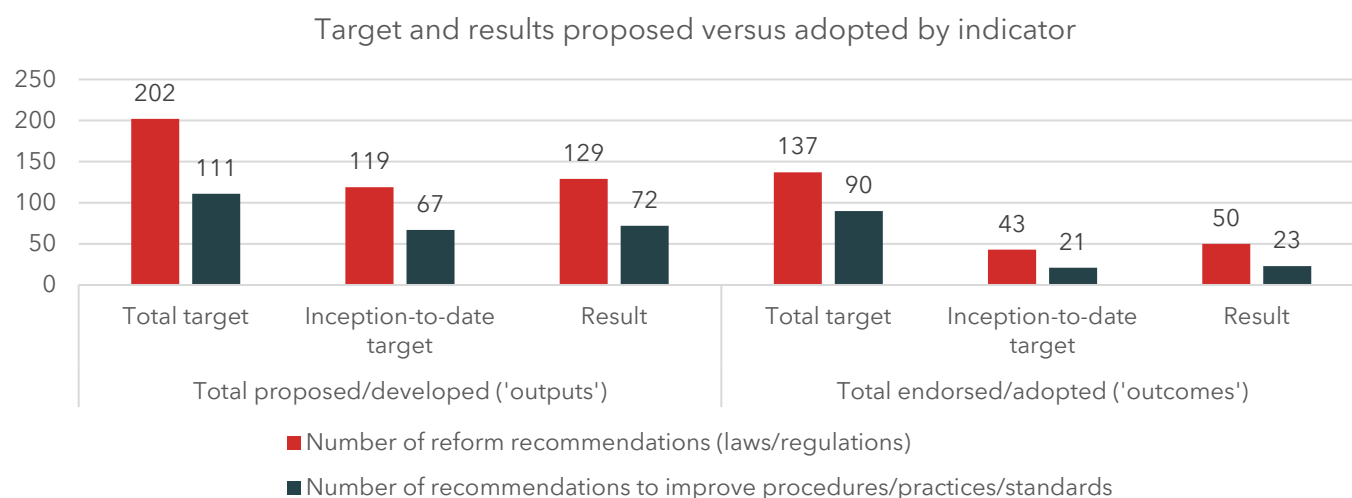
Table 2.3 Overview of inception-to-date targets, results and achievement rates as of June 2021.

As of June 30, 2021 (end of FY21)	Inception-to-date targets	Inception-to-date results	Achievement rate %
OUTCOMES			
1.1 Number of reform recommendations (laws/regulations) endorsed	43	50	116%
1.2 Number of recommendations to improve procedures/practices/standards adopted	21	23	110%
OUTPUTS			
1.1.1 Number of reform recommendations (laws/regulations) proposed	119	129	108%
1.1.2 Number of recommendations to improve procedures/practices/standards proposed	67	72	107%
1.1.3 Number of capacity building events conducted	136	113	83%
1.1.4 Number of public goods produced	36	35	97%
1.1.5 Number of KM documents and training tools prepared	17	24	141%

Source: GTP M&E Masterfile, Aggregated Results Framework.

²⁰ For some indicators, a GTP achievement score of 100% reflects a result without a specific target being assigned to it. An example is the achievement rate for indicator 2.2 in FY 2018.

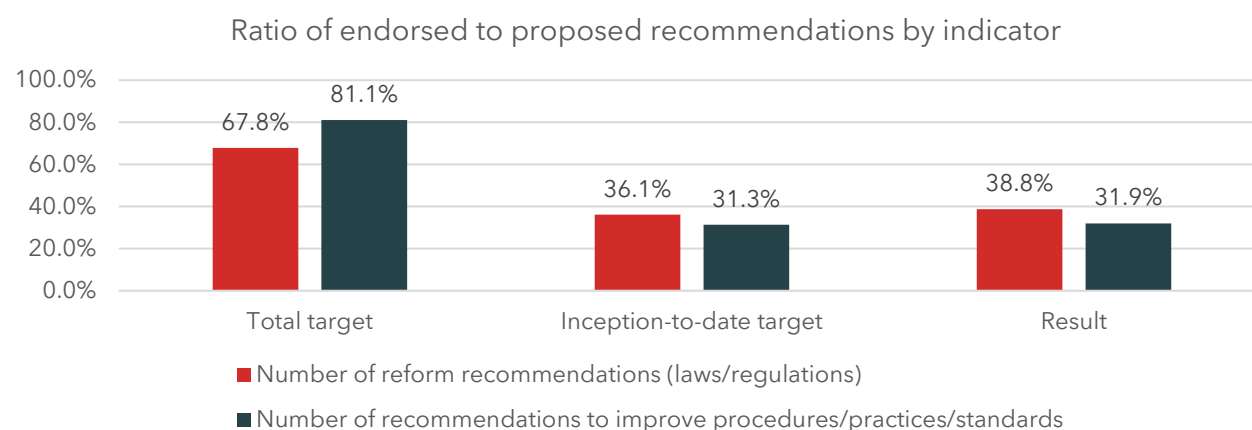
Table 2.4 The number of endorsed/adopted reforms was lower than the number of proposed reforms, as expected



Source: SEO Amsterdam Economics based on GTP M&E Masterfile (June 2021).

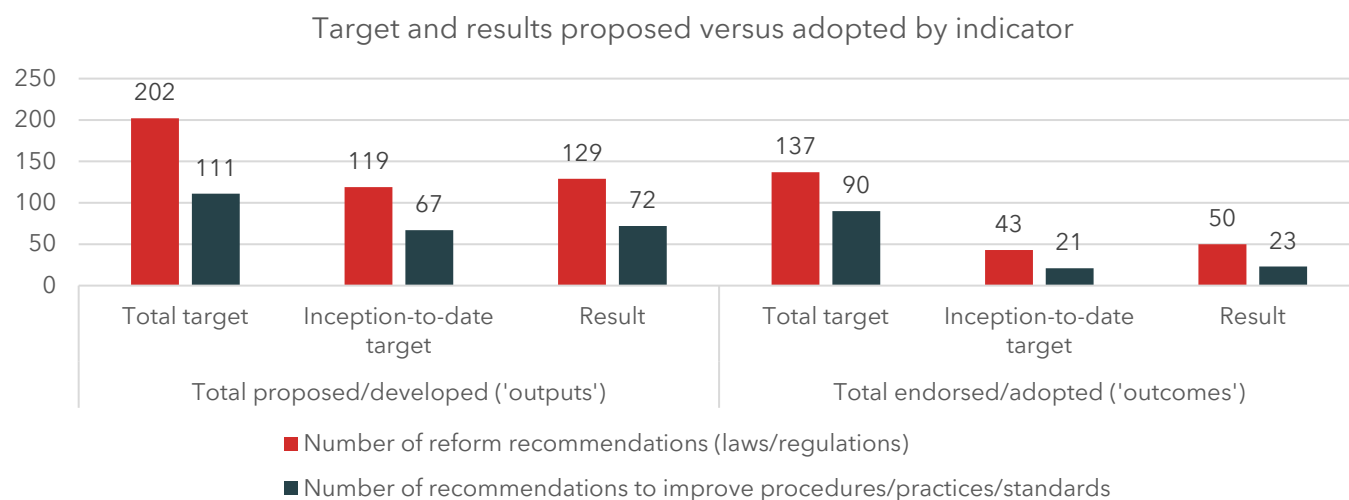
The ratio of endorsed to proposed reform indicators was in line with M&E targets, and both the proposed and endorsed reform targets exceeded the inception-to-date targets. The resulting percentage of proposed recommendations endorsed was in line with the inception-to-date targets (Figure 2.7), and as of June 2021 the data even showed overachievement of targets for both types of recommendation indicators (Figure 2.8).

Figure 2.7 The ratio of endorsed to proposed recommendations was in line with inception-to-date targets.



Source: SEO Amsterdam Economics based on GTP M&E Masterfile (June 2021).

Figure 2.8 As of June 2021, the M&E data illustrated overachievement of targets.



Source: SEO Amsterdam Economics based on GTP M&E Masterfile (June 2021)

2.3.2 Effectiveness: GTP-wide assessment

Effectiveness of program management

The GTP Secretariat has generally managed the GTP well. One indication is that most GTP projects achieved most of their targets in a timely manner. TTLs were sometimes thought to have set ambitious targets, especially in relation to the number of training sessions/trainees. The GTP Secretariat assisted TTLs with setting more realistic targets, monitored project performance, and reached out in case projects were faced with delays, to remind TTLs of their targets. COVID-19 naturally led to delays in many activities and projects, and countries were hesitant to proceed with tax reforms during the COVID-19 pandemic, leading to different in-country prioritizations. This meant that budget execution was low (in part due to travel cost savings) and that it was difficult to achieve all intended targets. The GTP Secretariat responded to this situation by allowing adaptations to the sequence and implementation of certain activities.

The progress reports adequately reflect the progress and delivery of activities. Project cards that trace progress, along with real-time updates, are used by the GTP Secretariat to prepare the semi-annual progress reports for the Steering Committee and the annual reports. The tracking of progress and writing of project reports or cards is a WB and GTP requirement and TTLs will monitor results indicators, sometimes complemented with custom indicators for the GTP projects.²¹ In practice, the M&E specialist of the GTP Secretariat plays a vital role in supporting TTLs with reporting and tracking the progress of each project, which then feed into the semi-annual and annual reports. Donor partners generally appreciate these reports and assess them as very accessible and of better quality than reports from similar organizations. One of the donor partners indicated that a narrative or story behind certain achievements would be appreciated (rather than only reporting achievements on numbers of documents and recommendations produced).

²¹ World Bank (2018) Bank Directive/Procedure: Advisory Services and Analytics (ASA), including Reimbursable Advisory Services (RAS), p. 7.

Dissemination of knowledge and information generated through the GTP was considered an essential step, but has not yet fully materialized. Results developed under Window 1, combined with lessons from country-level implementation under Window 2, and research output under Window 3 should be communicated internally to the WB and beyond. This was considered “a vital part of expanding the impact of the GTP and creating a multiplier effect beyond the countries...”²² The plan was to organize webinars and policy notes to stakeholders such as internal WB, the IMF, OECD, in-country stakeholders, and Non-Governmental Organizations (NGOs) and Civil Society Organizations (CSOs) that work on tax-related issues. However, this component has not yet effectively materialized in the sense that the GTP disseminates knowledge and information on a broad scale. Multiple stakeholders did indicate that dissemination of knowledge had recently improved, though. Targets set in the results framework for (externally oriented) reports and policy dialogues under Windows 1 and 3 show good achievements, albeit sometimes based on low or even no targets set.

Effectiveness of program strategy

The GTP has responded effectively to emerging issues. As previously indicated in the paragraph on relevance, the GTP has ensured alignment with IDA 20 and thus commitment to address issues such as climate change and gender equality, both thematically and within the project portfolio. With regard to the issue of gender, the GTP’s place within the WB enables a good track record and expertise on this subject. There is increased demand from countries to understand how fiscal policies, particularly revenue policies, can address climate change. Traditionally, there is a gap between demand and supply in this thematic area, and while the GTP recently incorporated projects on green/environmental tax it is unclear whether WB has sufficient in-house expertise to fully deliver on this climate change demand.²³ The COVID-19 pandemic has given further momentum to the GTP to incorporate emerging issues, and as such contribute to more fair, equitable and inclusive growth. Moreover, besides gender equality in taxation, the subject of health (or sin) taxes was incorporated thematically as evidenced by two workstreams on these topics. As of 2021, GTP teams collaborate with country counterparts in 15 countries to respond to government demands in the field of health tax policy, among other things.²⁴ Also viewing COVID-19 as an emerging issue, the GTP effectively and timely mobilized a temporal fiscal pillar to help countries address the economic and fiscal impact of the crisis, supporting 31 projects in 27 countries in FY21.²⁵

As previously introduced and elaborated upon in the section on coherence, strong linkages with the WB’s concessional lending operations increase the GTP’s effectiveness (and vice versa). Surveyed stakeholders positively assessed the GTP’s effectiveness and its potential contribution to results in the long term. Survey respondents were positive about the GTP’s contribution to the tax policy dialogue, the adaptation of technical recommendations and concrete reforms. Survey respondents were also positive about the GTP’s ability to contribute to higher revenues. Most of them foresee higher revenues only in the future, though, and the majority of providers does not see this happening in the short run.

More than 90 percent of the surveyed recipients indicated that the overall effect of the GTP TA was according to (or above) their expectations. Where applicable, survey respondents were positive about GTP TA’s contribution to better informed decisions, new guidelines and improved processes. Both recipients and providers saw GTP TA

²² Global Tax Program Umbrella Program Document, 2020, p. 42

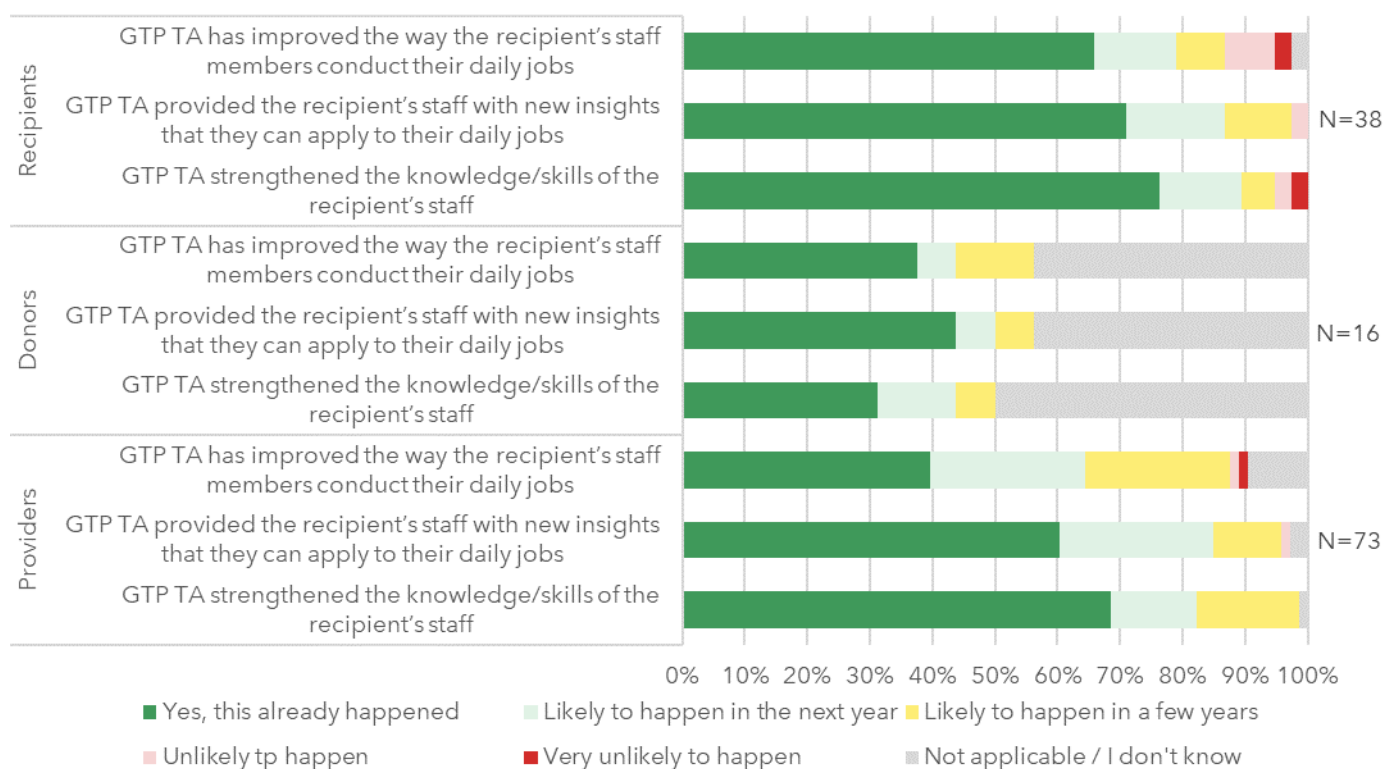
²³ Early in FY21, the GTP introduced two environmental tax projects to address this demand. The GTP will continue expanding the support by adding new projects. In April 2022, four new projects with a focus on climate change and sustainability were approved during the last call for proposal process

²⁴ GTP (2022) Donor Semi-Annual Update July-December 2021, p. 9

²⁵ GTP (2021) FY21 Annual Progress Report, p. 13

as contributing more to enhancing knowledge of the recipient's staff than to improving everyday practices, as illustrated in Figure 2.9.

Figure 2.9 Respondents saw GTP TA as contributing more to enhancing knowledge than to improving everyday practices.



Source: SEO Amsterdam Economics, based on a survey of 127 stakeholders

According to multiple stakeholders, the GTP's effectiveness could be increased if the Bank would be able to retain more high-level tax experts at the center. In general, it has proven challenging for the GTP to attract truly high-level tax experts -- the labor market is seen as "especially tight" for these skills and a viable career path for such experts may not always be obvious in the Bank, which often requires rotation. Moreover, hiring more high-level tax experts will also increase costs. Nevertheless, GTP management informed the evaluation team that possible measures to attract more tax experts at the center were being discussed with Senior Management at the time of writing this report.

2.3.3 Effectiveness: project-level assessment

Many case study projects are on track to meet objectives according to the output and outcome indicators in the M&E framework, a view supported by stakeholders. In the case studies, GTP projects were found to be effective in, for example, assisting authorities with tax and customs administration and facilitating a high-level dialogue between (federal) member states. Another major achievement is the development and adoption of an MTRS, which has contributed to national government objectives. Lastly, multiple case studies illustrate

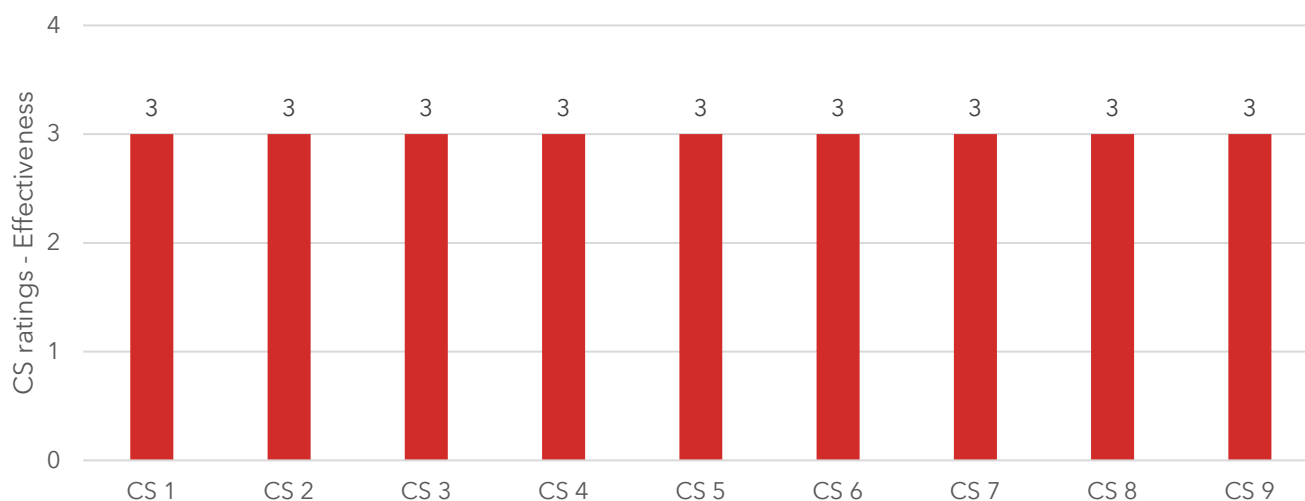
overachievement of inception-to-date targets of project outcomes and outputs, e.g., in the form of more reforms proposed or endorsed than originally targeted.²⁶

Most case studies indicated that moving to online missions, where possible, did not result in a loss of quality, although project documents and interviewed stakeholders signaled a few challenges. Connectivity issues (e.g., unanticipated power and network outages), difficulties in engaging the participants, and IT staff rotation were seen as the main impediments for effectively carrying out online missions. Nevertheless, moving to online missions helped increasing the frequency of contacts between the providers and the recipients, which is also likely to have positive spill-over effects on sustainability.

Nevertheless, the overall effectiveness was negatively affected by the COVID-19 pandemic as certain activities were postponed or even cancelled. This was mostly the case for capacity building activities, leading to related output targets not being met as planned. As a result, many case study projects had lower levels of achievement.²⁷

However, most projects reacted flexibly and successfully to constraints created by the pandemic. In some case studies, project teams were able to alleviate the pandemic's limitations by prioritizing and progressing with remote activities, while capacity-building activities were postponed. GTP projects were also found to be effective in assisting national authorities with their fiscal responses to the COVID-19 pandemic. As an example, the flexibility of the GTP program enabled the provision of just-in-time services to accommodate the changing need of national tax services during the pandemic. As a result, revenue increased in both 2020 and 2021 for one case study.

Figure 2.10 The effectiveness of case study projects was seen as satisfactory for all case studies.²⁸



Source: SEO Amsterdam Economics. 1 = not achieved; 2 = partially achieved; 3 = largely achieved; 4 = fully achieved.

²⁶ Overachievement of output and outcome targets could be overshadowed if the number of reform recommendations endorsed exceeds the recommendations that were adopted, which happened in one case study. This distinction between reforms endorsed versus adopted leads to the question whether a project can be considered effective already if recommendations are endorsed rather than directly adopted.

²⁷ For some case studies this low level of achievement could be explained by the rather recent start date of the project and/or relatively high number of targets set for future years.

²⁸ Assessments of effectiveness in case studies were based on interviews, WB ratings of project outcomes, and document review.

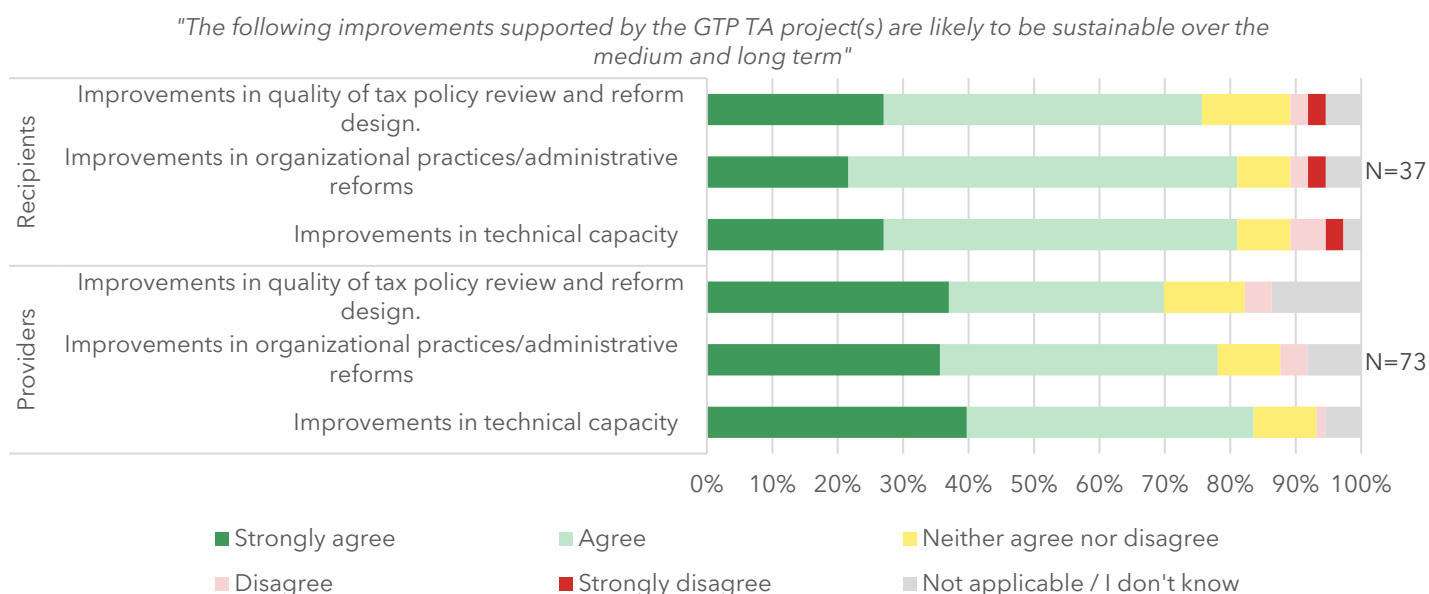
2.4 Sustainability

The WB's strong country presence and its ability to provide additional funding through concessional lending operations bode well for ensuring the sustainability of GTP results. However, risks to sustainability include high staff rotation, limited absorption capacity, political instability, and the short-term focus of the results framework.

2.4.1 Sustainability: GTP-wide assessment

In interviews and survey responses, GTP stakeholders were generally positive about the likelihood that GTP results will be sustainable over time. As Figure 2.11 shows, around 80 percent of the surveyed recipients and providers believed that the various improvements targeted by GTP-supported TA projects were likely to be sustainable over the medium to long term. However, the survey also showed that stakeholders tended to be more positive about the sustainability of improvements in individual technical capacity than about the longer-term improvements in organizational practices or tax policy review and reform design.

Figure 2.11 Respondents were positive about the sustainability of GTP results in the medium and long term



Source: SEO Amsterdam Economics, based on a survey of 110 stakeholders

The strong local presence through WB country offices increases the GTP's potential to achieve sustainable results. A key advantage of the GTP, often mentioned during interviews, is its ability to leverage the WB's strong country presence through country offices. One advantage of having local country teams present is that they can continue to monitor reform implementation and provide follow-up support where needed (provided that there is good knowledge transfer between GTP project teams and WB country teams). Another advantage is that the WB's strong country presence can contribute to developing closer ties between the provider and recipient of TA, which can positively contribute to reform implementation both during and after a GTP-supported project. A third

advantage, also pointed out under “internal coherence”, is that WB country teams can improve synergies between TA projects and WB concessional lending operations. An often-heard limitation of TA projects in general is the absence of funding for non-TA support that is fundamental to reaping the full benefits, such as investments in IT systems (software or hardware) that are often fundamental to improving the capacity of tax administrations. Moreover, engagement in concessional lending operations typically means that both the WB and the recipient country itself commit to a longer-term relationship, which in turn can give rise to additional requests for follow-up GTP projects (see also the discussion on synergies between TA and concessional lending operations under “internal coherence”).

Nevertheless, stakeholders also pointed out several risks to the sustainability of GTP-supported projects.

First, both interviewees and survey respondents sometimes questioned the country’s absorption capacity in terms of political support, staff expertise and resources, especially recipients themselves. Second, other often-mentioned impeding factors were the high staff rotation on the side of the local authorities and technical difficulties (e.g., IT capacity constraints). Third, while stakeholders agreed that TA projects should be seen in a longer-term perspective, their (often) short duration was flagged as a concern. Fourth, it was mentioned that external shocks such as COVID-19 can imply that project teams need to adapt TA activities to the more urgent needs of governments, which could jeopardize or delay longer-term structural reforms.²⁹

There are also risks to the sustainability at the strategic level:

- First, as discussed in our Effectiveness chapter, the GTP at present does not monitor long-term outcomes related to reform implementation, as these are seen as “aspirational” and beyond the GTP’s control. The risk of not monitoring these longer-term effects is that project teams may focus too much on achieving “endorsement” and “adoption” of reforms (e.g., adoption of a MTRS by Parliament) without necessarily including sufficient follow-up support on implementation of these reforms (e.g., implementing legislation).
- Second, covering a large number of countries and a large number of issues could potentially reduce the systemic and sustainable impact that the GTP could have at the individual country level. The demand for support is expected to remain high (including from “new” countries) and donor commitments are not expected to decrease (as a result of the ATI), therewith creating opportunities for expansion. However, expansion of the GTP to other countries (i.e., increasing its width) should not come at the expense of achieving a deeper impact within a particular country (depth). Given human and organizational constraints, a too strong focus on expansion to more countries poses the risk of more shallow support and thus of continued impact.

2.4.2 Sustainability: project-level assessment

In our case studies, sustainability typically had the lowest score, partly because it was too early to judge, but also because risks to sustainability were identified. As Figure 2.12 shows, six out of nine case studies rated sustainability as a “2” (partially achieved). On the one hand, this is logical, since many projects had only just finished or were still ongoing at the time of the evaluation. On the other hand, the case studies did indicate various risks to the (future) sustainability of the results.

²⁹ Our case studies did not yield much evidence on this point. Although output targets were delayed, the expected results (outcome targets) by the end of the projects did not change substantially. Many projects requested a “no-cost extension”, suggesting that they expected outcome targets still to be reached, but later than originally envisaged (the future will tell whether or not this will indeed be the case).

Two external risks to sustainability were mentioned in particular:

- High staff rotation on the side of the recipient was mentioned in about half of the case studies and was therefore seen as one of the major risks to the sustainability of GTP-supported projects. In two other case studies, low staff rotation was seen as a reason for high sustainability. This issue is typically outside of the direct sphere of control of the GTP project team.³⁰ Nevertheless, the GTP could consider working with WB country teams to identify opportunities for supporting the authorities with recommendations on measures to reduce staff rotation (e.g., as part of WB support for public sector reform in the country).
- Political unrest was seen as posing risks to sustainability in at least two case study countries. On the one hand, these risks can be expected when operating in low-income countries and fragile states, and they are typically outside of the sphere of control/influence of the WB. On the other hand, there were case studies where the risk of political unrest was high and predictable to some extent (e.g., it was known that elections were coming up). In these cases, the evaluation team felt that the risk of not meeting project targets could have been anticipated better and could have been mitigated by postponing or adjusting the project.³¹

In addition, the absence of a Medium-Term Revenue Strategy (MTRS) was flagged as a risk to sustainability in two case study countries. The purpose of an MTRS is typically to help a country with guiding its revenue reforms. The absence of such a medium-term strategy implies risks to sustainability because there is no guarantee that short-term reforms will continue to be followed up in the medium term. While this is partly outside of the control of the GTP, the GTP itself can take measures to reduce this risk, i.e., by seeking more often to support country authorities with developing an MTRS (typically in coordination with the IMF).³²

The case studies also yielded several examples of factors that promote sustainability of results:

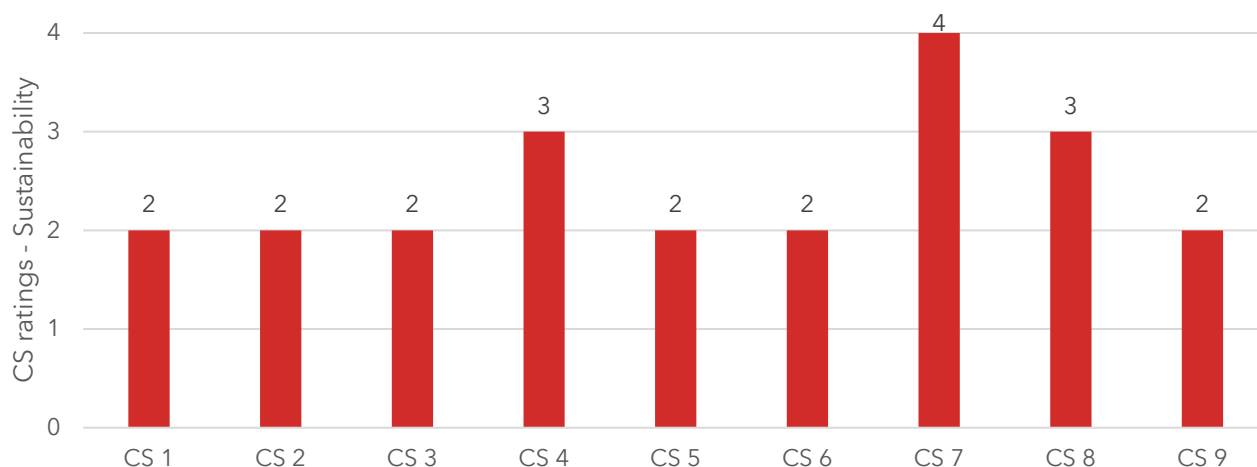
- High ownership and relevance. Sustainability of results was found to be promoted through high ownership and relevance in at least four case study countries. In one case study, it was mentioned that ownership was generally good at the higher (senior) level, but that there was room for improvement at the lower levels.
- A change of the legal framework was mentioned multiple times as a reform that bodes well for sustainability. For example, a formal adoption of tax code amendments makes it more likely that these amendments will continue to be in place over a longer time horizon.
- A combination of various types of support was found to also contribute to improved sustainability. Combining TA with policy advice and/or concessional lending operations naturally increases the sustainability of the results, since it increases the likelihood that reform recommendations are implemented. In two case studies, the combination of GTP-supported TA with WB concessional lending operations was mentioned as a key driver to sustainability by all parties involved.

³⁰ However, in one case the WB itself was actually mentioned as a cause of staff rotation, as higher salaries at the WB drew some local staff away from local government.

³¹ The Bank made a corporate decision to engage more heavily in fragile states, and this has continually been supported by the donor community through IDA replenishments. A consequence is that at times, political unrest (e.g. in Afghanistan and Myanmar) may resurface, jeopardizing the status of tax projects as well as many other development objectives.

³² The GTP team informed the evaluation team that there was a specific GTP-funded ASA in the past that supported MTRS development, which closed without meeting all its targets. It would be useful to assess why this project was ended and why it did not meet all its targets, given the fact that having an MTRS in place is generally seen as an important factor that bodes well for the effectiveness and coherence of tax reforms.

Figure 2.12 Sustainability is most at risk in the case studies, with “2” being the most frequent score.



Source: SEO Amsterdam Economics. 1 = not achieved; 2 = partially achieved; 3 = largely achieved; 4 = fully achieved.

2.5 Efficiency

The GTP generally manages its projects efficiently and in a timely manner. The evaluation team did not find evidence of any major issues regarding cost-effectiveness. However, operational efficiency could be further improved.

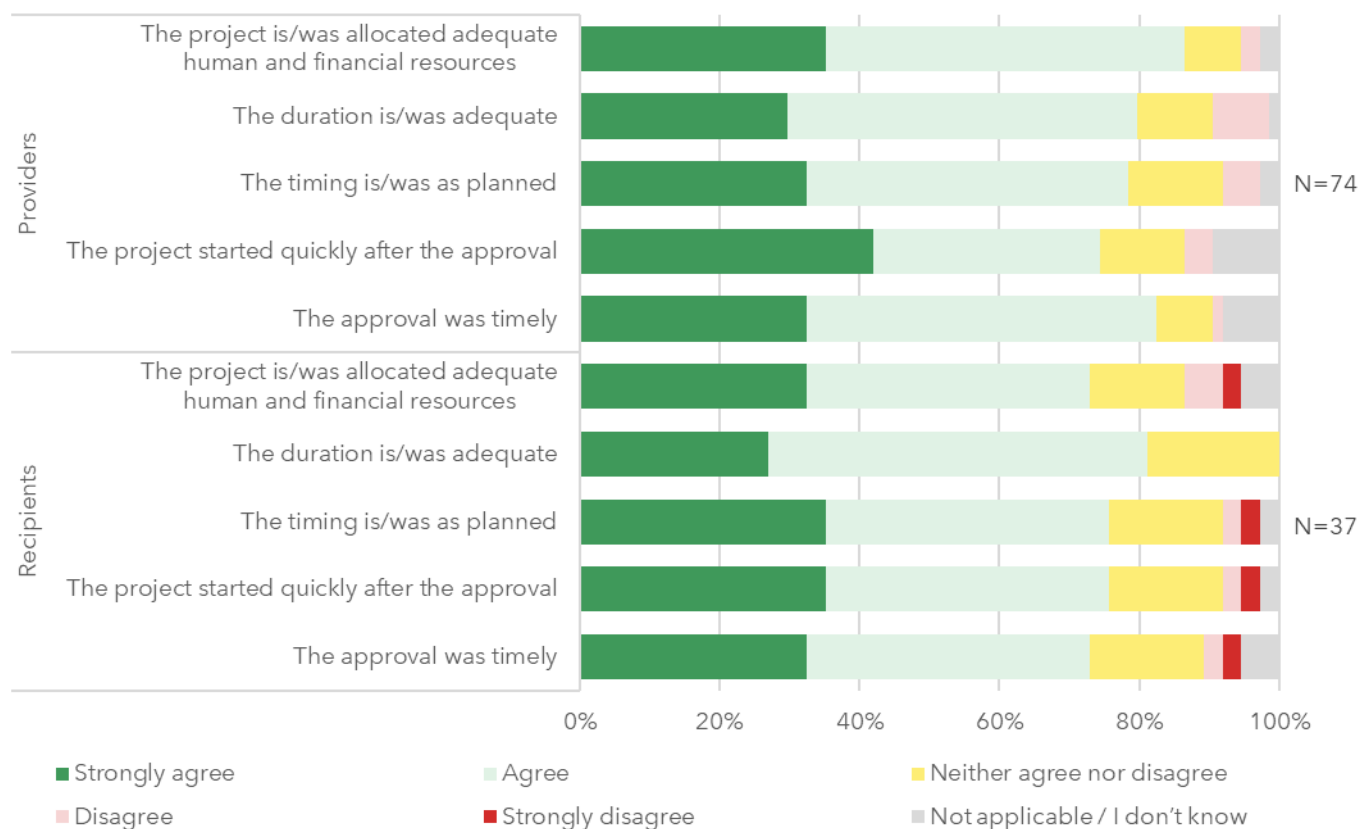
2.5.1 Efficiency: GTP-wide assessment

Recipients and providers were generally satisfied with the organizational efficiency of the GTP, despite some minor concerns about timeliness and duration (Figure 2.13). According to multiple interviewees, COVID-19 was the main cause of issues with timeliness, since many activities were delayed and projects asked for extensions. While case studies showed that project teams often managed to replace in-person activities with online activities, stakeholders also mentioned in interviews that this could have been done more efficiently and that other organizations were quicker than the GTP to move activities online and reallocate budgets in response to COVID-19. Confronted with these remarks, GTP stakeholders explained that “the system” does not easily allow to change previously approved budgets and that, due to the very nature of the TA supported by the GTP, some activities were difficult to move online (e.g., implementation of certain IT projects).

Stakeholders also mentioned some other operational inefficiencies stemming from organizational/governance mechanisms. These were, however, mostly related to the WB’s structure and therefore outside of the GTP Secretariat’s control. As pointed out by several stakeholders, the operational work of the GTP team was somewhat constrained by the WB’s structure. One issue in particular that was mentioned by at least two interviewees is the fact that taxation is not a core part of the WB’s mandate (as it is e.g., for the IMF) and that the WB does not have “a central tax unit” (e.g., like the Fiscal Affairs Department at the IMF). The fact that the GTP operates across different WB global practices (e.g., Governance, Macroeconomics Trade & Investment [MTI]) does not necessarily affect cost-effectiveness, but was seen by some as reducing its operational efficiency, since it requires additional coordination work for the GTP Secretariat. In addition, several stakeholders pointed out the relatively low number of specialized in-house tax experts at the WB. GTP management representatives acknowledged that this may be the

case but explained that attracting new high-level tax experts is not straightforward, given the tight labor market and career structures at the Bank. On the one hand, flexible access to much tax expertise can be garnered through the use of short-term consultants (STCs), who are also typically less expensive than full-time staff. On the other hand, developing and implementing procurement procedures for these STCs constitutes an additional administrative burden and does not necessarily guarantee timely availability of the required expertise for all GTP projects.³³

Figure 2.13 Recipients and providers were satisfied with the organizational set-up of GTP projects, despite some minor concerns with timeliness and duration.



Source: SEO Amsterdam Economics, based on a survey of 111 stakeholders

The GTP Secretariat monitors the “output efficiency” of GTP funded projects. The GTP defines “output efficiency” as the ratio between the “budget execution rate” and the “percentage of outputs achieved.” The budget execution rate is defined as actual expenditures to date relative to total planned expenditures. The output score is defined as outputs achieved relative to total outputs. An advantage of this measure is that it allows to disentangle efficiency from confounding factors (country or mission-specific factors, e.g., when higher costs of travel to country A than to country B make mission A more expensive than mission B).

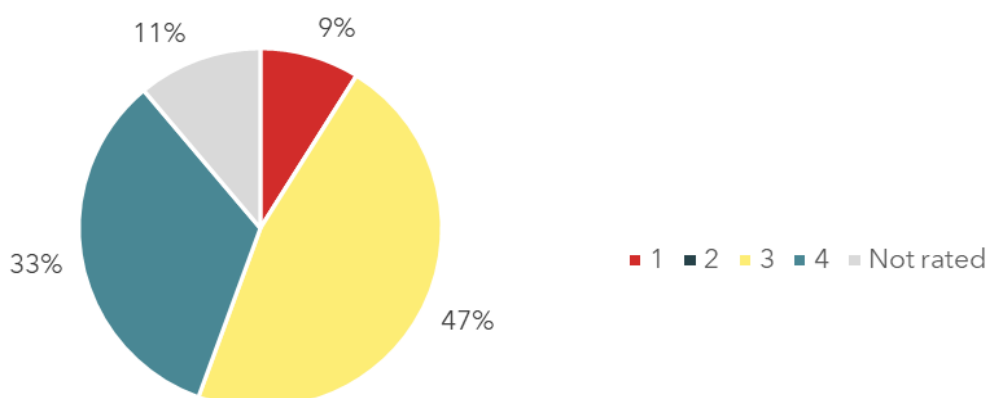
While monitoring output efficiency is a strong aspect of GTP project management, the definition of output efficiency could be further improved. According to the GTP’s own estimations, the output efficiency, defined as the ratio between budget inception to date execution rate (numerator) and percentage of outputs achieved

³³ One stakeholder pointed out that competition among TTLs to obtain the best short-term consultants sometimes also caused some inefficiencies.

(denominator), is currently good (3 out of 4, Figure 2.14). However, because of its definition, a high efficiency score could partly be driven by budget overestimation (budget inception to date execution rate is artificially low) or target output underestimation (output score is artificially high). As Figure 2.15 shows, project teams have consistently underspent their budgets,³⁴ even before the pandemic, resulting in a low inception to date execution rate. While including a “cushion” in the budget for unforeseen expenditures is a good management practice, using this “cushioned” estimate for measuring output efficiency artificially alters the efficiency score.

Figure 2.14 The GTP’s output efficiency ratings are good (mostly 3 and 4) but could potentially be the result of budget overestimation or low output targets.

Distribution of output efficiency ratings across all GTP-supported projects

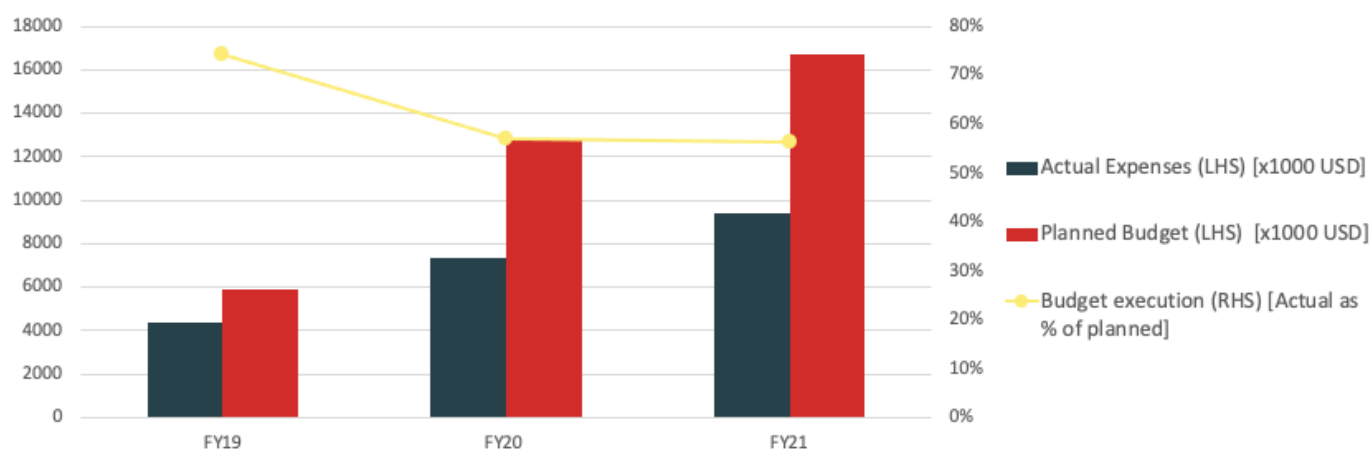


Source: GTP portfolio review assessment and SEO calculations. Data as of end FY21. Note: (Output) efficiency ratings are defined as the ratio between budget inception to date burn rate (numerator) and output score (denominator). The inception to date burn rate is measured as actual expenses compared to planned expenses (in %), whereas the output score is measured as achieved output compared to target output (in %). Efficiency ratings are 4 (0%-50%), 3 (50%-100%), 2 (100%-120%), 1 (>120%).

The GTP follows standard WB procedures for recruiting experts and setting daily rates, generally resulting in transparent processes and sufficient cost-effectiveness. However, further improvement is possible according to some stakeholders. The WB’s procurement system (econsultant2) is generally seen as professional, transparent and competitive. The stakeholder survey also showed that providers generally perceive rates of GTP consultants to be in line with (or even lower than) those of comparable institutions, confirming the procedure’s competitiveness. Nevertheless, some interviewees noted that exceptions to procurement procedures are often made in practice. One stakeholder noted that the GTP typically hires “familiar” consultants, following a special procedure (“direct hire”), when an argument can be made that the expert has highly specialized tax expertise. The evaluation team was not able to ascertain how often these exceptions are made, but also did not find evidence that the use of such special procedures is unjustified (given that tax expertise can indeed be highly specialized). Regarding diversity, the survey highlighted that a considerable share of providers believed that the GTP should more actively recruit female experts for TA projects.

³⁴ Another issue related to overbudgeting is that, as reported by one stakeholder, there is a tendency among TTLs to “pre-book” consultants without a clear prior plan on how to use them, to ensure their availability for a specific project. While this does not necessarily inflate budgets, it reduces efficiency in that these consultants can then sometimes be allocated less easily to other projects.

Figure 2.15 Budget execution was low (74 percent of planned in FY19, 57 percent in FY20, 56 percent in FY21). The drop in FY20 and FY21 was likely mainly due to COVID-19.



Source: GTP annual reports and SEO calculations.

2.5.2 Efficiency: project-level assessment

The evaluation team assessed all case study projects as moderately efficient and cost-effective.³⁵ Many of the projects were characterized by substantial underspending, which is consistent with the above-mentioned finding that (a) TTLs tend to prefer “cushioned” budgets (leading to systematic underspending and, in turn, to overestimation of the efficiency rate), and (b) it was not always possible to move all activities online (in a timely manner). Stakeholder interviews and a document review also highlighted that the pandemic led to underspending, mostly due to savings on travel costs (due to postponing or cancelling previously planned missions or replacing them with online activities).

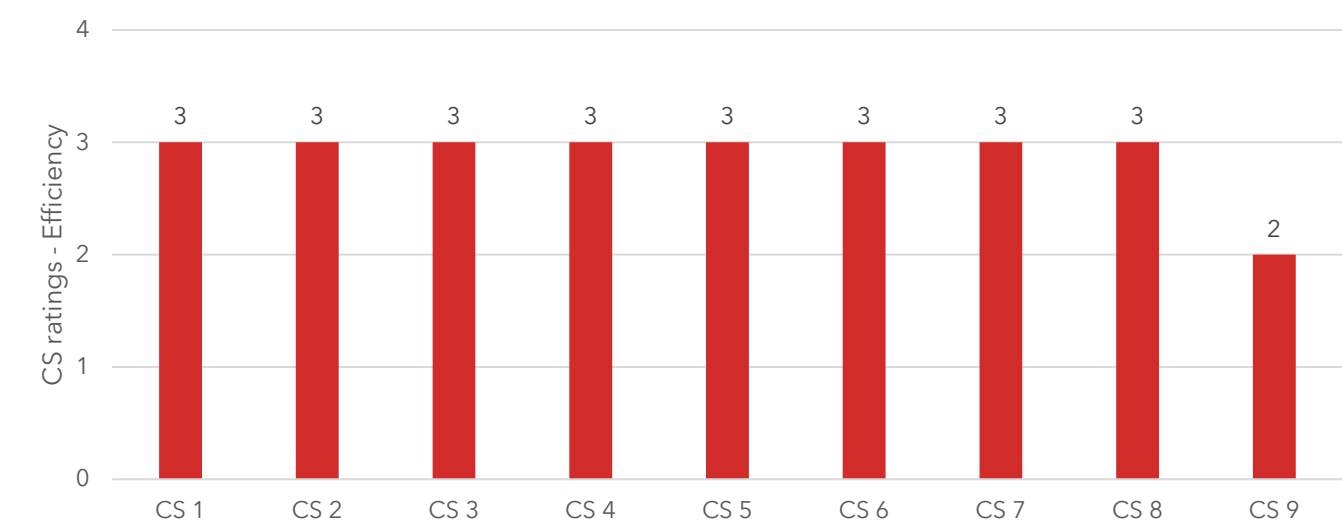
Most project teams delivered the results in a timely manner and gave realistic explanations for any delays.

The main reason for delays was the COVID-19 pandemic, although some projects also faced delays due to country-specific situations, for example political conflict or early elections. The presence of strong country teams was often seen as key to mitigating delays during the pandemic. For example, country teams helped to maintain regular contacts with government officials, including via regular discussions on project timelines, thereby lessening the negative effects of the pandemic on project timeliness.

³⁵

In a few cases, the evaluation team’s ratings were slightly lower than the GTP’s own output efficiency ratings. This was due to the different nature of the two ratings. The GTP rated *output* efficiency numerically, as explained above. Instead, the evaluation team’s efficiency rating also considered *outcome* efficiency and timeliness, and was based on multiple sources, such as document analysis, the survey and interviews, as well as the GTP’s own output efficiency ratings.

Figure 2.16 Most case study projects scored well on efficiency (3 out of 4)



Source: SEO Amsterdam Economics. 1 = not achieved; 2 = partially achieved; 3 = largely achieved; 4 = fully achieved.

3 Conclusions and recommendations

This chapter contains the main conclusions of the mid-term evaluation. In addition, we include a list of recommendations to further improve the GTP going forward.

3.1 Conclusions

RELEVANCE

Various sources confirm that GTP-supported projects are well aligned with the needs and priorities of recipient country authorities. The GTP has also responded well to emerging issues such as COVID-19, climate change and gender inequality.

- Stakeholders interviewed and surveyed widely agree that GTP-supported TA projects are demand-driven and relevant to the recipient country, despite a few concerns about absorption capacity.
- Case study projects scored well on relevance in all cases, since they were generally found to be (1) demand driven (i.e. responding to country requests), 2) addressing important gaps (e.g. revenue gaps or capacity gaps), and (3) in line with national priorities.
- One key advantage of the GTP is that it can leverage the World Bank's strong local presence through its network of country offices, which can facilitate the alignment of projects with the needs and priorities of TA recipients.
- The GTP has thus far responded well to emerging issues, by addressing topics such as climate change, gender inequality and health taxation. The WB's general strength in such thematic areas enables the GTP to be especially relevant in these fields.
- The GTP generally adapted well to the COVID-19 pandemic, including by replacing in-person activities with online activities where possible. However, travel budget savings could have been reallocated more quickly.

COHERENCE

The GTP actively monitors related activities by other WB teams and development partners. Internal and external coordination is taking place, and overlap and conflicting policy advice with other development partners is usually avoided. However, there is room to further increase complementarities and strengthen synergies, particularly with other development partners.

Internal Coherence

- GTP project teams have an eye for complementarities and synergies with other Bank operations in the tax domain and fiscal planning and budgeting domain, mostly other WB concessional lending operations or ASAs.
- Stakeholders generally saw GTP TA as coherent and synergetic with other WB operations (especially concessional lending operations), although some indicated that the GTP may be constrained somewhat due to the fact that not all WB teams see taxation as a key development tool.
- Case studies confirmed that internal coherence was generally largely or fully achieved, but that the linkages between TA and concessional lending operations could be further streamlined.

External Coherence

- The GTP actively monitors complementarities with other development partners, such as the IMF or the OECD.

- Multiple information sources (document review, stakeholder survey, case studies) confirmed that the GTP has generally coordinated at an acceptable level with other development partners.
- Overlap and conflicting policy advice is usually avoided, but there is room to increase complementarities and further strengthen synergies with other development partners. This applies in particular to the IMF and the OECD, including through the PCT.

EFFECTIVENESS

Despite COVID-19, most GTP-supported projects were effective in that they achieved their targets for outputs and short-term outcomes. However, longer-term outcomes were difficult to assess, since most projects were still ongoing and the GTP results framework does not yet include longer-term indicators.

- Both the GTP-wide portfolio analysis and the in-depth case studies confirmed that the targeted outputs and short-term outcomes were largely achieved, and that most projects were on track to meeting their objectives, despite some delays and constraints caused by COVID-19.
- The GTP results framework is well developed and appreciated by donor partners. The underlying logical framework shows an operationalized intervention logic, but the overall “Theory of Change” can be improved.
- Most indicators are formulated at the output level (e.g. “recommendations proposed”) or the short-term outcome level (e.g., “recommendations endorsed/adopted”). The indicators are well formulated, although the terms “endorsed” and “adopted” could be better defined.
- One shortcoming is that the results framework does not yet include longer-term outcome indicators related to the implementation of reform recommendations, since they are seen as “aspirational” and beyond the GTP’s control.

SUSTAINABILITY

The WB’s strong country presence and its ability to provide additional funding through concessional lending operations bode well for ensuring the sustainability of GTP results. However, risks to sustainability include high tax official rotations in country administrations, limited absorption capacity, political instability, and the short-term focus of the results framework.

- Stakeholders were generally positive about the likelihood that GTP results will be sustainable over time.
- The strong local presence through WB country offices increases the GTP’s potential to achieve sustainable results.
- In our case studies, sustainability was typically seen as only partially achieved. This is partly due to the fact that it was too early to judge, but also because risks to sustainability were identified.
- Risks to sustainability identified during case studies included the high turnover of local officials, the sometimes limited absorption capacity of recipients, and political instability.
- Risks to sustainability at the strategic level included the short-term focus of the results framework, and a possible loss of focus if the GTP would cover too many countries and areas.
- Factors that were seen as promoting the sustainability of results in case studies included high ownership and relevance, a change in the legal framework, and the combination of various types of support.

EFFICIENCY

The GTP generally manages its projects efficiently and in a timely manner. The evaluation team did not find evidence of any major issues regarding cost-effectiveness. However, operational efficiency could be further improved.

- The fact that the GTP itself monitors the “output efficiency” of its projects is a strong aspect of GTP project management, but the definition of output efficiency could be further improved.
- The evaluation team assessed all case study projects as moderately efficient and cost-effective.
- Stakeholders were generally satisfied with procurement processes and the organizational efficiency of the GTP, apart from some minor concerns about timeliness, exceptions to procurement procedures, and some issues related to the way in which the GTP fits inside the WB’s structure.
- Most project teams delivered the results in a timely manner and gave realistic explanations for any delays (e.g. due to COVID-19).

Table 3.1 Aggregate GTP ratings by OECD-DAC criteria

Criteria	GTP-wide assessment	Project-level assessment	Overall score
Relevance	4	4	4
Internal Coherence	3	3.5	3.25
External Coherence	3	2.5	2.75
Effectiveness	3	3	3
Sustainability	3	2.5	2.75
Efficiency	3	3	3

1 = Not achieved; 2 = Partially achieved; 3 = Largely achieved; 4 = Fully achieved. The overall score was determined by using equal weights for the assessments at both levels.

Table 3.2 Case Study ratings by OECD-DAC criteria

Criteria	CS 1	CS 2	CS 3	CS 4	CS 5	CS 6	CS 7	CS 8	CS 9	CS 10	Overall rating
Relevance	4	3	4	4	3	4	4	4	4		4
Internal Coherence	3	3	3	4	2	3	4	4	4		3.25
External Coherence	3	3	3	2	2	3	3	2	4		2.75
Effectiveness	3	3	3	3	3	3	3	3	3		3
Sustainability	2	3	2	3	2	2	4	3	2		2.75
Efficiency	3	3	3	3	3	3	3	3	2		3

1 = Not achieved; 2 = Partially achieved; 3 = Largely achieved; 4 = Fully achieved. The overall score was determined by using equal weights for the assessments at all levels and rounding the resulting figure to the nearest quarter point.

3.2 Recommendations and lessons learned

Although the GTP is already scoring relatively high on multiple OECD-DAC criteria, the evaluation team has the following recommendations to further improve the program based on the lessons learned from this mid-term evaluation.

RELEVANCE

- **Fine-tune the criteria for assessing relevance in project proposals.** For example, project teams could be asked to include evidence that (a) the revenue gaps and capacity gaps that the project aims to address are significant, (b) the project is in line with country priorities; (c) there is sufficient ownership and political will to implement recommendations in this area.
- **Consider more actively soliciting requests from countries where revenue gaps and potential DRM impact are highest:**
 - Increase the use of the WB Revenue Dashboard and other diagnostic evidence to assess revenue gaps;
 - Work with country teams to more actively assess the demand from countries with the highest revenue gaps and better assess ownership and the political will to adopt and capacity to implement recommendations.
- **Ask project teams to pay more attention in project proposals to assessing political economy risks and absorption capacity risks, while not turning the Bank's back on fragile states.**

COHERENCE

Internal coherence

- **Further improve the exchange of information with WB country teams** (since case studies revealed that they were not always fully informed about GTP activities).
- **Streamline and institutionalize requirements for assessing complementarities with concessional lending and/or ASA operations.** For example:
 - Ask project teams to include more information on (existing or possible future) complementary concessional lending operations, including on the type of complementarities, both in project proposals and during annual monitoring.
 - Identify complementary sources of funding for constraints that cannot be addressed via the GTP (e.g., software or hardware).
- **Continue to support efforts to include taxation as a key part of the Bank's development solution toolbox.** This could potentially be done by increasing some internal and external outreach activities, such as:
 - Providing regular briefings to senior management on progress on the tax agenda;
 - Devoting more resources to creating an intellectual center of excellence, reflected by additional drafted and disseminated research papers, policy papers and other reports;
 - Seeking senior management support for a World Development Report or other flagship publications on taxation and development.
- **Deepen activities in areas where there may be clear synergies with other Bank work:**
 - This may allow the GTP to specialize more in specific thematic tax areas, based on the WB's expertise in fields such as climate, gender, infrastructure, subnational governance, health, urban, extractives and poverty;
 - This could also improve external coherence (more complementarities with other DPs).

External coherence

- **Further strengthen synergies with other development partners (DPs) offering TA in the DRM domain (e.g., IMF and OECD).** This could be achieved, for example, by:
 - Increasing the exchange of information, e.g., through more frequent sharing of (draft) project proposals, TA reports and other project documents;
 - Inviting relevant other DPs to review project proposals (and offer to review theirs) and conduct joint meetings with local authorities where possible;
 - Co-designing joint TA projects from the start;
 - Making more active use of existing donor coordination working groups (jointly with country teams and local authorities).
- **In coordination with the PCT, explore increasing complementarities between the GTP and other DPs operating in the DRM domain, while taking into account that different DPs have different mandates and delivery models.** The existence of the PCT has already led to greater clarity between the four IOs concerning their tax work. Going forward, it may be useful to:
 - Agree on a set of mutually recognized areas of comparative expertise and comparative advantages (value added), taking into account the respective organizational strengths, accountabilities and costs;
 - Based on these agreed comparative advantages, aim for a clearer allocation in focus areas, while still ensuring that each organization pays regard to its mandates and stakeholders;
 - In pursuing this, an obvious area for the GTP to focus on would be to take the lead on some specific thematic tax areas, thereby leveraging the WB's wider expertise in broader development areas such as climate, gender, health and poverty.

EFFECTIVENESS

- **Simplify the Theory of Change** and distinguish clear impact pathways and underlying assumptions.
- **More clearly define the concepts of "endorsed" and "adopted".** This could be done, for example, by creating a formal list of feasible actions that would qualify as endorsement or adoption).
- **Expand the results framework by including indicators of longer-term outcomes.** In particular, it is important to monitor not only if a reform recommendation is endorsed and adopted (short-term outcomes), but also to what extent steps are taken to implement these reforms (medium-term outcomes), and to what extent revenue gaps are eventually reduced (long-term outcomes).
 - Review the already developed "aspirational" indicators and consider formally including them in the results framework.³⁶
 - Develop additional medium-term and long-term indicators as needed.³⁷
 - Involve WB country teams in the monitoring of these longer-term outcome indicators. Such monitoring should ideally be continued beyond the project's lifetime (particularly when implementation processes can take a long time) and could make a useful contribution to the development of new projects.
 - Explore whether the WB could adopt a similar strategy on targeting and monitoring long-term outcomes at the institutional level.
- **Further improve the results framework.** While the results framework is already well designed, there are still various opportunities to further increase its usefulness. For example:
 - Put less emphasis on aggregate numbers (e.g., total number of recommendations), because the definition of "one recommendation" is somewhat arbitrary and the "size" of a recommendation can differ greatly by

³⁶ For example, "number of countries that raise the tax-to-GDP ratios above 15 percent".

³⁷ For example, "number of recommendations implemented" in addition to "number of recommendations adopted".

project (more guidance on when to split one large recommendation into two parts can also be helpful here).

- Rather than focusing on the aggregate number of recommendations endorsed or adopted, report the share of recommendations endorsed/adopted/implemented as a percentage of the recommendations proposed.
- Consider expanding the results framework with “contribution scores” that assess to what extent the GTP contributed to a reform being endorsed, adopted, or implemented. This could be complemented with a qualitative description of which other external factors contributed positively or negatively to the adoption or implementation of a reform. Such a “contribution analysis” can also help to improve future project design.

SUSTAINABILITY

- Increase the likelihood that project results will be sustainable by systematically assessing the need for follow-up support with implementation, either from the GTP, other WB funds, or from other development partners. For instance:
 - Towards the end of a GTP-supported project (or even on an annual basis), systematically assess the need for follow-up support, including by asking project teams to identify any remaining knowledge, financial or technical capacity gaps, in coordination with the recipient authorities.
 - Following the adoption of reforms, assess the need for follow-up support (from the GTP or other sources) with reform implementation (e.g., developing relevant by-laws in case of new tax legislation).
 - Increase the role of World Bank country teams in ensuring follow-up (e.g., by allocating a certain share of the budget to follow-up activities).
- **Enhance efforts to achieve longer-term outcomes by leveraging the strength and continuity of the Bank’s strong country presence.** In particular, GTP project teams could be asked to allocate a certain share of the budget to follow-up activities by WB country teams with:
 - monitoring of reform implementation efforts (even if there is no new project/operation);
 - follow-up implementation support / formulation of new GTP projects and WB operations.
- Improve sustainability of GTP results by further strengthening synergies with other WB initiatives (see recommendations under “internal coherence”) and with other development partners (see recommendations under “external coherence”).

EFFICIENCY

- **Improve the measurement of output efficiency by avoiding overstated budget estimates.** In particular, project teams could be asked to explicitly budget for “unforeseen expenses”, so that these can be excluded (if not materialized) when calculating output efficiency.
- **Improve the allocation of consultants across GTP-supported projects.** In order to avoid long tender procedures or prevent that the most valued tax consultants are “pre-booked” for particular tax projects (which is not efficient), the GTP Secretariat could consider the following measures:
 - Ensuring that, at the time of “pre-booking” consultants, the TTL makes clear arrangements on their ToR (providing more security to consultants) and informs other TTLs about them;
 - Allow for budgets and consultants to be reallocated more quickly in case of budget savings or delays (e.g., due to COVID-19 or political instability);

- Consider hiring short-term tax experts through a “GTP framework agreement” for which consultants can pre-qualify (through a competitive tender), after which hiring procedures for specific projects can be much quicker (provided that this is possible within the World Bank’s overall procurement procedures).
- **Further increase diversity within GTP project teams**, by scaling up efforts to include more women, youth, and local consultants (including via targeted recruitment and advertisement of staff or short-term consulting positions).

Appendix A Evaluation questions

Table A.1 Evaluation questions

Category	No.	Question
Relevance	1.1	Were the objectives derived from capacity gaps identified by the authorities, the WB or other partners/institutions?
	1.2	Do the national authorities consider the project's objectives among the priorities of the country?
	1.3	Were the design and implementation sensitive to the context (e.g., economic, political economy, technical capacity of the beneficiaries, qualifications of experts) in which they took place?
	1.4	Were the objectives and risks detailed in the proposals realistic at entry?
	1.5	Were the objectives and design successfully adapted to changing circumstances (e.g., COVID-19)?
	6.1	Was the quality of the proposals effectively vetted prior to approval (in terms of the WB's comparative advantages, results framework and client needs)? (Relevance of program management)
	7.1	Is the current GTP portfolio well aligned with the strategic priorities set for the Program? Are there any gaps? (Relevance of program strategy)
	7.4	What changes, if any, would be necessary to better respond to developing countries' needs post-COVID-19? Can GRID and IDA20 be better reflected in the Program strategy? COVID (Relevance of program strategy)
Coherence	2.1	Internal: To what extent are there synergies and interlinkages between the project and other interventions carried out by the government?
	2.2	Internal: To what extent is the project consistent with other WB initiatives?
	2.3	Internal: To what extent were any complementarities with ongoing WB projects properly identified during the design phase of the projects?
	2.4	Internal: How well-coordinated were the GTP funded activities with other Bank and DP efforts?
	2.5	Internal: To what extent are tax frameworks and diagnostic tools (from the WB and other development partners) used to define and inform tax reform priorities and project design in the GTP-funded projects?
	2.6	External: To what extent is the project consistent with interventions carried out by development partners?
	2.7	External: To what extent were any complementarities with ongoing interventions by development partners properly identified during the design phase of the GTP funded activities?
	2.8	External: How well-coordinated were the GTP funded activities with other DP efforts?
Effectiveness ³⁸	3.1	Were the inception-to-date results, as defined in the results framework, achieved or are they likely to be achieved?
	3.2	To what extent can the observed direct results be attributed to the project? (CS)

³⁸ The ToR also notes that: "Evaluators are encouraged to the best extent possible to isolate/differentiate the COVID-19 impact on the Program from other factors (internal to the intervention or external) and identify areas for improvement to guide the intervention during COVID-19 and future scenarios with normal conditions."

	3.3	To what extent did the COVID-19 pandemic adversely impact program results and delivery (versus other factors)?
	6.2	Are the progress reports useful to stakeholders for reporting on the progress and delivery of the activities? (Effectiveness of program management)
	6.3	Were activities at risk of not meeting their identified objectives and were issues addressed in a timely manner? (Effectiveness of program management)
	6.5	How well have the GTP's results been disseminated and reported to donor partners as well as at international fora where DRM is discussed? (Effectiveness of program management)
	6.6	How effective has the dissemination of knowledge and information generated through the GTP been? (Effectiveness of program management)
	7.2	How well has the Program adapted and/or flexibly responded to emerging issues and changing external circumstances, e.g., addressing issues such as climate change, gender equality, and health taxation (sin taxes)? (Effectiveness of program strategy)
	7.3	How well has the GTP facilitated and/or encouraged collaboration across teams and with other development partners to develop a comprehensive Program where activities can be coordinated and sequenced strategically? (Effectiveness of program strategy)
	7.5	What are the options for strengthening the synergies between GTP activities and other Bank operations, notably tax and other fiscal projects, debt operations, governance / digital projects, and sectoral operations? (Effectiveness of program strategy)
	4.1	To what extent does the intervention deliver, or is likely to deliver, results in an economic and timely way?
	4.2	Were the project inputs converted to outputs, outcomes, and impact in the most cost-effective way possible, as compared to feasible alternatives? <i>(if possible, comparing both TA delivery modalities (physical missions/phase to phase vs remote TA))</i>
Efficiency	6.4	Are the staffing, internal mechanisms and procedures sufficient for managing the Program? (Effectiveness of program management)
Sustainability	5.1	Discuss the likelihood of the net benefits continuing over the medium and long term.
	5.2	Assess the capacity of the recipient country's system to sustain the benefits of the TA received over time.

Appendix B Case study selection

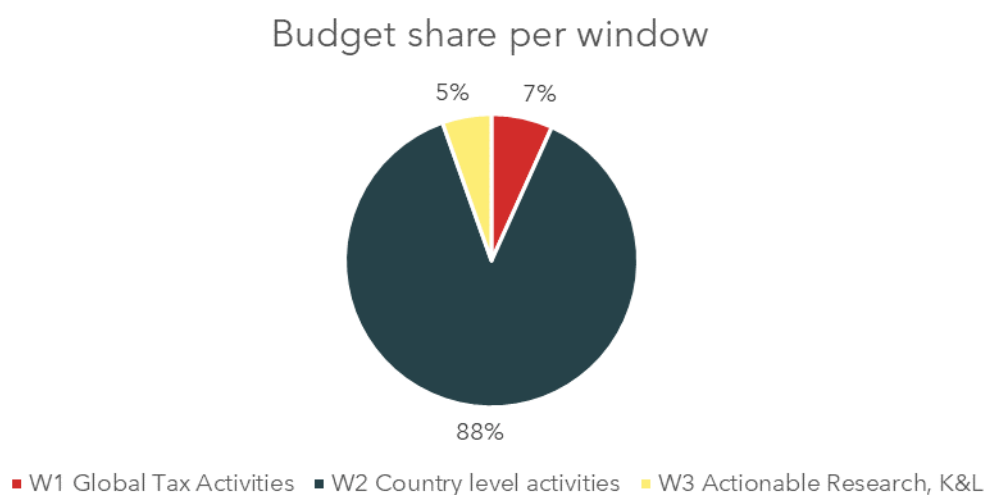
The evaluation team made a considerable effort to select a representative sample for its case studies.

Conclusions derived from case studies can be extrapolated to the program level only if the case study sample can be considered to be reasonably representative in certain important respects. To ensure a representative sample, the evaluation team therefore applied the following methodology during the inception phase of the evaluation:

- **Step 1:** Analysis of the distribution of key characteristics within the GTP portfolio for the period under evaluation.
- **Step 2:** Determination of selection criteria based on portfolio characteristics.
- **Step 3:** Application of selection criteria and selection of case study samples (in coordination with the Evaluation Committee).

As of June 2021 (end of evaluation period), the GTP portfolio included 45 projects in Windows 1-3. In terms of budget distribution, 88 percent of the total budget was allocated to country-level activities (Window 2). Windows 1 (global tax activities) and 3 (actionable research) covered 5 and 7 percent of the budget, respectively.

Figure B.1 88 percent of the budget was allocated to country-level activities (Window 2).

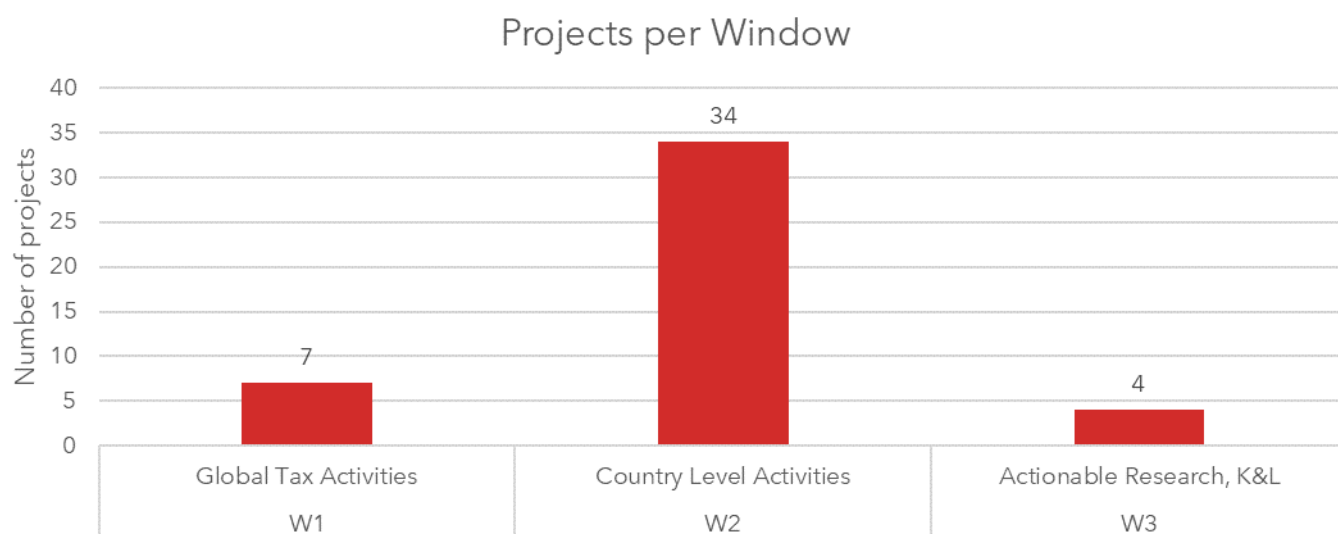


Source: SEO Amsterdam Economics, based on GTP Portfolio data and GTP Annual Report FY20.³⁹

Note: Estimation excludes Window 4 budget (program management)

³⁹ Note that these figures do not include the temporary Fiscal Pillar that was added to the program during FY20.

Figure B.2 As of June 2021, most projects were part of Window 2.



Source: SEO Amsterdam Economics, based on GTP Portfolio data.

A brief description of the four windows:

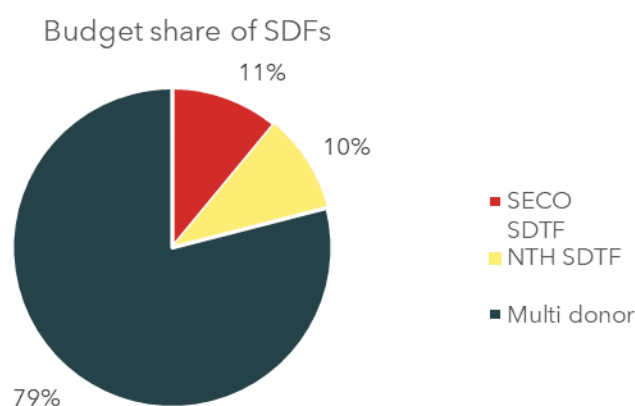
- **Window 1: Global Tax Activities and Global Public Goods.** The work on Global Public Goods includes the development of diagnostic tools, engagements and the provision of support for improving DRM in developing countries. This includes the Tax Policy Assessment Framework (TPAF) and the Tax DIAMOND framework⁴⁰ for identifying drivers and solutions of performance weaknesses in revenue administrations.
- **Window 2: Country Level Activities.** This window supports countries in the improvement of revenue collection through medium-term revenue strategies (such as the MTRS) that are guided by country-focused diagnostic work such as TADAT or DIAMOND, MTRS, flagship research work on Innovations in Tax Compliance, PCT toolkits on international tax, DRM data analytics and policy tools.
- **Window 3: Actionable Research and Data, Knowledge and Learning.** This window brings together various strands of research, data analytics and experimental methods to promote shared learning and push the research frontiers. This window includes research reports with the dual objectives of: (i) pushing the research frontier in an area of DRM that leverages the GTP's comparative advantages; and (ii) creating actionable research that will ultimately improve operational results and inform the design of future WB projects. Knowledge management, sharing and training were also identified as priorities under Window 3.
- **Window 4: Program Implementation and Supervision.** This window supports the management of the Multi-Donor Trust Fund (MDTF) and Single Donor Trust Funds (SDTFs) through the GTP Secretariat.⁴¹ The MDTF manages funds collected from different donor partners as a single program, while the activities within Single-Donor Trust Funds (SDTFs) are determined by a bilateral agreement between the donor partners and WB. The two SDTFs are supported by Switzerland/SECO and the Netherlands, respectively.

⁴⁰ <https://www.taxdiamond.org/>

⁴¹ According to the Annual Report FY20, the GTP Program is "one of the Umbrella 2.0 pilot programs for Trust Fund (TF) Reforms recently undertaken by the WB. One of the main objectives of this TF Reform is to improve the efficiency and effectiveness of TFs at the WB. A more centralized approach is expected to help better match the priorities of countries and donor partners with WB comparative advantages and simultaneously improve the management of TFs (from both donor partners' and the WB's perspective)."

The GTP budget consisted of both the multi-donor trust funds (MDTF) and single-donor trust funds (SDTFs). Nearly 80 percent of the total budget corresponds to the MDTF. The SDTF SECO and SDTF NTH correspond to a minority of 11 and 10 percent, respectively (see Figure B.3).

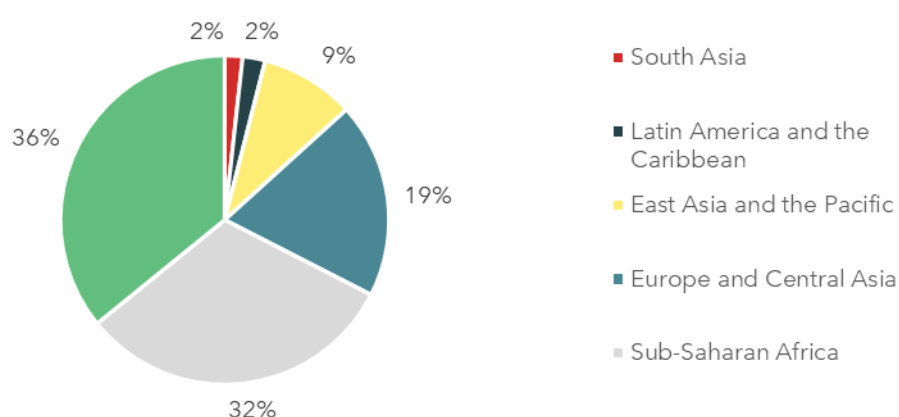
Figure B.3 Three quarters of GTP activities were funded by multi-donor trust funds.⁴²



Source: SEO Amsterdam Economics, based on GTP Portfolio data. Note: Estimation excludes Window 4 budget (program management)

In geographical terms, the largest part of the budget was allocated to Sub-Saharan Africa (SSA) and Europe and Central Asia (ECA). Of the total budget, 32 percent was directly linked to individual countries in SSA; if we exclude projects which were identified as “multi-country”, the corresponding weight of AFR is 54 percent.

Figure B.4 The largest part of the country-specific budget was allocated to Africa and the Middle East (AFR).⁴³



Source: SEO Amsterdam Economics, based on GTP Portfolio data.

Note: Estimation excludes Window 4 budget (program management)

The following selection criteria were used to arrive at a representative case study selection:

- Representative spread across regions.

⁴² Estimation excludes Window 4 budget (program management)

⁴³ Estimation excludes Window 4 budget (program management)

- Representative spread across SDTF and MDTF distribution.
- Representative spread across windows.
- At least 2 smaller GTP projects (< US\$ 750,000) and at least 2 larger GTP projects (>US\$ 3 million).
- Project start date in the period of June 2018 through December 2020.⁴⁴

Table B.1 Based on the above criteria, we selected the following sample of ten case studies

Name	Window	Region	Start date	Expected End Date	Budget US\$ x 1000	Funding source
1.6 Carbon Pricing Assessment Tool	1	Multi-country	Jul/20	Jun/23	500	MDTF
2.04 International Tax Program	2	Multi-country	Jul/17	Jun/23	6800	MDTF
2.32 Senegal MTRS Tax and Customs	2	AFR	Jan/20	Dec/22	966	NTH
2.34 Ethiopia Tax Policy	2	AFR	Jun/20	Jun/24	1960	NTH
2.26 Vietnam Tax Reform	2	ASIA	Dec/18	Dec/22	1075	SECO
2.27 State Tax Service Kyrgyz Republic	2	ECA	Dec/18	Mar/24	2000	SECO
2.20 Malaysia Tax Policy Reform	2	ASIA	Oct/20	Mar/24	500	MDTF
3.4 Ethiopia Gender and Taxation	3	AFR	Jun/20	Mar/24	1290	MDTF
2.10 Somalia Customs	2	AFR	Nov/19	Jun/23	2000	MDTF
2.05 Tax Reform in Uzbekistan	2	ECA	Nov/18	Mar/24	3200	MDTF

Table B.2 This proposed case study sample met all selection criteria

Selection criterion		Portfolio*	Case study sample	Score
Representative spread across regions	AFR	42%	31%	✓
	ECA	25%	26%	
	ASIA	6%	8%	
	LAC	1%	0%	
	Multi-country	26%	36%	
Representative spread across windows	W1	6%	2%	✓✓
	W2	87%	86%	
	W3	7%	9%	
Representative spread across SDTF and MDTF	MDTF	72%	72%	✓✓
	SECO	14%	14%	
	NTH	14%	14%	
At least 2 smaller GTP projects (< US\$ 750,000)				✓✓
At least 2 larger GTP projects (>US\$ 3,000,000)				✓✓

* Percentages exclude GTP projects that fell outside the evaluation period.

⁴⁴ 10 TA projects fell outside the time frame and were therefore excluded.

Appendix C GTP Theory of Change





"Solid research, Sound advice."

SEO Amsterdam Economics carries out independent applied economic research on behalf of national and international clients – both public institutions and private sector clients. Our research aims to make a major contribution to the decision-making processes of our clients. Originally founded by, and still affiliated with, the University of Amsterdam, SEO Amsterdam Economics is now an independent research group but retains a strong academic component. Operating on a nonprofit basis, SEO continually invests in the intellectual capital of its staff by granting them time to pursue continuing education, publish in academic journals, and participate in academic networks and conferences. As a result, our staff is fully up to date on the latest economic theories and econometric techniques.

SEO report No. 2022-150
ISBN 978-90-5220-251-8

Information & Disclaimer

SEO Amsterdam Economics has not performed any research on the obtained information and data that would constitute an audit or due diligence. SEO is not responsible for errors or omissions in the obtained information and data.

Copyright © 2022 SEO Amsterdam.

All rights reserved. Data from this report may be used in articles, studies and syllabi, provided that the source is clearly and accurately mentioned. Data in this report may not be used for commercial purposes without prior permission of the author(s). Permission can be obtained by contacting: secretariaat@seo.nl.

Roetersstraat 29
1018 WB, Amsterdam
The Netherlands

+31 20 525 1630
secretariaat@seo.nl
www.seo.nl/en/