**Executive summary**

**Introduction**

The chief purpose of this meta-evaluation is to assess, in a systematic manner, the quality of evaluations commissioned by the Czech Ministry of Foreign Affairs (MFA) in 2012 and 2013 and obtain an objective picture of the state of the Czech system of development-project evaluations, in an effort to improve the system and ensure that it follows international evaluation standards. The evaluations under scrutiny are 20: nine were conducted in 2012 and 11 in 2013. Four evaluations are of projects implemented in the Czech Republic, three of projects in Georgia, two each of projects in Mongolia, Moldova, Ethiopia, and Kosovo, and one each of projects in Palestine, Zambia, Cambodia, and Burma/Myanmar.

The contracting authority of the meta-evaluation is the MFA; the implementer is an evaluation team of Ivica Petríkova, MA (principal evaluator) and Dhruv Chadha, MSc. (evaluation expert).

**Methodology**

*Brief analytical framework*

In judging the quality of the primary evaluations, we follow an adjusted framework developed by the Czech Evaluation Group (2013) on the basis of international quality criteria of evaluations – utility, feasibility, propriety, and accuracy. To these four principles we added a fifth one - evaluation system quality-which aims to assess the quality of the Czech evaluation system as a whole**.**

Utility standards aim to ensure that evaluations satisfy the information needs of their intended users in a timely and credible manner. Propriety standards have to do with the evaluations’ ethical aspects. Accuracy standards seek to guarantee the factual accuracy of evaluations. Finally, within the evaluation system qualitycategory, we examine whether the process of selecting evaluators is fair and efficient, whether the terms of the evaluations and the support provided by government officials to the evaluators are adequate, and whether the evaluations’ recommendations shape further policy in any significant manner.

*Data sources and analysis*

In this meta-evaluation, we work with both secondary and primary data. The main source of the primary data are the main texts and appendices of the twenty primary evaluations listed in the previous section along with their Terms of Reference. We also consulted a range of relevant secondary documents. Other primary data were collected through interviews, both group and individual, with evaluators, implementers, and Czech government officials, and through an on-line survey. The data collected were analyzed using primarily qualitative methods (desk research, content analysis), with some simple quantitative methods (percentages and proportions) utilised in the analysis of the survey data.

*Methodological limitations*

One methodological limitation has been the timing of the meta-evaluation during the summer months of July and August, when many relevant stakeholders were on vacation. Another limitation relates to the suspected low validity of data obtained via the online survey. As an attempt at remedy, we did assess the survey’s results but crosschecked them with the findings attained from other data sources, to which we gave preference in the case of disagreement.

Language constituted another problem for us – neither of us can write in Czech and hence we had to rely on hired translators without the ability to ensure that the translation was carried out correctly. Finally, despite our best efforts at objectivity, as people we are by nature subjective. Consequently, the views expressed in this evaluation are those of the evaluators only and do not necessarily represent those of the Czech Ministry of Foreign Affairs or of any other stakeholder in the Czech evaluation system.

**Most salient findings**

*Utility*

In examining the evaluations’ utility, as the main positive characteristic we found the good level of communication between evaluators and other stakeholders and the fact that all the evaluations fulfilled their Terms of Reference. All evaluations also applied the OECD/DAC criteria but some had problems with the understanding of ‘impact’, which some evaluators apparently interpreted as ‘effectiveness’, and of ‘sustainability’, which some evaluators interpreted as sustainability of outputs while others as sustainability of impacts. A more serious misunderstanding has been noted in the application of the cross-cutting issues. Regarding the evaluation of the projects’ implementation of human rights and gender equality, the situation should improve following the 2014 publication of relevant guidelines by FfDC. The application of the environment and climate protection and good governance criteria would unify and likely improve as well if similar guidelines were published also on these issues. Further, the quality of some evaluators’ appraisal of projects’ effectiveness and impact was reduced when they judged the projects’ outputs, goals, and purpose against a logical framework that they themselves re-constructed. Lastly, several evaluators did not assess the degree of the projects’ coordination with other projects in the area, which is key in evaluating the projects’ relevance and which has been increasingly emphasized by international and European guidelines as crucial in the efforts to bolster aid effectiveness.

Other drawbacks of the primary evaluations from the utility perspective have been a recurrent lack of involvement of all relevant stakeholders, low quality of some recommendations, and a relatively low accessibility of results. Several evaluators left out local recipients from stakeholder analyses and even if they included them, they frequently failed to solicit their views on what the evaluations should focus and on how the projects affected them. The implementing organizations, both Czech and local, have also expressed a desire to play a greater role in the evaluation process, which would likely enhance the quality and ‘value added’ of the evaluations’ recommendations. In the case of local partners, the desire to be involved in evaluations is further complicated by the fact that to date most evaluation reports have been published only in Czech, with English executive summaries. Even among the Czech stakeholders the evaluation reports have not been widely publicized. They are available online on the MFA’s website but due to the structure of the website their visibility is rather low.

*Sensibility*

On feasibility, the evaluations have fared better than on utility. They generally fulfilled the timelines set by the MFA even though those were often quite short and sometimes seasonally wrongly set up. For example, evaluations of Southeast-Asian projects took place during the monsoon season, when it is difficult to travel, and the evaluations of a project aimed at schoolchildren took place during the summer holidays, when evaluators could not interview or survey the children targeted. Upon evaluators’ recommendations the MFA, however, took these complaints into account and already modified the timelines of evaluations so as to avoid similar mistakes.

The evaluations were also quite cost effective. Even though the range of per-diems and daily rates charged by the different evaluators was very wide, even the most expensive evaluations did not exceed 4 per cent of project costs or 30,000 USD, which are both below international averages. Finally, even though, as mentioned under utility, many evaluators failed to take into account the interests of local project recipients, they were mostly aware of the different interests that stakeholders might have had and did their best to reach politically balanced conclusions.

*Propriety*

From a propriety perspective, the evaluations have also been overall satisfactory. Through our research, we did not discover any grave ethical concerns. In general, evaluators respected human rights and human dignity at all stages of the evaluations process and did their best to Nefunkční přístup do inet pcminimize the potential negative consequences of evaluations, by offering constructive solutions to things that did not work well in projects rather than recommending the cessation of funding. No conflicts of interests were uncovered amongst the evaluators either, which owes its due also to the MFA’s requirement that evaluators are in no way (familiar or business) related to the implementer or commissioner. However, we did discover several discrepancies between some evaluations’ findings on the OECD criteria and corss-cutting issues and their corresponding ratings. In particular, in several cases the ratings seemed inflated in light of the findings’ discussion. Other weaknesses in the propriety dimension have included the frequent lack of acknowledgement of methodological limitations on evaluations’ findings and in one case, a comparison of another project to the Czech one as less effective and efficient without convincing evidence.

*Accuracy*

On the accuracy front, the evaluations were again a bit more deficient. As a group, they did have several strengths even here, particularly in terms of making an effort to examine all relevant project documentation - an effort, in which, however, they were hindered by the fact that sometimes the necessary documentation was missing. Another positive characteristic has been the relative clarity with which many evaluations managed to describe the context of the projects evaluated and sometimes also the projects themselves. The same cannot be said about the evaluations’ methodology where the descriptions were often brief, unclear, and sometimes describing data sources and/or methods that were never actually used.

The primary deficiency within the evaluations’ accuracy rests with their methods of data collection and data analysis. Only a few evaluators made an effort to collect data randomly so as to obtain a representative sample of the population of project recipients. It is true that particularly for qualitative data analysis non-random methods of data sampling, such as maximum variation, extreme, snowball, and criterion sampling, are acceptable and even encouraged but on the basis of such data one cannot draw inferences about the population (of project recipients), which most evaluators did. Another problem with the data collection has been the frequent lack of baseline data (however, that is a problem of the implementers not evaluators) and infrequent collection of data on a comparison/counterfactual group, without which establishing impact is very difficult if not impossible.

That brings us to the problems observed with the analytical approaches used. Most of the primary evaluations did not use impact-evaluation methods and hence their findings on impacts cannot be considered valid. If one views the definition of ‘sustainability’ as the long-term continuance of impacts (not outputs), the conclusions about sustainability cannot be consequently deemed valid either. The fact that financial monitoring reports were often not available to the evaluators could have compromised their findings on the projects’ efficiency as well.

*Evaluation system quality*

Some deficiencies encountered within the primary evaluations are closely linked with the deficiencies within the Czech evaluation system. To mention its positives first, however, it is admirable that the Czech Republic, a small DAC donor in terms of the amount of aid provided, has an evaluation system at all. None of the other new EU donors, including Poland that provides more total aid, has established anything as sophisticated or institutionalized. Furthermore, the Czech evaluators are external and independent, which affords the system credibility; the MFA, CzDA, and implementing agencies are all open about their support for the process; and there is even an institutionalized process for responding to and incorporating the evaluations’ recommendations. There are still some deficiencies within the system as it stands, but the desire to improve further is clear as is evidenced also by the commissioning of this meta-evaluation.

To turn to the drawbacks of the Czech evaluation system, which are at least partially responsible for the lower quality of some of the evaluations, they include primarily short time and low funding available for the evaluations, the wrong timing of some of the evaluations, the inability to attract more qualified candidates, and the evaluations’ often too generic Terms of Reference.

**Main recommendations**

|  |  |  |
| --- | --- | --- |
| Aspect/  importance | Recommendation | Addressee |
| *Utility* | * If possible, assess the effectiveness of projects by evaluating them according to their original, not the reconstructed, logical frameworks. * Assess how well the project evaluated worked with and complemented/contradicted other similar projects in the area * Include local ‘common-people’ project recipients – i.e. people affected by the project who neither participated in it directly nor made decisions about it - in your stakeholder analysis, in your interviews and focus groups, and if possible, also in identifying additional evaluation questions * If possible, share – or ask the MFA to do so – the final evaluation reports with local partners. For local project recipients consider organizing a de-briefing or some other engaging way of presenting preliminary results before you leave the field | Evaluators (and in the last case, also MFA) |
| *Feasibility and propriety* | * Make sure that the ratings on the OECD/DAC criteria and the application of the cross-cutting issues tie in with the findings discussed. * Do not compare the ‘success’ of the project evaluated with other projects’ unless you have good evidence backing up such a claim. | Evaluators |
| *Accuracy* | * Try to describe the projects evaluated in such a way that even those previously not familiar with them can grasp their main concepts and activities. * Describe in detail what methods were used in the evaluation and do not discuss methods that were not used without * Consider using random-sampling methods of data collection or some other method aimed at increasing the samples’ representativeness. * Consider collecting data on a counterfactual/comparison group. * Consider using analytical methods suitable for impact analysis. | Evaluators |
| 1 | * Consider changing the official language of non-Czech evaluation projects to English, with executive summaries in Czech and in the most prevalent local language. | MFA |
| 2 | * Consider publishing the list of planned evaluations at least one year in advance and pre-paying at least some of the evaluations’ costs. | MFA |
| 2 | * Consider writing more detailed Terms of Reference. | MFA |
| 1 | * Consider well the timing of the evaluations (enough time, suitable time of year, suitable time in the project cycle for an evaluation). | MFA |
| 2 | * Consider clarifying suitable application of the OECD/DAC criteria and of the cross-cutting issues in projects as well as in evaluations. | MFA |
| 3 | * Consider requiring a common graphic evaluation template; consider increasing the maximum number of pages allowed in the main text of evaluations; do not limit the number of recommendations allowed. | MFA |
| Nefunkční přístup do inet pc | Nefunkční přístup do inet pc | Nefunkční přístup do inet pc |
| 2 | * In order to further increase visibility of the evaluation system and its results, consider encouraging a more open discussion about the evaluations’ findings, recommendations and their intended or executed incorporation (including publishing related documents on www.mzv.cz). | MFA, CzDA |
| 2 | * In relation to the above, consider carrying out, by 2016, an analysis focused specifically on the incorporation of evaluations’ recommendations, with the aim to identify the main obstacles that prevented partially or fully the implementation of some of the recommendations. | MFA |
| 2 | * Consider allowing evaluators to switch costs between partial budget items by at least 15-20% and ensuring that upon winning the bid, they receive a CD/USB with the final versions of all relevant documentation along with a checklist. | MFA, CzDA |
| 3 | * For the future, consider creating an evaluation department within the MFA, which would be staffed by people not subject to periodic diplomatic rotations. | MFA |

Recommendations for the MFA and CzDA are labelled 1, 2 or 3 depending on the degree of their perceived importance (1 being the most important). Recommendations for the evaluators are not labelled in such manner because the evaluations differ significantly and hence recommendations that mught be very relevant for some might not be equally relevant for others.