

## Executive summary

### Introduction

The Policy and Operations Evaluation Department (Internationaal Onderzoek en Beleidsevaluatie (IOB)) of the Ministry of Foreign Affairs of the Netherlands commissioned a synthesis study on Dutch support for education in Official Development Assistance (ODA) countries. The study includes evaluations of Dutch-funded initiatives in education, published between 1 January 2015 and 31 of December 2023. IOB assessed these evaluation reports on the basis of thirteen IOB Quality Criteria for Evaluations. This resulted in nine reports that were deemed to have sufficient methodological quality to be included in the synthesis. The findings from these nine evaluation reports have been assessed in this report and their results have been synthesised.

The overarching questions for the synthesis are:

1. What do existing evaluations of Dutch-funded initiatives for education in ODA countries say about the extent to which the initiatives were effective in achieving their goals and why?
2. What do existing evaluations of Dutch-funded initiatives for education in ODA countries say about the extent to which the initiatives were relevant, coherent, efficient, and sustainable, and why this was the case?
3. Based on this, what overarching lessons can be drawn?

Responding to nine sub-questions, the study has described the interventions and their intervention logic and based on the evidence in the nine selected evaluation reports, assessed the interventions against the DAC criteria effectiveness, impact, efficiency, relevance, coherence and sustainability. The sub-questions also helped assess to what extent the interventions had been gender sensitive, and what overarching findings could be derived.

The synthesis study faced some limitations. As it was based on a desk review, it drew solely on secondary data, limiting the triangulation of findings through primary research. The review identified several limitations in the quality of the evaluation reports, including extensive reliance on stakeholder perceptions, lack of input from respondents not directly involved in the interventions, absence of contribution or attribution assessments, potential document preselection, insufficient time for comprehensive evaluation, and non-random respondent selection. These factors may have introduced bias, affecting the reliability and validity of the synthesis study's results. For some of the nine evaluation reports, findings on the intervention's effectiveness were excluded from the synthesis, as the methodology was not assessed as sufficiently robust to support reliable claims about effectiveness. For all of the evaluation reports, impact assessments were either absent or not methodologically reliable, and most findings at impact level were therefore excluded from the synthesis report. The interventions also exhibited significant differences in scope, size, and design, complicating aggregation. Lastly, as the evaluation reports span 2015 to 2023, some information may be outdated, and relevant findings published after 2023 were omitted, meaning adjustments in response to earlier evaluations are not reflected in the synthesis.

### Findings

#### The assessed evaluation reports and programmes

Five assessed evaluation reports are for multilateral programmes. Two evaluations concern the Global Partnership for Education (GPE), which is the largest educational global fund supporting reforms in primary and secondary education across approximately 90 countries. Established in 2002, GPE aims to enhance national education plans, improve aid effectiveness, and facilitate collaboration among diverse stakeholders.

Three of the selected evaluations are about Education Cannot Wait (ECW), a global fund established to address education in emergencies like conflicts and natural disasters. Launched in 2016, ECW links humanitarian and development efforts, providing emergency aid and focusing on long-term solutions. It has three funding windows: First Emergency Response, Multi-Year Resilience Program, and Acceleration Facility, each targeting immediate and systemic challenges to quality education in crisis situations.

Four evaluations concern five bilateral programmes: The Nexus Skills and Jobs Programme (NSJP), seeks to promote links between education and employment, supporting skills development and job opportunities for young men and women. The Middle East North Africa Scholarship Programme aims at building capacity through scholarships for mid-career professionals. The Netherlands Initiative for Capacity Development in Higher Education (NICHE) aims to strengthen institutional capacity in partner countries. The Netherlands Fellowship Programmes (NFP) offers scholarships for study and training in the Netherlands. The successor of both NICHE and NFP, the Orange Knowledge Programme, continues similar interventions, directed at collaboration, training, scholarships, and alumni support.

### Effectiveness

For the multilateral funds, for the GPE, the effectiveness was rated from mixed to good, improving over the years. For the ECW, the evaluations assess the effectiveness reasonably positively, although not every evaluation assesses all aspects or objectives.

In the bilateral programmes, evaluators' assessments of effectiveness also varied considerably. For NSJP, objectives were partially achieved, the assessment was positive for the objective of increasing employability but less so for linking trained youth to employment. The evaluation of NFP deemed the programme to have been effective at the individual as well as the organisational level. The evaluation of NICHE only looked at output level (for which effectiveness was assessed as good) since the activities had only recently started. MSP and OKP were not assessed for effectiveness in this synthesis study due to quality issues in their evaluation reports for this criterion.

Almost all programmes used capacity building to achieve their objectives. The multilateral funds were reported to have used capacity building to achieve their objectives and results at country and organisational level. For the bilateral funds, NFP was reported to have used capacity building to achieve objectives at individual and organisational level, whereas under NSJP, this was mainly reported at individual level. Project management had its bearing on effectiveness, in particular through having good structures, and providing clarity about roles, responsibilities, and processes. Implementation-related factors, such as too short duration and quality of training also affected effectiveness. Additionally, evaluators found having a well-defined ToC and a high-quality results framework essential for tracking progress and providing information for potentially necessary adaptations.

### Impact

In the assessed evaluation reports, impact was either not measured or the impact measurement was deemed insufficiently robust during the IOB quality assessment. Hence, this synthesis study cannot report on the extent to which the initiatives contributed to achieving goals at impact level. Some positive unintended effects have been reported, mostly relating to the intervention positively influencing the effect of other implementing agencies' activities. Negative side effects were seen in terms of funding replacement. In the NSJP evaluation, respondents perceived that an uneven coverage of all internally displaced peoples' communities had created a sentiment of exclusion. It should be kept in mind though, that such findings are only anecdotal.

### Efficiency

Findings on efficiency were generally positive, with some exceptions. Process and management received positive evaluations in almost all cases. Where interventions had not been timely, the reason was mostly found outside of the programme's scope of control, such as procurement complications. Though evaluators were generally positive about the relation between funds spent and achievements made, they did not provide clear conclusions related to value for money, cost efficiency and cost effectiveness. Having a clear division of roles and responsibilities was seen as conducive to efficiency. Lengthy and untransparent procedures and lack of availability of capable human resources had hampered efficiency.

### Relevance

The findings of all assessed evaluations are strikingly positive on relevance, with all interventions deemed highly relevant. All programmes were found to respond to the country needs related to education, as well as to the beneficiary needs. Most programmes adapted well to changes at various levels. Still, the needs were not always assessed in detail by the evaluators and alternative support channels were rarely discussed. Including participation of governments and local partners, considered crucial for relevance by the evaluations, was not always sufficient. ECW programmes especially were reported to struggle to involve civil society organizations and national NGOs as stakeholders and funding applicants. A factor that hampered relevance was dedicating insufficient time to programme preparation, including not having conducted adequate analyses at the onset of a programme or intervention. A positive factor was working with institutions that had already worked in a similar subject area and that thus had a certain experience, expertise and trained manpower.

### Coherence

For the multilateral funds, GPE and ECW were found internally as well as externally reasonably coherent, be it with imperfections. For the GPE, stakeholder participation was found a positive factor for promoting coherence. ECW evaluations deemed that, even if there is potential for the combination of FER and MYRP to address the humanitarian-development nexus, insufficient understanding and inclusion of this nexus in the results framework hampered coherence.

For the bilateral programmes, the evaluations observed some overlap between Dutch-funded interventions, but they were positive on the interventions' alignment with Dutch development objectives.

### Sustainability

For the multilateral funds, for GPE, sustainability is not part of the ToC. The fund strives to achieve sustainability through systems strengthening but has not specifically defined the necessary conditions for making improvements sustainable. Two of the ECW interventions are exclusively focussed on emergency situations, and thus do not pursue sustainability. In the ECW MYRP, some first steps towards sustainability have been made, but insufficiently addressing of sustainability in applicants' proposals was found to hamper overall sustainability.

Under the bilateral programmes, positive observations were made on sustainability of academic achievements of individuals, jobs created, behaviour change, and improvements in academic performance. Institutional sustainability was assessed positively in one of the programmes but less so in another.

For all assessed interventions, evaluations highlighted the absence of collaboration and coordination with public stakeholders, and limited capacity and existence of strong regulatory frameworks as a constraint to sustainability. Also all evaluations found financial sustainability of the interventions less likely.

### Gender

The multilateral funds met the planned requirement of gender equal targeting. Some also included activities aimed at improved gender knowledge and awareness. The employment of the Girls' Education Accelerator in the GPE approach was deemed gender sensitive. While ECW has made significant efforts towards mainstreaming gender, implementation has been challenging. A consistent observation for the multilateral funds was the lack of attention to addressing specific needs of male and female students and staff. Evaluators found that including gender into programme goals, policy, result frameworks and action plans was conducive to achieving contributions to gender equality. The evaluations assessed the bilateral programmes as gender sensitive. Most bilateral programmes also met their gender equal targeting requirement, although limited reporting of gender disaggregated data at outcome level have challenged creating fundamental changes.

### **Lessons learned**

Evaluations stress the importance of having a well-defined theory of change and a robust results framework, which are frequently lacking. These components facilitate accurate measurement, prompt adaptations, and enhance inclusivity and efficiency, making the programmes more attractive to potential donors.

While measuring impact at the level of beneficiaries and education system improvement is crucial, it was either not done or not done adequately in the incorporated evaluations. Addressing this requires baseline studies, follow-ups with counterfactuals, and integrating impact measurement into intervention designs, supported by sufficient funding.

For a number of interventions, there is scope to better engage local actors, also to achieve localization, which is an agreed goal under the Grand Bargain. Local engagement is deemed vital for achieving longer-term objectives and meeting DAC criteria.