



# Synthesis of UNRWA Evaluations 2016-2023

Department of Internal Oversight Services  
Evaluation Division

02/2024  
Evaluation Report



## Team members

Leslie Thomas, Former Chief, Evaluation Division, DIOS

Ronnie MacPherson, Evaluation & monitoring officer

Ref: DIOS/EVAL/2024/02

Dated: 24 December 2024

© 2024 UNRWA, all rights reserved

## About UNRWA

UNRWA is a United Nations agency established by the General Assembly in 1949 and mandated to provide assistance and protection to a population of over 5.7 million registered Palestine refugees. Its mission is to help Palestine refugees in Jordan, Lebanon, Syria, West Bank and the Gaza Strip achieve their full human development potential, pending a just and lasting solution to their plight. UNRWA services encompass education, health care, relief and social services, camp infrastructure and improvement, and microfinance. UNRWA is financed almost entirely by voluntary contributions.

**Table of Contents**

List of Acronyms.....IV

I. Introduction ..... 1

II. Purpose ..... 1

III. Approach ..... 1

IV. Common themes and findings.....2

1: Resource limitations are compromising performance, results and sustainability..... 2

2: Internal communication and coordination gaps ..... 3

3: Accountability to Affected Populations mechanisms are in place, but are not sufficiently developed ..... 4

4: More qualitative and outcome-level monitoring would be beneficial..... 4

5: Limited consultation during programme / project design..... 5

6: External partners would appreciate more strategic engagement and clearer communication..... 5

7: UNRWA’s symbolism amongst refugees is an enabler but also a constraint, particularly when changes are made to services ..... 6

VI. Observations on evaluation quality ..... 7

1. Evaluation report structure ..... 7

2. Evaluation report content and length ..... 7

3. Evaluation recommendations ..... 8

4. Quality assurance process ..... 8

5. Best practice reports ..... 8

Annex A: Evaluations reviewed ..... 9

## List of Acronyms

|       |  |
|-------|--|
| AAP   | Accountability to Affected Populations   |
| DAC   | Development Assistance Committee   |
| DIOS  | Department of Internal Oversight Services                                      |
| GFO   | Gaza Field Office  |
| HQ    | Headquarters   |
| JFO   | Jordan Field Office  |
| LFO   | Lebanon Field Office   |
| MTS   | Medium-Term Strategy   |
| OECD  | Organisation for Economic Co-operation and Development                         |
| QA    | Quality Assurance  |
| RBM   | Results-Based Management   |
| SEA   | Sexual Exploitation and Abuse  |
| SFO   | Syria Field Office   |
| TOR   | Terms of Reference   |
| UNEG  | United Nations Evaluation Group  |
| UNRWA | United Nations Relief and Works Agency for Palestine Refugees in the Near East |
| WBFO  | West Bank Field Office   |

## I. Introduction

The Evaluation Division of the Department of Internal Oversight Services (DIOS) assesses UNRWA’s policies, programmes and activities to determine the extent to which the Agency’s interventions achieve their intended results, and to identify factors that can strengthen or constrain performance. In addition to providing accountability, evaluations generate evidence and learning that supports management decision-making, informs planning processes, and strengthens the Agency’s ability to plan strategically.

A core part of the Evaluation Division’s work is the delivery and/or oversight of **independent evaluations**. These comprise **centralized evaluations** (managed directly by the Evaluation Division) and **decentralized evaluations** (managed by UNRWA programmes, field offices or other operational units, with the support of the Evaluation Division). Reflecting the diversity of UNRWA’s operations, evaluations can be targeted at a variety of themes and organisational levels, ranging from discrete projects (e.g. an evaluation of UNRWA’s ‘Engaging Youth in Syria’ project), to programmatic processes (e.g. an evaluation of the Family Health Team approach), to overall strategy (e.g. an evaluation of UNRWA’s Medium-Term Strategy).

## II. Purpose

While most evaluation findings will only be of direct relevance to the subject being assessed, **evaluation syntheses** can be used to identify **generalisable findings, themes and recommendations** that repeatedly arise across different evaluation reports. In turn, this process can help to identify broader **lessons, strengths, and areas for improvement**. The following paper firstly documents the central findings from a synthesis of the centralised and decentralised evaluations undertaken across UNRWA from 2016-2023.

Following that synthesis, the paper then presents summary observations on the **quality** of the evaluations reviewed for the synthesis. Common **strengths and weaknesses** are identified, along with some potential **considerations** to take into account when designing and commissioning future evaluations.

## III. Approach

The synthesis was based on a **desk review** of all 20 evaluation reports that were completed during the period 2016-2023, and that met UNRWA’s internal evaluation quality standards (see Annex A for full list). With the support of qualitative data analysis software, individual datapoints (e.g. quotes, findings, recommendations) were classified against a series of **synthesis categories**.

| Synthesis category       | Description   |
|--------------------------|---|
| Strategic objective (SO) | Datapoints were coded against the relevant 2023-2028 SOs. Even though most evaluations preceded this strategic period, there is considerable overlap between SOs used in different periods and – in any case – the synthesis aims to inform work in the current strategic period. |
| Cross-cutting themes     | Whether the datapoint was relevant to cross-cutting themes of accountability to affected populations (AAP), sexual exploitation and abuse (SEA), gender equality, disability inclusion, youth, environmental protection, value for money, resource mobilization.                  |
| Office                   | Whether the datapoint was relevant to headquarters (HQ), Gaza (GFO), Jordan (JFO), Lebanon (LFO), Syria (SFO), West Bank (WBFO).  |
| Enabler or constraint    | Where appropriate, whether the datapoint represented an element that enables UNRWA’s work, or constrains UNRWA’s work.  |
| Synthesis unit           | Whether the datapoint was a finding, conclusion, recommendation, lesson or result.  |

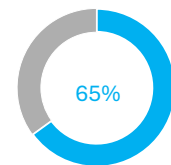
These classifications supported the **organisation and aggregation** of all datapoints, which in turn enabled the identification and analysis of commonalities and trends arising across the evaluation reports.

The paper mainly focuses on discussing the **seven most frequently identified themes and generalisable findings**. For each finding the proportion of evaluations containing relevant datapoints is also confirmed, serving as a proxy indicator of the **prevalence of each finding** across the reviewed evaluations. Given that evaluations inherently focus on **accountability** and **learning**, the most commonly identified findings all relate to **areas for improvement** across UNRWA's work.

## IV. Common themes and findings

### 1: Resource limitations are compromising performance, results and sustainability

% reports identifying this finding



Evaluations invariably found that UNRWA's work was highly relevant to its mandate and the populations that it serves. However, the majority of evaluations also directly linked UNRWA's current and recent resource limitations to tangible, negative and sometimes significant effects on the performance of programmes, projects and activities. In all instances, it is clear that resource limitations compromised the extent to which results were achieved. Many evaluations specifically identified the loss of funding from UNRWA's largest donor, the USA, in 2018 as a particularly damaging juncture. Typically, the effects of resource constraints were compounding, undermining multiple facets of any given operation. The most commonly identified negative consequences included:

- Short-term projects were having to be used as a means for delivering long-term services. In many instances this necessitated appointing staff on short-term contracts, undermining the ability to attract suitable candidates and resulting in considerable uncertainty for those staff that were hired. Short project cycles also resulted in staff time being diverted into frequent fundraising.
- Having to address escalating needs with reduced resources resulted in significant strain on UNRWA staff. This was worsened by the lack of resources available for UNRWA to work on staff well-being.
- Staff felt that the unpredictability of services resulted in reduced satisfaction amongst UNRWA beneficiaries, potentially extending to an erosion of trust in UNRWA. While this issue was directly noted in the e-card evaluation and the Family Health Team evaluation, several other evaluations identified service unpredictability as a significant risk for UNRWA.
- UNRWA's work on emergency preparedness was compromised, particularly within designated emergency shelters. There were insufficient resources to maintain non-food item stocks, to ensure shelter security, or to adequately monitor the condition of facilities.
- Limited resources reduced UNRWA's capacity for adaptive management, with little room to adjust programmes. As noted in – for example – the protection audit evaluation, this sometimes undermined UNRWA's capacity to address recommendations: *"the financial crisis has impacted the ability of fields and programmes to respond to and implement recommendations"*.



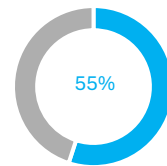
“UNRWA’s chronic funding crisis is having an impact on the quantity of assistance, the ability to maintain service quality and the ability to invest in operations. While UNRWA could take steps to strengthen resource mobilization, ultimately, it is reliant on donors to deliver on the recommendations and commitments to which they have signed up.”

Evaluation of the Medium-Term Strategy 2022-2016



## 2: Internal communication and coordination gaps

% reports identifying this finding



Most evaluations generally commended UNRWA’s coordination, with the Medium-Term Strategy (MTS), strategic objectives and Results Based Management (RBM) system commonly identified as effective structures for directing and organising UNRWA’s work. A core finding of the *processes that ensure that the MTS guides operational plans, and that field office and programme operational plans are coherent*. Despite those strong high-level constructs, a majority of evaluations also identified some shortcomings with internal coordination and – in particular – internal communications.

While UNRWA’s decentralized approach was not seen as a weakness in itself, several evaluations attributed performance shortcomings to internal communication gaps between UNRWA’s various operational tiers (frontline staff, Area Offices, Field Offices, HQ). On occasion, these gaps resulted in a lack of trust between the tiers, and undermined the efficiency and effectiveness of coordination. Some evaluations also identified coordination risks and trade-offs arising due to UNRWA’s decentralized model. The Security Risk Management evaluation found that *“the determination of acceptable risk is not currently determined in a systematic manner”* due to *“the operationalization of safety and security [differing] between the fields of operation”*. Similarly, the evaluation of the Family Health Team reform identified risks around inconsistent gender-based violence referral pathways: *“they have been developed to meet the requirements of each field rather than being centrally designed”*.



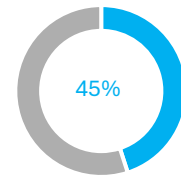
“[Interviewees] believe that Field Offices should make a greater effort to engage with [frontline] staff, by phone or through virtual meetings. Weak communication between some Field Offices, Area Offices and [frontline] staff has resulted in their feeling undervalued and underappreciated.”

Evaluation of the Family Health Team reform



### 3: Accountability to Affected Populations mechanisms are in place, but are not sufficiently developed

% reports identifying this finding



Evaluations routinely found that UNRWA had put in place feedback and complaint mechanisms for refugees. However, those evaluations that explored these mechanisms invariably *also* found that the mechanisms were inadequate, and/or that UNRWA’s broader approach to Accountability to Affected Populations (AAP) could be strengthened. The most commonly identified shortcoming was a lack of transparency and feedback. Refugees were generally unaware as to whether and how their issues were being addressed, which in some instances had resulted in a loss in refugees’ confidence in the AAP processes. Several evaluations also found that mechanisms were not fully inclusive and were inadequate for harder-to-reach groups and populations, particularly those with disabilities and refugees living outside camps.



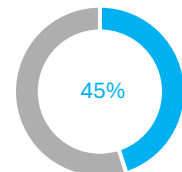
“Beneficiary awareness and confidence in Agency feedback and complaint mechanisms is weak... feedback gathered through AAP activities is not adequately synthesized or shared within the field or headquarter management structures to inform decision-making within the Agency”

Evaluation of the monitoring and reporting of the Medium-Term Strategy



% reports identifying this finding

### 4: More qualitative and outcome-level monitoring would be beneficial



Evaluations generally found the Agency’s RBM and results reporting to be a strength. However, many evaluations also identified limitations of the system being skewed towards activity and output monitoring, and an insufficient emphasis on outcome-level and qualitative reporting. Most commonly, evaluations found a notable gap being the lack of data being gathered around beneficiary satisfaction and perceptions. Some evaluations also identified missed opportunities to undertake longer-term, post-intervention outcome monitoring around – for example – the career trajectories of participants in UNRWA vocational training.



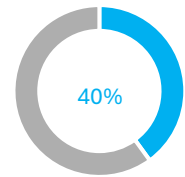
“While [the intervention’s results frameworks] provide a relatively solid basis for assessing the effectiveness of delivery at the output level, they cannot be used to provide an assessment of outcomes”

Evaluation of the Emergency Appeals for the oPt and Syria Regional Crisis



% reports identifying this finding

## 5: Limited consultation during programme / project design



Many evaluations found that interventions were undermined by a lack – and often complete *absence* – of consultation with beneficiaries during the design phase. On occasion, this resulted in solutions and responses that were inappropriate, or even contrary to community expectations. For example, the evaluation of the Jabal Al Taj school replacement project found that negative reactions to a new school site were largely due to poor consultation: “*the entire community involvement and community participation was missing...there was no meaningful involvement of children or participation in the building of their own school*”. Equally, evaluations also found that beneficiaries often had no involvement during interventions’ monitoring and evaluation.

Gaps in consultation with UNRWA staff were also evident. Highly relevant frontline staff perspectives were often absent from intervention design, ultimately resulting in design gaps, but also a degree of resentment amongst the staff that were delivering the actual work.



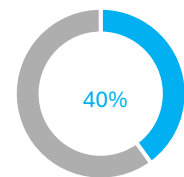
“While partners acknowledged UNRWA’s strength as a service provider, they were critical of UNRWA as a partner. This was in relation to UNRWA’s lack of participation in national forums, its information sharing, lack of communication, accountability mechanisms, and transparency.”

Mid-term evaluation of Madad phase III



% reports identifying this finding

## 6: External partners would appreciate more strategic engagement and clearer communication



External partners routinely praised UNRWA’s work, but evaluations also frequently found a degree of frustration amongst partners. UNRWA’s approach to partnership was commonly viewed as being reactive and transactional rather than strategic. For example, the Madad Phase II evaluation found that there was “*limited cooperation with government authorities in [Jordan and Lebanon] in terms of planning and setting complementary strategies*”. The lack of long-term planning with partners had resulted in missed opportunities, and there was sometimes a perception that UNRWA was failing to appreciate or capitalise on the technical competencies and comparative advantages of other agencies. Partners also called for more systematic communications from UNRWA, particularly around lesson learning: there was a sense that other agencies (and their beneficiaries) had much to gain from a better understanding of UNRWA’s knowledge and experience.



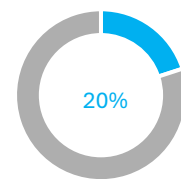
“Frontline staff spoke of their minimal involvement in the [intervention] design... being the staff members with the most contact with [beneficiaries], their views about a potential vulnerability assessment scheme could be most useful, but... their input was not requested or utilised”

Final evaluation of Madad phase II



## 7: UNRWA’s symbolism amongst refugees is an enabler but also a constraint, particularly when changes are made to services

% reports identifying this finding



Some evaluations found that one of UNRWA’s most significant, continuing impacts was the Agency’s “*stabilizing influence [which] creates a basis for social stability and cohesion among vulnerable populations*” (Evaluation of the MTS). However, the criticality of UNRWA to many beneficiaries’ lives and livelihoods was also found to place constraints on the Agency’s ability to remove or even change existing services. Evaluations found that service developments were often perceived by beneficiaries to “*portend reduced support from the international community and spell the ‘beginning of the end’ of UNRWA*” (Evaluation of the e-card modality).



“Any changes in programme delivery is perceived by some beneficiaries as another step away from caring, respecting and protecting Palestinian basic rights to education, health, relief and sanitation services. Any real or potential reduction of access or quality of services affects the attitude of the beneficiaries since it has a strong symbolic impact on them.”

Evaluation of Security Risk Management



## V. Sector-specific findings

Although not as prevalent as the generalisable findings identified above, some sector-specific findings were found in multiple evaluations:

- Those evaluations that focused on **vocational training** invariably found that UNRWA’s vocational training offerings were not always well-linked to market demands and would have benefited from deeper consultation and market analysis during the design phase.
- Only two evaluations focused on **infrastructure projects**. However, both identified significant shortcomings with UNRWA’s management of procurement, and of the construction process. Contracts were awarded above

market rates, and construction management did not always make use of industry standard oversight controls and processes, sometimes resulting in delays and poor-quality outputs.

- Evaluations generally found UNRWA's work on **gender** to be strong, with – for example – the MTS evaluation finding that “UNRWA has mainstreamed gender into its programmes and, despite the lack of funding, has strengthened staff capacity on gender issues and Gender-Based Violence”. Gender disaggregated monitoring was also a strength, with the MTS monitoring and reporting finding that “[UNRWA’s] results framework and...reporting provides rich levels of information by gender”. However, evaluations also often found a tendency to conflate ‘gender’ with ‘women’, and that men’s involvement and perspectives were rarely accounted for during the design or implementation of gender-focused activities. This was subsequently linked to a lack of support amongst some men for UNRWA’s gender-focused activities.

## VI. Observations on evaluation quality

### 1. Evaluation report structure

Evaluation reports that were structured against **Organisation for Economic Co-operation and Development's (OECD) Development Assistance Committee (DAC)** criteria (or against a similarly simple, limited set of criteria/questions) tended to be of better quality. Criteria-based report structures helped to ensure clarity of analysis and kept the narrative focused on discussing evaluative findings, rather than straying into (e.g.) detailed activity or process descriptions. Conversely, when evaluation criteria were *not* used, reports tended to be sprawling and less focused, were consequently more difficult to follow, and often shifted away from presenting evaluative findings towards discussing interventions in more general terms.

### 2. Evaluation report content and length

The main body of several evaluations exceeded (sometimes greatly) the current UNRWA guidance of 40 pages. Page length is by no means a proxy of evaluation quality, and in some instances (for example the MTS evaluation) the length was justified by the scope and content, and did not actually compromise accessibility or readability. However, in several instances **report length was excessive and unnecessary**, ultimately reducing the readability and utility of the reports.

A number of factors contributed to reports becoming too lengthy. Most commonly, several reports allocated too much space to highly detailed descriptions of the intervention, sometimes extending to forensic descriptions of processes and activities: this level of detail was rarely (if ever) necessary. The target audience (i.e. staff and donors involved in the intervention’s oversight and delivery) will be well-familiar with the intervention, and won’t require such a detailed background. An unfortunate consequence of this approach tended to be that the actual evaluative findings and analysis got lost in the thickets of contextual detail. In some instances the reporting was exclusively descriptive, with no evaluative substance at all. This tended to happen when subject experts rather than evaluators took the lead in report authoring. There is clear, significant value in having subject experts (e.g. construction professionals, health experts) as part of an evaluation team, but analysis and reporting should almost always be led by an experienced evaluator.

As above, lengthy reports are sometimes justified by an evaluation’s scope and content. However, a useful heuristic is that the longer a report gets, the less useful it becomes for the target audience. DIOS could consider reducing their page limit guidance from 40 to 30 pages: such a reduction *should* help evaluators to better focus their narrative, and would also serve as a clear signal around DIOS’s expectations. Along the same lines, DIOS could consider implementing a hard limit of 3 pages for executive summaries, even if this necessitates removing elements such as methodology. Where multi-faceted reports are unavoidably (justifiably) lengthy, DIOS could consider requiring evaluators to provide short (2-3 page) briefing notes on the report’s main findings. For example, if a lengthy evaluation report covers multiple UNRWA strategic objectives, perhaps standalone briefings could be generated for each strategic objective.

### 3. Evaluation recommendations

There was substantial variance in the scope and framing of evaluation recommendations. Some reports presented over-generalised recommendations that essentially had no utility for UNRWA (e.g. “Continue the work to secure donor funding for the future”). Conversely, other reports provided numerous and/or highly detailed recommendations that placed a burden on UNRWA that was sometimes disproportionate to the shortcomings identified. Overly prescriptive recommendations also left little room for UNRWA to adjust and improve according to their own (closer) familiarity with the context and work.

More consistency would be beneficial here: a uniform approach to recommendation formulation should help to ensure interventions are treated equally and that the recommendation ‘burden’ is proportionate to the issues identified. A possible approach would be to provide **example recommendation formats** as an annex to evaluation TORs. It may also be worth taking a **more consultative approach**: evaluations identify shortcomings and the reasons for those shortcomings, evaluators propose headline recommendations to address the shortcomings, but the recommendations are then refined and finalised on the basis of close consultation with the intervention team – what is realistic and achievable in the intervention’s context?

### 4. Quality assurance process

UNRWA’s template-based evaluation quality assurance (QA) process is directly comparable to those applied by other UN and development agencies. The QA template covers all the core elements of evaluation quality, and helps to ensure a well-structured, systematic review process. QA reports were available for 10 of the 20 evaluations reviewed for the synthesis. All QA reports were sufficiently clear and generally in line with the synthesis author’s informal assessments of report quality. If anything, the QA reports skewed slightly positive (but there is always going to be a degree of subjectivity here!).

QA reports identified a recurring (though not universal) evaluation report weakness as being the limited extent to which evaluations incorporated and addressed **gender, disability, and/or human rights**. This gap in evaluation reports could plausibly be linked to a lack of requirement / reference to these themes within evaluation TORs. DIOS may want to ensure that future evaluation TORs (centralised and decentralized) formally include requirements around addressing gender, disability and human rights. TORs should also point tenderers towards the evaluation guidance that UNRWA applies for these themes, whether that guidance is internal or from another source (e.g. UNEG) – signposting prospective evaluators towards such guidance can help with setting expectations.

The most recent iteration of the QA template includes an assessment criterion asking “**Is the report accessible and written in a clear manner?**” (this criterion was not applied within earlier versions of the QA template). However, the criterion is one amongst many and – on its own – does not influence the overall QA score greatly. But given the above-noted shortcomings with evaluation length, focus and readability, DIOS may want to consider placing more weight on this criterion. While many of the other criteria are far more important from DIOS’s perspective, this ‘accessibility’ criterion is arguably one of the most important from the target audience’s perspective.

### 5. Best practice reports

Across the cohort of 20 evaluations reviewed, the following reports were of particularly good quality. Their structure, approach and content could be instructive for UNRWA teams that are in the process of commissioning, overseeing or reviewing evaluations.

- **Evaluation of the Medium-Term Strategy 2016-2022**: Although relatively lengthy, the report is focused and comprehensive, with all findings and recommendations very clearly linked to evidence.
- **Evaluation of UNRWA’s Gender Initiative 2014-2017**: A well-structured report with clear analysis throughout. The report also provides an informative cost efficiency analysis, which is rarely present in UNRWA evaluations. However, the report’s recommendations are overly general.
- **Evaluation of the UNRWA Engaging Youth Project in Syria**: Another well-structured, comprehensive report, benefiting from use of the OECD-DAC evaluation criteria.

## Annex A: Evaluations reviewed

### I. Centralized evaluations

Evaluation of the E-Health Project (March 2016)

Evaluation of the Security Risk Management Programme (March 2016)

Evaluation of UNRWA's Transition to the E-Card Modality in the Jordan, Lebanon and West Bank fields (September 2018)

Evaluation of UNRWA Monitoring and Reporting of the Medium-Term Strategy 2016-2022 (March 2020)

Real time evaluations of UNRWA responses to the COVID 19 emergency (April-July 2020)

Evaluation of the UNRWA Protection Audit Mechanism (April 2021)

Evaluation of the UNRWA Medium-Term Strategy 2016-2022 (September 2021)

Evaluation of the UNRWA Family Health Team Reform (October 2021)

Review of the UNRWA Management of the Emergency Response to the 2021 Gaza-Israel Conflict (October 2021)

Evaluation of UNRWA Support to Vulnerable Children in Gaza funded by Islamic Relief USA 2015-2021 (February 2023)

Evaluation of the UNRWA Emergency Appeals for the Occupied Palestinian Territories and Syria Regional Crisis 2016-2021 (March 2023)

### II. Decentralized evaluations

Final Evaluation of Improvement of Living Conditions of Vulnerable Palestinian Refugees in Jerash Camp Project (2017)

Decentralized Evaluation of UNRWA Gender Initiative 2014-2017 (February 2018)

Evaluation of UNRWA's Engaging Youth in Syria Project (June 2018)

Project Evaluation Report: Replacing Rented Schools at Jabal Al Taj (September 2018)

Final Evaluation of Madad Phase I (January 2019)

Evaluation of UNRWA Child and Family Protection Services (June 2019)

Final Evaluation of Madad Phase II (July 2020)

Mid-term Evaluation of Madad Phase III (March 2022)

Final Evaluation of the Project 'Support to the Training and Professional Development of Teachers and the Teaching of French for Quality Education in UNRWA Schools, Year 1 and Year 2' (May 2022)