

EVALUATION DEPARTMENT



REPORT 9/2015



Photo: Ken Opprann

Evaluation Series of NORHED: **Evaluability Study**

NORHED Evaluability Study

Development Portfolio Management Group

November 2015

ISBN: 978-82-7548-819-8

Acknowledgements

This report rests on the contributions of several individuals. Norad's Evaluation Department, the FORSK staff that manages NORHED, and the staff of Norad's Results Management Section generously met with the team, answered email queries, and ensured that the team had access to the documents needed.

The Development Portfolio Management Group managed the work. The DPMG team consisted of Sue Berryman, who managed the task, Juan Saavedra, Milda Jonusaite Nordbø, and Tarra Kohli.

Xavier Legrain provided a quality assurance review of the draft document.

This report is the product of its authors, and responsibility for the accuracy of data included in this report rests with the authors. The findings, interpretations, and conclusions presented in this report do not necessarily reflect the view of Norad's Evaluation Department.

Table of Contents

Acknowledgements	ii
Table of Contents	iii
Acronyms and Abbreviations	v
Executive Summary	vi
I. Introduction	1
A. Purpose.....	1
B. Limitations.....	1
C. Organization of the report	2
II. Framework for Assessing the Evaluability of NORHED	3
A. The purpose of the evaluation determines the criteria for judging evaluability	3
B. Learning is the primary purpose of evaluating NORHED, with an ancillary accountability purpose	4
III. Results for Re-analysis of Adequacy of Grantees' Results Frameworks	10
A. Prior Analysis	10
B. Current Analysis.....	11
C. Conclusions	17
D. Recommendations	18
IV. Analysis of Quality of the Standard Indicators	19
A. Prior Analysis	19
B. Relevance of NORHED's Standard Indicators Revisited	20
C. NORHED's Standard Indicators Revisited: Findings	21
D. Conclusions	28
E. Recommendations	29
V. NORHED Supervision Processes	30
A. Introduction	30
B. Findings	31
C. Conclusions	35
D. Recommendations	35
VI. Can NORHED be adequately monitored and evaluated?	37
A. Specific Conclusions.....	37
B. Overall Conclusion.....	37
References	38
Annex A: Terms of Reference	40
Annex B: Amendment to DPMG Proposal	45
Annex C: Annual Progress Report for NORHED Projects	47

Figures

Figure 1: Norad's Results Chain	7
Figure 2: Overall Quality of RFs – Original Evaluation Framework	11
Figure 3: Overall Quality of RFs – Reduced Evaluation Template	12
Figure 4: Overall Quality of RFs – Original Evaluation Framework with Updated #11 and #12	14
Figure 5: Overall Quality of RFs – Reduced Template with Updated #11 and #12.....	14
Figure 6: Overall Quality of RFs – Original Evaluation Framework with Revised Items #5 and #8.....	15
Figure 7: Overall Quality of RFs – Original Evaluation Template with Revised Items #5 #8, #11, #12.....	16

Tables

Table 1: Data required for evaluating NORHED and NORHED data sources by stage of the project cycle	8
Table 2: Evaluation Template (Yes/No) with Percentage of Projects Complying	10
Table 3: Seven Norad Requirements with Percentage of Projects Complying.....	11
Table 4: Rejected Evaluation Criteria and Corresponding Requirements of NORHED's RF	12
Table 5: NORHED Standard Indicators with Numerical Values.....	13
Table 6: Original and Revised Mean for Criteria 11 and 12 in Evaluation Template	14
Table 7: Original and Revised Mean for Criteria 5 and 8 in Evaluation Template	15
Table 8: Evaluation Template (Yes/No) with Percentage of Projects Complying for Original and Revised Scoring	16
Table 9: Mean and Standard Deviation of Relative Score	17
Table 10: Steps and products of FORSK's annual supervision process	30
Table 11: Crosswalk between data requirements during implementation and template items in NORHED annual progress report	32

Acronyms and Abbreviations

CREAM	Clear, Relevant, Economic, Adequate and Monitorable
DPMG	Development Portfolio Management Group
EOI	Expression of interest
FORSK	Section for Research, Innovation and Higher Education (Norad)
ICT	Information and Communication Technology
IDA	International Development Association
IRC	Independent Review Committee
LMIC	Low and middle income country
M&E	Monitoring and Evaluation
NOK	Norwegian Kroner
NOKUT	Norwegian Agency for Quality Education
NOMA	Programme for Master Studies (Norad)
Norad	Norwegian Agency for Development Cooperation
NORHED	Norwegian Programme for Capacity Development in Higher Education and Research for Development
NUFU	Norwegian Programme for Development, Research and Education
PDO	Project Development Objective
RFF	Revised Results Framework

Executive Summary

A. Purpose

The purpose of this report is to summarize the evaluability status of the Norwegian Program for Capacity Development in Higher Education and Research for Development (NORHED). This status is assessed relative to a theory-based evaluation model. Against standards of the types and quality of information required for evaluating NORHED, this report assesses the results frameworks for the 45 NORHED projects, the 14 NORHED indicators common across the projects, and, within narrow limits, the processes that Norad's Section for Research, Innovation and Higher Education (FORSK) uses to supervise the projects.

B. Methodology

The methodology for assessing the evaluability of NORHED consists of the data required at each stage of the project cycle in order to judge NORHED's performance. This methodology is based on Norad's evaluation intent for NORHED and, given the design of NORHED, on the only evaluation model available: a theory-based evaluation model.

Under this model, the burden of assessing attribution or contribution has to be carried by: a) a clear and plausible theory of change for the program; and b) a careful monitoring of whether and how the links in the causal path of this theory of change play out. This approach to attribution assumes that if the causal path is plausible and the links in the chain expected to produce certain results can be demonstrated to have materialized, the project/program can tentatively be concluded to have caused the observed changes. The greater the fidelity of how the program/project unfolds to the theory of change, the greater the confidence that results by the program's end can be attributed to the program.

NORHED in fact consists of the overall program's theory of change and 45 project-specific theories of change. All projects have the same development objectives, albeit with different targets. However, each project has a different theory of change that is shaped by the sector for the grant—e.g., health or natural resource management—and the means that the grantee selects to build capacities within the partnership.

These realities put a premium on: a) a well-defined and well-measured theory of change for NORHED overall, and b) since the achievements of the individual grants define the success of the overall program, on well-defined and well-measured theories of change for the individual projects. Operationally, the projects' results frameworks are treated as their theories of change.

C. Limitations

This report has certain limitations.

- 1) This report was not based on an in-depth review of a representative sample of universities and projects. In March 2015 a member of the DPMG team conducted a quick field visit to an unrepresentative and small sample of countries, universities, and projects that excluded visits to the Norwegian partners for the projects. This quick scan surfaced problems with the indicators and potentially serious NORHED program design issues that could have affected NORHED's evaluability. Although these tentative findings would have warranted an in-depth review of a larger and more representative sample of universities and projects, budget limitations precluded the fieldwork required to carry out such a review.
- 2) DPMG was unable to pretest proposed modifications of the NORHED standard indicators.
- 3) Desk reviews should be complemented by field visits to a small sample of projects in order to optimally assess the supervision process. However, budget limitations for this exercise precluded both desk reviews and field visits. This reality translates into providing an incomplete picture of the evaluability of NORHED. FORSK, which manages NORHED, will lack the feedback that it needs to

affirm the quality of its supervision process and/or to modify particular aspects of the process to increase its validity and reliability.

D. Key Findings

Quality of projects' results frameworks

Unclear criteria for judging project results frameworks. The criteria against which grantees' results frameworks should be judged are not clear. The results matrix (section 3) in NORHED's main application form works with the entire results chain except for inputs/resources, but Norad's Results Management Section indicated that they held grantees accountable for only outputs and outcomes. Thus, their frameworks should not be judged against criteria associated with the fuller results chain—for example, whether the causal connections between activities and outputs were traceable and plausible.

Good news from the evaluation of grantees' results frameworks. More than 8 out of every 10 grantees had a clear statement of the problem or challenge that the project was trying to solve. This result is to the credit of the grantees, FORSK, and Norad's Results Management Section.

Less good news from the evaluation of grantees' results frameworks. Grantees' results frameworks showed a pattern of weak causal connections between the links in their results chain. Given a theory-based evaluation model, clarity about the entire results chain is crucial for evaluability purposes. Weak results frameworks attenuate the evaluability of NORHED. Outcomes are clearly stated and linked to the problem in only 3 out of every 10 projects. In only four out of every 10 projects the outputs are clearly linked to the outcomes. Although reviewing the implementation plans improved the rating for the links between activities and outputs, again only about 3 out of every 10 projects showed acceptable linkages.

Reviewing the first year's annual reports for baselines for outputs and outcomes markedly improved the ratings of projects on these dimensions. However, despite significant efforts by FORSK and the Results Management Section to obtain baseline data so critical to the evaluation of the projects, a third of the projects still lack complete baselines for outputs. Two out of every 5 projects lack complete baselines for outcomes. Only 2 out of every 5 projects had final year targets for their outputs and outcomes.

Relevance and quality of the NORHED standard indicators

In consultation with grantees, Norad developed 14 standard indicators for capturing the outputs and outcomes of the projects. This was an excellent process, and the 14 indicators were found to be substantively relevant, although incomplete relative to NORHED's objectives. The indicators did not measure quality, as opposed to amount, of research; quality of educational programs as a check on the quality of graduates; sustainability of NORHED-funded capacity developments, and system and institution-level conditions that enable or impede capacity building at the university level.

The main problems lay with how the indicators are operationalized. Criteria known as CREAM (Clear, Relevant, Economic, Adequate and Monitorable) were used to evaluate the indicators as specified. The first annual progress reports were also reviewed to determine how grantees were interpreting (or misinterpreting) requests for updated M&E data on each indicator. Against these standards, only four of the 14 indicators met CREAM standards (indicators 1, 3, 5, and 8), although three others can probably be quite easily modified to eliminate measurement problems. Chapter IV reports the detailed results for all 14 indicators.

The indicators had problems that commonly arise in survey design, the most problematic being:

- 1) Ambiguity in what NORHED is requesting from grantees.
- 2) Ceiling effects--i.e., where the desired outcomes have already been achieved by the time that the projects start, leaving little or no opportunity for variation between projects or over time within projects;
- 3) Measurement error when grantees are asked to make calculations;

- 4) Questions framed in ways that encourage selective reporting by grantees; and
- 5) Lack of guidance to grantees on how to respond to qualitative items. Qualitative items are not always asked to elicit from grantees a detailed discussion of the change process. If the change process is poorly detailed, the causal connections between inputs/activities and output/outcomes cannot be credibly established.

Quality of NORHED supervision processes

Within the limited scope of this exercise, the following findings emerged.

The annual narrative report template collects data on an appropriate set of topics. It does not collect data on the topics important to its goals or for interpreting results identified earlier. In some cases the questions as phrased should yield valid and reliable data, but in a number of cases, especially for items that request updated data on the standard indicators, they will not. In most cases the M&E data collection schedule is defined and makes sense, with the data collection schedule being undefined for only two standard indicators.

FORSK has no dashboard that summarizes "at a glance" each project's implementation status and likelihood of reaching its development objectives. FORSK has commissioned annual summaries of the results of project implementation relative to the standard indicators and other Norad priorities. These summaries will yield important information for the NORHED management, but they aggregate data across the projects for the program as a whole. This is not the same as the proposed dashboard idea, which would discriminate between projects that seem on track to succeeding and those that, for whatever reason, are in trouble and thus warrant particular FORSK attention.

Grantees self-report indicator data. The literature identifies potential validity problems with self-ratings. Although FORSK staff review these data, these reviews have limits. How FORSK writes the TORs for independent mid-term and/or final reviews of the program will determine if they will obtain independent checks on the validity of grantee-provided data.

FORSK is considering commissioning independent tracking of whether aid-funded interventions actually benefit intended beneficiaries. Commissioning such studies is good aid management practice.

Conclusions: Evaluability of NORHED

Conclusions about whether NORHED can be adequately monitored and evaluated are based on the data required to evaluate NORHED, as identified in the framework specified in this report for judging the evaluability of NORHED.

Specific Conclusions

Theory of change for NORHED. With some caveats, the implicit theory of change for NORHED has been judged to be adequate, relative to the international literature on theories of change for capacity development at the tertiary level (Norad, 2014). The major omissions were system-level and institution-level conditions that enable or impede capacity development at the university level.

Theories of change for the NORHED projects. The evaluation of the results frameworks that operationalized grantees' theories of change found a pattern of weak links along the results chain. Since clarity about the entire results chain for each project is crucial for evaluating the overall NORHED program, the weak results frameworks of many of the projects attenuate program evaluability.

Performance indicators common to all projects. The indicators are substantively relevant, but did not measure four variables germane to NORHED's objectives and evaluability. Only four of the 14 standard indicators were well operationalized, meaning that the data yielded by the remaining 10 indicators will not be interpretable or will not register the range of change that NORHED intends to allow.

Baseline data for output and outcome indicators. Although baselines are the *sine qua non* for evaluating project and program achievements, a third of the projects lacked complete baselines for outputs, and two out of every five projects lacked complete baselines for outcomes.

Data to monitor and evaluate project implementation. Measurement problems with the standard indicators reduce the validity and reliability of data generated during implementation. Depending on how it writes the TORs for independent midterm/final evaluations of the program, FORSK may or may not have ways to protect against self-report bias on the part of grantees.

Overall Conclusion

Using a binary scale to judge whether NORHED can or cannot be evaluated is inappropriate. This review shows that some aspects of NORHED can be properly evaluated. It also shows that several standard indicators for the program do not meet criteria for well-designed indicators; that baselines for output and outcome indicators for the standard indicators are often missing; and that grantee-generated data on outputs and outcomes may or may not be independently verified. Revising the standard indicators to meet criteria for valid and reliable measurement would have the single biggest effect on improving NORHED's evaluability.

Recommendations

Recommendations 1 and 2 are addressed to Norad as a whole.

- 1) **For most programs, Norad should require results frameworks that reflect the full results chain.** Many Norad aid programs will be able to sustain only a theory-based evaluation model. This model requires attention to the entire results chain in order to draw even remotely credible conclusions about attribution/contribution of the aid program/project. It assumes that if the causal path is plausible and the links in the chain expected to produce certain results can be demonstrated to have materialized, the project/program can tentatively be concluded to have caused the observed changes.
- 2) **A Norad program should use a consistent result framework.**

Recommendations 3 to 7 are directed to FORSK and Norad's Results Management Section.

- 3) **Revise the standard indicators in the annual progress report template to maximize the reliability and validity of M&E data.**
- 4) **Add data collection items to the annual progress report template to measure the quality of NORHED-funded new/revised educational programs as a check on the quality of graduates, sustainability of NORHED-funded capacity developments, and system and institution-level conditions that enable or impede capacity building at the university level.**
- 5) **Consider creating a dashboard that summarizes the health of the individual NORHED projects by having FORSK officers assign summary ratings annually to each grant.**
- 6) **If the TORs for the planned independent midterm and/or end of project evaluations do not check the accuracy of the data provided by grantees, protect the program against the inevitable positive bias associated with self-reported data** by commissioning independent checks of output and outcome data for a random (or purposive) sample of projects.
- 7) **FORSK is now considering commissioning independent tracking of whether NORHED interventions actually benefit intended beneficiaries. Pursue this possibility.** The delivery chain from resources to benefits for intended beneficiaries can often be a long one, with multiple points along the chain that can cause benefits to fail to materialize.

I. Introduction

A. Purpose

1.1. The purpose of this report is to summarize the evaluability status of the Norwegian Program for Capacity Development in Higher Education and Research for Development (NORHED). Against standards of the types and quality of information required for evaluating NORHED, this report assesses the results frameworks for the 45 NORHED projects, the 14 NORHED indicators common across the projects, and, within narrow limits, the processes that Norad's Section for Research, Innovation and Higher Education (FORSK) uses to supervise the projects.

1.2. Annex A presents the original terms of reference in which this activity was embedded. Annex B presents the amended proposal accepted by Norad's Evaluation Department for this particular exercise.

1.3. NORHED has two immediate objectives. These are to produce more and better research relevant to six identified areas/sub-programs¹ and to produce more and better qualified graduates, men and women, in these same areas/sub-programs. By strengthening capacity in higher education institutions in low- and middle-income countries (LMICs), NORHED's longer term objective is to sustainably contribute to a) a more and better qualified workforce, b) increased knowledge, c) evidence-based policy and decision-making, and d) enhanced gender equality.

B. Limitations

1.4. This report has a few methodological limitations.

- 1) This report was not based on an in-depth review of a representative sample of universities and projects. In March 2015 a member of the DPMG team conducted a 12 day "issues identification" mission in 3 out of 23 countries² (a 13% sample) and 6 out of 47 South universities (a 13 percent sample). The Norwegian universities partnered with the 6 South universities were not visited. The mission touched on issues such as the NORHED standard indicators assessed in this report, enabling conditions, and student views of new or revised NORHED-funded programs. In terms of the indicators, DPMG asked about: a) the process through which project report data are generated, probing for quality; b) the effort required to gather the data and whether data collection was imposing unnecessary burdens on the projects; and c) whether the standard indicators were creating any perverse incentives or unintended consequences for project managers. Although this quick scan identified issues about the indicators and program design issues, it by no means constituted the more in-depth field work required to pursue the several lines of enquiry that this quick scan surfaced.
- 2) DPMG was unable to pretest proposed modifications of the NORHED standard indicators.
- 3) At a minimum, assessing the supervision process requires desk reviews of the documentary trail for supervision of a sample of projects to determine the nature and quality of the information generated during supervision. Optimally, desk reviews should be complemented by field visits to a small sample of projects. However, budget limitations for this exercise precluded both desk reviews and field visits. This reality translates into providing an incomplete picture of the evaluability of NORHED. FORSK, which manages NORHED, will lack the feedback that it needs to affirm the quality of its supervision process and/or to modify particular aspects of the process to increase its validity and reliability.

¹ The six priority programme areas are: Education and Training; Health; Natural resource management, climate change, environment; Democratic and economic governance; Humanities, culture, media, and communication; and Capacity development in South Sudan.

² Two of these 23 countries had no institutions participating in the grant (Colombia and Ecuador).

C. Organization of the report

1.5. Chapter II presents a framework for judging the evaluability of NORHED. It consists of the data required at each stage of the project cycle in order to judge NORHED's performance. It is based on Norad's evaluation intent for NORHED and on the only evaluation model available, given the design of NORHED: a theory-based evaluation model. This model puts a premium on: a) a well-defined and well-measured theory of change for NORHED overall, and b) since the achievements of the individual grants define the success of the overall program, on well-defined and well-measured theories of change for the individual projects. The projects' results frameworks are treated as the operationalization of their theories of change.

1.6. Chapter III assesses the quality of the results frameworks data for the 45 NORHED projects, using different results management practices. Chapter IV is central to the issue of NORHED's evaluability. It assesses the quality of the NORHED standard indicators that structure the M&E of the projects and of the overall program. It proposes a number of ways to revise the indicators so as to improve the quality of the data that they will generate. Chapter V conducts an initial assessment of the availability and quality of implementation and supervision data. Chapter VI summarizes the evaluability of the NORHED as it now stands.

II. Framework for Assessing the Evaluability of NORHED

2.1. This chapter frames the basis for judging the evaluability of NORHED.

A. The purpose of the evaluation determines the criteria for judging evaluability

2.2. Judgments about the evaluability of a program depend on the purpose of the evaluation.

- Purpose affects the evaluation questions asked and thus the type and quality of data required to answer them.
- The data requirements for an evaluation constitute the criteria for judging whether the type and quality of data being generated by the program meet these requirements. The fit between data demand and data supply define the evaluability of a program.

2.3. The evaluation literature categorizes evaluation purposes variously, but distinctions between three types of evaluations are serviceable (Rossi, Lipsey, and Freeman, 2004).

2.4. Program improvement: to help design or revise the program to perform better. This type of evaluation focuses on issues such as tradeoffs between different implementation processes³ or the effects of particular social, political, physical, and/or economic environments on aspects of an intervention. Typical audiences for such evaluations are program planners, administrators, oversight boards, or funders with an interest in optimizing the program's effectiveness.

2.5. Accountability: to arrive at judgments about the performance of (often taxpayer-funded) programs. This type of evaluation assesses the program's efficiency, its effectiveness, its sustainability, and/or its relevance, with the specific questions differing depending on the stage in the program cycle. For example, it might judge if the program, as designed, is more or less likely to achieve its intended outcomes. It might assess the program during implementation to determine if the program is likely to achieve its intended outcomes or can be restructured so as to increase those chances. It can assess the results of the program at its conclusion and judge whether these can plausibly be attributed to the program or to unrelated events outside of the program.

2.6. Typical audiences for this type of evaluation are decision-makers with major roles in program oversight, such as the funding agency, governing board, legislative committees, political decision-makers, or upper management. Accountability evaluations may influence significant decisions about the continuation of the program, allocation of resources, restructuring, or legal action. They thus require information that is sufficiently credible relative to scientific standards to provide a confident basis for action and to withstand criticism aimed at discrediting the results.

2.7. Knowledge generation: to describe the nature and effects of an intervention as a contribution to knowledge. This type of evaluation focuses on issues such as determining the need for a program, piloting an intervention to see if the results warrant going to scale, or testing a theory via a program based on the theory. It is more apt to focus on why particular program events occurred, not just on whether they occurred. Typical audiences for this type of evaluation are the sponsors of the research as well as a broader audience of interested scholars and policymakers. The findings of the evaluation are most likely to be disseminated through scholarly journals, research monographs, conference papers, and other professional outlets.

2.8. Obviously, evaluations conducted primarily for one purpose shed light on questions that arise under other purposes. However, different purposes require somewhat different evaluation designs and different kinds of information.

³ For example, how do different procurement and contract management approaches to securing school construction services affect costs, delivery time, the ability to scale up, or the ability to reach rural and remote areas?

B. Learning is the primary purpose of evaluating NORHED, with an ancillary accountability purpose

2.9. As specified in the Terms of Reference for the framework agreement under which this study is conducted, learning is the primary evaluation purpose for NORHED.⁴ However, the TORs for call-offs under this framework agreement and the documents that govern NORHED imply an ancillary purpose of accountability for results. The TORs for call offs A-1 and A-2 have both been concerned about whether NORHED can be judged to have at least contributed to any observed results. Although this question is germane to learning, it is usually central to evaluations with an accountability purpose. Documents that govern Norwegian ministries and agencies, such as the 2003/2010 *Regulations on Financial Management Central Government*, and NORHED specifically (the 2011 *Grant Scheme Rules for Support for Capacity Building within Research and Higher Education* - Chapter/Item 165.70) either directly require or imply the need for accountability.

2.10. Project cycle questions

2.11. Both learning and accountability objectives put a premium on collecting data to answer these questions.

- Are the project and projects within the program designed to assess results?
- Are the intended results (outputs and outcomes) for the program and projects within the program occurring?
- If the intended results are not occurring, why not? This question is central to the learning objective, but as an "early warning" question, links to accountability. It is unnecessarily risky to wait until the end of a program to determine its effects, given the messiness of the real world in which programs such as NORHED are implemented. By monitoring the performance of projects and the overall program during implementation, the administrators of NORHED can identify or anticipate problems that, if unresolved, will jeopardize the program's and projects' chances of success.
- Have the intended results (outputs and outcomes) for the program and projects within the program occurred?
- If not, why not?

Attribution question

2.12. As noted, Norad is also concerned about attribution or, given the frequently multiple sources of funding for change, about contribution (Norad, 2008, p.11). Establishing attribution/contribution is fundamental to conclusions about aid effectiveness. The OECD DAC (n.d.) distinguishes between outcome and impact evaluations: "The proper analysis of impact requires a counterfactual of what those outcomes would have been in the absence of the intervention....There is an important distinction between monitoring outcomes, which is a description of the factual, and utilizing the counterfactual to attribute observed outcomes to the intervention" (p.1).

2.13. Randomized controlled trials (RCT) are the gold standard for counterfactual comparisons. Quasi-experimental impact evaluation designs can also create observationally similar comparison groups, but comparability along unobserved dimensions remains a challenge. NORHED was not designed as a RCT.

⁴ "The purpose of the evaluation programme under this framework agreement is to enable learning by relevant stakeholders within the field higher education and development, so that future investments can become more effective in building capacity in higher education institutions as a contribution to development. In particular the evaluation programme will increase knowledge about the extent to which NORHED builds/develops capacity in higher education institutions, how capacity development of higher education institutions can be conducted most effectively, and how higher education institutions affect development." [Section 3, p.4 of Annex 7, Terms of Reference for Call for framework agreement on evaluation of the Norwegian Programme for Capacity Development Higher Education (NORHED)]

In fact, project grantees were deliberately selected on a number of dimensions potentially related to the desired outcomes of the project.

2.14. Under call-off A1 and Phase I of call off A2, DPMG examined a number of options for creating a control group in a quasi-experimental impact evaluation research design. DPMG reluctantly recognized that no quasi-experiment with credibly comparable treatment and control groups could be constructed for the dissimilar 46 grants.

2.15. The more "adjacent" the control group is to the treatment group, the lower the risk of important, but unmeasured, initial differences between the two. However, the risks of contamination--"leakage" or "spillover" of treatment effects from the treatment group into the control group--significantly increase. Grantees represent sub-groups of their higher education institutions, leaving those not benefiting from a NORHED grant in the university as potential control groups. However, activities pursued by grantees—such as revisions in the department's curriculum that are intended to improve the learning of the department's students--can easily affect the thinking and activities of the control group. In fact, an implicit objective of NORHED is such "leakage" or "spillover".

2.16. Other control group options included: a) departments in the same field as the grantee (e.g., health), but at a different university, and b) applicants to NORHED that had not won grants at different or the same university as the grantee. The unmeasured or poorly measured variable problem loomed large in all options such as these—completely aside from whether such potential control groups could be expected to cooperate with any data collection. Control groups at other universities (whether they had unsuccessfully applied for a NORHED grant or not) are subject to different enabling/disabling institutional conditions that can affect the outcomes of interests. Losing applicants at the same university as the winning applicant are subject to the same institutional enabling/disabling conditions as the winning applicant. However, the non-random allocation of grants among applicants means that, as judged by NORHED's Independent Review Committees, losing applicants differed from winning ones in important ways.

2.17. All of these options had to be discarded as being highly vulnerable to: a) initial differences between the treatment and control option that were either unmeasured or only imperfectly measured and that could bias the outcomes of interest; or b) contamination between the treatment and control groups.

2.18. This reality means that evaluating NORHED for performance and/or impact purposes has to be based on a theory-based evaluation design (White and Phillips, 2012). In other words, the burden of assessing attribution or contribution has to be carried by: a) a clear and plausible theory of change for the program; and b) a careful monitoring of whether and how the links in the causal path of this theory of change play out. This approach to attribution assumes that if the causal path is plausible and the links in the chain expected to produce certain results can be demonstrated to have materialized, the project/program can tentatively be concluded to have caused the observed changes. The greater the fidelity of how the program/project unfolds to the theory of change, the greater the confidence that results by the program's end can be attributed to the program.

2.19. Although not as rigorous as approaches such as randomized controlled trials, this evidence provides some reassurance about the source of results. White (2009) identifies principles that, if followed, increase the rigor of these types of evaluation designs. For example, understanding the social, political and economic context in which the program takes place can identify possible causes of the results observed that have nothing to do with the project or program.⁵

2.20. This approach puts a premium on the theory of change, but NORHED in fact consists of the overall program's theory of change and 45 project-specific theories of change.⁶ All NORHED projects are expected

⁵ The system-level and the institution-level enabling conditions identified in the theory of change for the first call-off for the NORHED evaluation represent contextual factors expected to affect the development of capacities at the tertiary level.

⁶ All analyses in this report for the universe of NORHED grants refer to 45 grants only. FORSK awarded 46 grants. However, one of these grants was put on hold because of a corruption case pending against an official of the

to meet the same development objectives: developing capacities that produce more and better research relevant to the six identified areas/sub-programs listed earlier and to produce more and better qualified graduates, men and women, in these same areas/sub-programs.⁷

2.21. However, the *means* by which each project reaches these shared objectives vary by project. As the *NORHED Programme Presentation* states: "Based on a needs assessment in the relevant country...the range of interventions potentially eligible for NORHED support is broad, with some degree of flexibility as long as project outputs meaningfully contribute to defined higher level outcomes and longer-term impacts."⁸ Thus, although all projects have the same development objectives--albeit different targets, each project has a different theory of change that is shaped by the sector for the grant--e.g., health or natural resource management--and the means that the grantee selects to build capacities within the partnership.

2.22. These realities put a premium on: a) a well-defined and well-measured theory of change for NORHED overall, and b) since the achievements of the individual grants define the success of the overall program, on well-defined and well-measured theories of change for the individual projects.

What a theory-based evaluation implies for the nature of the results framework

2.23. Theories of change in the development world are normally operationalized in results frameworks. In a legitimate desire to focus grant schemes on outputs, outcomes, and impacts, the 2011 *Grant Scheme Rules for Support for Capacity Building within Research and Higher Education* has a somewhat restricted view of a results framework (Section 3, p.2, Criteria for achievement of objectives).

2.24. However, the results framework section (section 3) of the application form for those applying for NORHED grants required an expanded form of a results framework. It asks applicants to specify activities and assumptions/risks to their achievement; outputs, output indicators, and assumptions/risks to their achievement; outcomes, outcome indicators, and assumptions/risks to their achievement; the development goal/intended impact on society; and baselines for intended outputs and outcomes.

2.25. The 2008 *Results Management in Norwegian Development Cooperation – A Practical Guide* works with the full results chain, as does the excellent presentation by Norad's Department for Quality Assurance at the March 2014 Addis workshop for NORHED grantees on Managing for Results and Risk Management.

2.26. Figure 1 displays a results chain that is more expansive than that specified in the 2011 *Grant Scheme Rules for Support for Capacity Building within Research and Higher Education*. Although we are entirely sympathetic to Norad's attempt to stress outputs, outcomes, and impact, the full version of the results framework is required by the only evaluation model available for NORHED: a theory-based evaluation model. This model requires attention to the entire results chain in order to draw even remotely credible conclusions about attribution/contribution of the aid program/project. In other words, clarity about the entire results chain is crucial for evaluability purposes. As the program/project unfolds, it is also important to learn what is working and what is not working and why. If outputs occur as expected, it can be inferred that inputs and activities probably unfolded as expected. However, if outputs are not occurring as expected, the focus has to shift upstream in the results chain to inputs and activities to determine

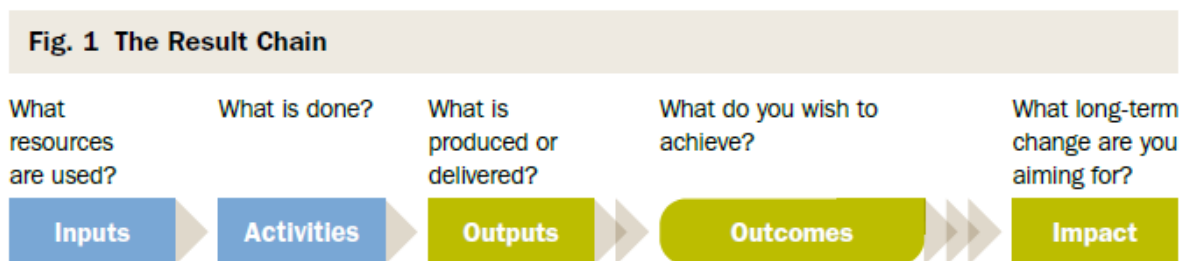
university. That case has now been resolved, and FORSK hopes to finalize the contract with this grantee in November, 2015.

⁷ When projects meet these objectives, they are expected to support NORHED's longer term objectives of sustainably contributing to a) a more and better qualified workforce, b) increased knowledge, c) evidence-based policy and decision-making and d) enhanced gender equality.

⁸ The *Presentation* document lists "typical" interventions that projects might pursue in-country/regional Masters (and bachelor) education programmes, Ph.D. studies and post-doctoral fellowships, joint research projects in line with overall NORHED programme aims and sub-programme areas, institution and systems strengthening, systems for knowledge management, information and dissemination of results, and scientific equipment and small scale infrastructure.

bottlenecks--for example, procurement delays or inadequate budget that make it impossible to start key activities.

Figure 1: Norad's Results Chain



Source: Norad and Norwegian Ministry of Foreign Affairs, 2008. *Results Management in Norwegian Development Cooperation: A practical guide*, p.10.

Data requirements to evaluate NORHED

2.27. Thus, the question about the evaluability of NORHED becomes a question about how well the program is generating the data of the type and quality required to monitor the status of each project at each stage of the program cycle relative to its intended objectives. Table 1 shows the data required and the data available. It shades cells where, to our knowledge, a) data are not available, or b) in the case of project completion, the data that Norad plans to analyze or to collect are unknown.

2.28. The subsequent chapters evaluate the quality of the data or discuss missing data. Chapter III evaluates grantees' results frameworks, using sources of information additional to the results frameworks themselves. These frameworks should provide the projects' theories of change and baselines that are critical to being able to evaluate NORHED. Chapter IV evaluates the quality of the NORHED standard indicators that structure the M&E of the projects and of the overall program. It proposes a number of ways to improve the quality of the data that they are likely to generate. Chapter V evaluates NORHED's supervision processes, although within the very narrow limits allowed by the budget available for this report.

Table 1: Data required for evaluating NORHED and NORHED data sources by stage of the project cycle

Data required	Data source
Design	
1. Theory of change for NORHED	Explicated in first DPMG evaluation of NORHED ⁹
2. Theory of change for each grant described in sufficient detail that the project's causal path can be identified and its clarity and plausibility assessed	Grantees' applications (main application and implementation plans).
3. Data on contextual variables (system and institution-level) that might enable or impede grantees' attainment of program objectives. ¹⁰	Although grantees might identify some of these conditions as risks in their applications for NORHED funding, systematic data on contextual variables are not available.
4. Specification of risks to achieving project's objectives	Grantees' main applications for NORHED funding
5. Measurable performance indicators common to all projects ("standard" indicators)	NORHED defines "standard" performance indicators common across projects
6. Project-specific performance indicators	Grantees' results frameworks and implementation plans were expected to include both standard and project-specific indicators, with implementation plans specifying activities
7. Baseline data on all standard indicators	Grantees' revised results frameworks were expected to include baselines for all standard indicators. In some cases grantees were allowed to submit baseline data with their first annual reports, submitted at least a year after the projects started.
8. Baseline data for the project-specific indicators	Grantees' revised results frameworks were expected to include baselines for all project-specific indicators. In some cases grantees were allowed to submit baseline data with their first annual reports, submitted at least a year after projects started.
Implementation	
1. Updated disbursement data	Norad's computerized management system called Plan, Tilskudd, Avtale (PTA)
2. Updated data on standard indicators	Grantees' annual narrative reports
3. Status updates on project's causal path or theory of change, including updated data on project-specific indicators	Grantees' annual narrative reports
4. Updated information on system and institution level conditions that might enable or impede results achievement	Not available. Annual narrative report template does not ask for data on enabling conditions.
5. Updated information on risks	Grantees' annual narrative reports
6. Updated information on sustainability of NORHED-funded capacity developments	Not available

⁹ Evaluation Series of NORHED: Theory of Change and Evaluation Methods:

<http://norad.no/en/toolspublications/publications/2014/evaluation-series-of-norhed-theory-of-change-and-evaluation-methods/>

¹⁰ Ibid. See figure 1, which identifies the generic enabling conditions for reaching NORHED's objectives. As explicated by DPMG, the NORHED theory of change implicitly included some of these.

Data required	Data source
7. Assessment of implementation status: issues, solutions, successes	Grantees' annual narrative reports
8. Independent verification of performance data	<p>The status of this data requirement is not now known.</p> <p>Article X (Evaluation and Review) in the generic legal agreement with NORHED grantees specifies these requirements:</p> <p>"A final and /or mid-term review of the Project shall take place during the Support Period at a time agreed between the Parties. The review/s shall ascertain to what extent the Project has delivered the results set out in the result framework. The review shall also assess whether any adjustment of the result framework or organizational set up of the Project is needed. Norad shall draft the terms of reference for the review which shall be submitted to the Grant Recipient for information."</p> <p>The question is whether the TORs for these independent final and/or mid-term reviews will require the evaluators to collect their own data, verify the accuracy of the data provided by grantees, rely on the data provided by grantees, or use some combination of these approaches. If evaluators collect their own data or verify the accuracy of the data provided by grantees, these assessments will check for potential positive bias in the data that grantees provide. However, for reasons of time, budget, or feasibility, the TORs may have to ask evaluators to rely on the data generated by the projects. In this case the potential for positive bias remains unchecked and the need for independent checks remains.</p>
9. Independent tracking of whether inputs are benefitting intended beneficiaries	Not available. FORSK is considering adding such studies
10. Data on performance of overall NORHED portfolio, based on summaries of project-specific performances	Not available.
Project Results	
<p>1. Final status of each grant on each of the NORHED standard indicators relative to baselines and targets</p> <p>2. Analysis of results for individual grants to arrive at results conclusion for NORHED</p> <p>3. Data from design and implementation stages to interpret findings on results</p> <p>4. Special studies to assess accuracy of reported results, effects of projects for intended beneficiaries, factors that explain success or failure.</p>	See discussion of the data sources for item 8 under "implementation", above. It is not known what data sources Norad wishes to use to judge the results for each project and thus for NORHED overall. Evidence from the design stage (grantees' applications, revised results frameworks, and baselines from the results frameworks or the first annual reports) and from the sequence of grantees' annual reports is available. However, Norad may conduct special audits, case studies, surveys, or other types of studies to augment these basic sources of data.

III. Results for Re-analysis of Adequacy of Grantees' Results Frameworks

A. Prior Analysis

3.1. In its Inception Report, DPMG assessed the adequacy of grantees' results frameworks as a way to assess the clarity and plausibility of the theory of change or logic chain underlying the grants. Clarity and plausibility of such theory of change is crucial for the evaluability of NORHED given that the absence of credible counterfactual scenarios and the need rely on factual analysis to potentially establish attribution. The versions evaluated were not the grantees' original results frameworks, but ones that they had revised after a workshop conducted by Norad to help them improve the results frameworks.

3.2. DPMG assumed that a theory-based evaluation model—the only option for assessing attribution in NORHED projects—required the specification of the path from funds to interventions (activities) to outputs to outcomes. Accordingly, DPMG created an evaluation template that covered this path. DPMG's evaluation template consisted of 14 criteria that DPMG considered essential in a results framework.¹¹ After validating the assessment tool,¹² 45 results frameworks were coded against the evaluation template.¹³

3.3. This original evaluation template used a binary coding that measured the overall quality of a project's results framework (RF) from an evaluability perspective. If a criterion was adequately met, it was rated as "Yes;" if it was not met, it was rated as "No." The "yes" responses were then coded as 1 and the "no" responses as zero, as is common practice in statistical models of limited qualitative (or dichotomous) variables. Therefore, the maximum possible score for a project's RF was 14. The minimum was zero. Table 2 below shows the evaluation criteria and the percent of projects that met each criterion.

Table 2: Evaluation Template (Yes/No) with Percentage of Projects Complying

Criteria	% of Projects scored Yes
1. Does the Results Framework (RF) clearly state the problem or challenge that the project is trying to solve?	84%
2. Are the stated assumptions required to reach the outcomes reasonably within the span of control of the project?	36%
3. Do project activities realistically address project risks?	0%
4. Are the outcomes that the project wants to achieve clearly stated and linked to the problem?	31%
5. Does the RF clearly identify the project's critical activities?	38%
6. Are the projects' outputs measurable?	31%
7. Is the overall structure of the results framework clear and coherent?	27%
8. Are the causal connections between activities and outputs traceable and plausible?	16%
9. Are the causal connections between outputs and outcomes traceable and plausible?	42%
10. Does the RF include activity indicators to measure progress?	0%

¹¹ DPMG created the evaluation template based on the 2013 guidance note used to internally evaluate results frameworks of World Bank projects, a review of the current literature on results frameworks, and section 3 of the template that NORHED used to solicit applications to the program. The World Bank quality standard is used by other multinational development banks, by policy research groups such as the MIT Jameel Poverty Action Laboratory, and by governments in countries such as Colombia and Chile.

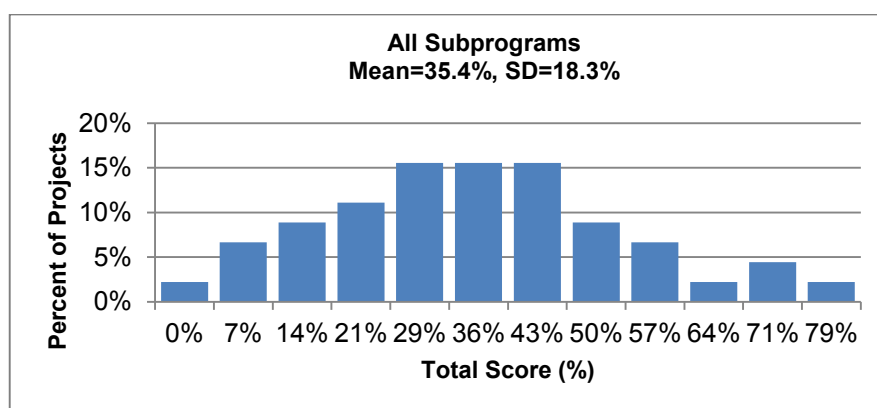
¹² DPMG applied common practice in psychometric testing to establish the validity and internal consistency of its evaluation template and to resolve ambiguities. Three independent evaluators from the DPMG team separately used the instrument to code a sample of five updated RFs. The three DPMG evaluators then compared answers and clarified questions. Inter-rater reliability was very high relative to established psychometric standards. Specifically, the evaluators consistently responded to a given item in the same way for over 90 percent of questionnaire items

¹³ One results framework was still missing at the time this task was conducted.

Criteria	% of Projects scored Yes
11. Does the RF include output indicators with numerical baseline values (if applicable)?	31%
12. Does the RF include outcome indicators with numerical baseline values (if applicable)?	27%
13. Does the RF include <u>final-year</u> targets for all output and outcome indicators?	60%
14. Does the results framework include any arrangements for measuring the project's output and outcome indicators once the project starts to implement?	73%

3.4. To summarize our analysis of the overall quality of results frameworks, we constructed a relative index based on the total score by converting scores into percentages. In this way, a relative score of 100% represents an RF with 14 out of 14 essential criteria. Figure 2 shows that the average project's RF score was of fairly low quality: 35.4%. It met only 5 of the 14 criteria. No RF met all of the essential criteria of an RF. Figure 2 shows that one project (or 2.5% of projects) received a score of zero (i.e., it met none of the criteria.) The highest score received was 79% (i.e., met 11 of 14 essential criteria) and was scored by only one project (2.5%).

Figure 2: Overall Quality of RFs – Original Evaluation Framework



B. Current Analysis

Norway's Results Management Practice for Results Frameworks

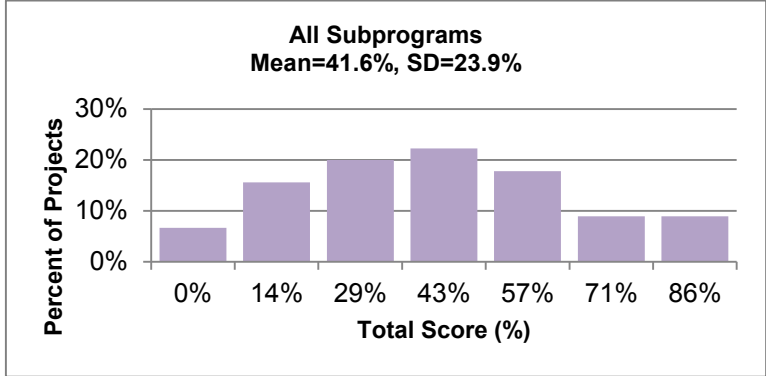
3.5. Criteria for results frameworks specified by Norad. Aid agencies differ in how they define a results framework. In its response to the DPMG inception report, Norad felt that the original evaluation template had not reflected its results management practices and that the grantees were scored against criteria that Norad had not required of them. Items 1, 2, 3, 4, 5, 8, and 10 were identified as not corresponding to Norad's requirements for applicants' results framework matrixes. To address Norad's concerns, a "**reduced evaluation template**" of the seven Norad criteria was created. See Table 3.

Table 3: Seven Norad Requirements with Percentage of Projects Complying

Criteria	% of Projects scored Yes
6. Are the projects' outputs measurable?	31%
7. Is the overall structure of the results framework clear and coherent?	27%
9. Are the causal connections between outputs and outcomes traceable and plausible?	42%
11. Does the RF include <i>output</i> indicators with <u>numerical</u> baseline values (if applicable)?	31%
12. Does the RF include outcome indicators with numerical baseline values (if applicable)?	27%
13. Does the RF include <u>final-year</u> targets for all output and outcome indicators?	60%
14. Does the results framework include any arrangements for measuring the project's output and outcome indicators once the project starts to implement?	73%

3.6. Figure 3 indicates that even after reducing the number of essential criteria, the average project RF remains of fairly low quality: 41.6%. The average project RF met less than half of the criteria considered--about 3 of the 7 Norad criteria. Figure 3 below summarizes the distribution of the revised RF scores for the 45 funded projects for which we had RFs.

Figure 3: Overall Quality of RFs – Reduced Evaluation Template



3.7. When assessed against Norad’s seven criteria, 9% of projects (4 projects) obtained a high score of 86% (6 of 7 essential criteria), compared to only one project getting the highest score (79%) when the RFs were assessed against the original template. Assessed against the original evaluation template, only one project scored a zero; however, 7% of projects (3 projects) obtained a score of zero when assessed against the reduced template.

3.8. **Criteria for results framework specified by NORHED's application template.** Although DPMG responded to the request to assess grantees' RFs against reduced criteria, **the results matrix in the NORHED application form (section 3) requested information on almost all of DPMG's original criteria.** The Inception Report had evaluated grantees' *revised* results frameworks, and the revised frameworks did not use ONE structure. **However, the majority of grantees followed the structure of section 3 in the application form.**

3.9. Table 4 below shows where the evaluation template’s criteria that Norad identified as not required actually were requested in NORHED’s application form RF matrix.

Table 4: Rejected Evaluation Criteria and Corresponding Requirements of NORHED’s RF

Evaluation Template Criteria	Application Results Framework Section
1. Does the Results Framework (RF) clearly state the problem or challenge that the project is trying to solve?	3.1 Development goal/ intended impact on society
2. Are the stated assumptions required to reach the outcomes reasonably within the span of control of the project?	3.4. Assumptions associated with delivering planned activities 3.3. Assumptions associated with reaching the outputs 3.2 Assumptions associated with reaching the outcomes.
3. Do project activities realistically address project risks?	3.4. Risks associated with delivering planned activities 3.3. Risks associated with reaching the outputs 3.2. Risks associated with reaching the outcomes.
4. Are the outcomes that the project wants to achieve clearly stated and <u>linked</u> to the problem?	3.2 Project Goal (purpose/intended outcomes)
5. Does the RF clearly identify the project's critical activities?	3.4 Planned Activities
8. Are the causal connections between activities and outputs traceable and plausible?	3.4 Planned Activities 3.3 Expected results/services/products (outputs)
10. Does the RF include activity indicators to measure progress?	NA

Other Results Management Tools

3.10. Although many grantees provided baseline data for outputs and outcomes in their revised RFs, Norad allowed grantees whose RFs lacked baselines to provide those in the first annual progress report.¹⁴ Norad also felt that the applicants' implementation plans should be reviewed for data on intended activities, although in fact section 3 of the main application form clearly asks for activities. Accordingly, we supplemented our analysis by reviewing the annual report for numerical baseline values and implementation plans for activities for all 45 projects.

3.11. Annual reports and baseline values for indicators. DPMG reviewed the annual progress reports to review the quality of the data on baselines in order to reflect Norad's decision about allowing baselines to be reported in the first annual report.¹⁵ The new analysis affects criteria 11 (baselines for output indicators) and 12 (baselines for outcome indicators) in the evaluation template.

3.12. DPMG updated the scores for the projects that did not report baseline values for output and outcome indicators in their RFs, but did report them in their annual reports. This did not penalize grantees who reported baseline values in their results framework, but did not in the annual report.

3.13. We reviewed the annual reports for their reporting of NORHED's standard indicators (see Chapter IV for a more detailed discussion of NORHED's standard indicators). Of the 14 standard indicators, eight have numerical values (see Table 5). We coded those as either output or outcome indicators. We then assessed whether baseline values were reported in the annual report by the projects that had not included baseline values for output and outcome indicators in the RF. To receive a Yes score on criteria 11 and 12, **all** output and outcome indicators, respectively, had to have baseline values.

Table 5: NORHED Standard Indicators with Numerical Values

Indicator	Type	
1	Number of new and number of revised bachelor/master/PhD programs/modules supported by NORHED	Output
2	Number of bachelor/master/PhD programs /modules supported by NORHED with a gender perspective included	Output
3	Capacity to enroll and graduate students in NORHED-supported programmes (bachelor/master/PhD)	Output
5	Number of academic staff with strengthened qualifications (master/PhD) by relevant institutional level and gender	Output
6	Ratio of qualified academic staff (master/PhD) to students by relevant unit supported by NORHED	Output ¹⁶
7	Retention rates of qualified academic staff at relevant unit supported by NORHED	Output
8	Number of scientific publications	Outcome ¹⁷
9	Number and type of other dissemination activities by type (media, policy briefs, outreach) and gender	Outcome

3.14. Twenty projects improved their scores for baseline data once the annual reports had been reviewed. Of those 20, six annual reports had not yet been approved by FORSK. Table 6 shows: a) the original score; b) the updated score including the revised score for only the 14 projects that included baseline values for indicators in their **approved** annual reports; and c) the updated score including the revised score for all 20 projects that included baseline values for indicators in their annual reports regardless of the annual report's approval status.

¹⁴ Constructing good results frameworks is challenging, and, as a group, grantees had limited experience with them. FORSK and Norad's Results Management section worked hard to help grantees improve their RFs.

¹⁵ As late as this summer, three grantees had not yet submitted their annual reports, and FORSK had not yet formally approved 10 of those that had been submitted.

¹⁶ Because standard indicator #6 can serve only as a weak proxy for quality, we coded this as an output indicator. Strictly speaking it is an output as it is not a direct measure of the quality of graduates

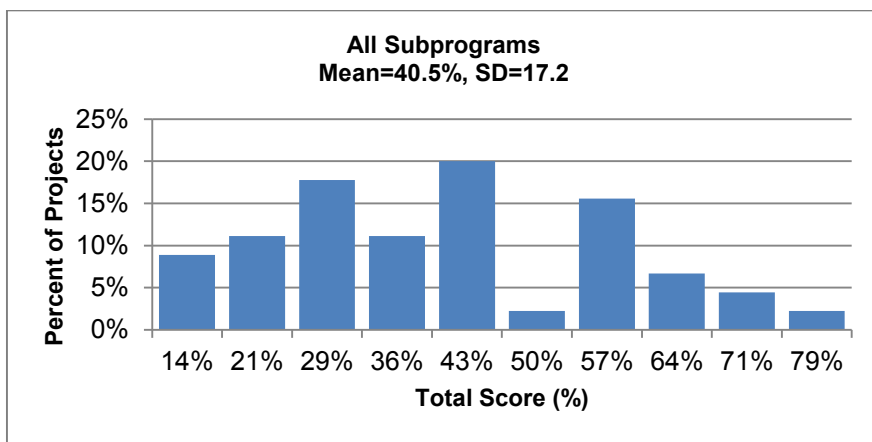
¹⁷ This is unlikely to be observed in the time period

Table 6: Original and Revised Mean for Criteria 11 and 12 in Evaluation Template

Criteria	Original	Revised (excl. unapproved)	Revised (incl. all annual reports)
11. Does the RF include <i>output</i> indicators with numerical baseline values (if applicable)?	0.31	.56	.67
12. Does the RF include outcome indicators with numerical baseline values (if applicable)?	0.27	.53	.62

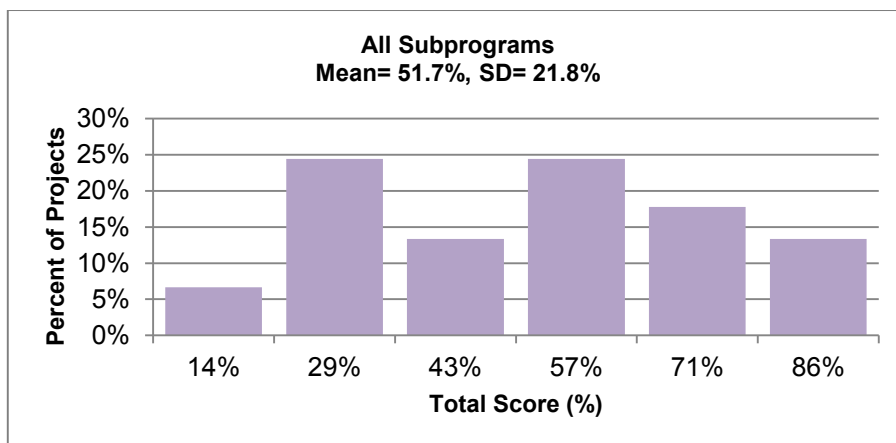
3.15. When including the revised scores for criteria 11 and 12 based on all annual reports (not just approved), the average score in the original evaluation template of 14 criteria went up to 40.5% (i.e., the average project met just under 6 of the 14 essential criteria) (see Figure 4). No project scored a zero. The lowest score obtained was 14% (meeting 2 of the 14 essential criteria) by four projects.

Figure 4: Overall Quality of RFs – Original Evaluation Framework with Updated #11 and #12



3.16. When including the revised scores for criteria 11 and 12 based on all annual reports (not only those that have been approved), the average score in the reduced evaluation template (per Norad’s RM practices) went up to 51.7% (see Figure 5). In other words, the average project met 3.7 of the 7 Norad criteria. The highest scored obtained was 86% (meeting 6 of the 7 Norad criteria) by six projects.

Figure 5: Overall Quality of RFs – Reduced Template with Updated #11 and #12



3.17. Implementation plans and activities. Almost 40 percent of the grantees included activities in their results frameworks (see Table 7 below). However, Norad's Result Management section felt that grantees' implementation plans specified activities more completely. DPMG reviewed all implementation plans and revised the scores for items 5 and 8 of the evaluation template (see Table 7 below).

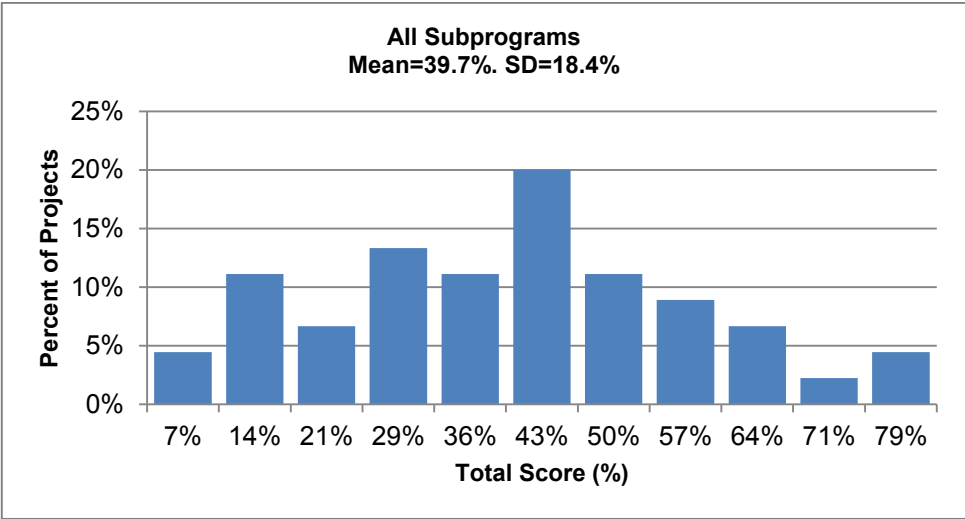
Table 7: Original and Revised Mean for Criteria 5 and 8 in Evaluation Template

Criteria	Original	Revised
5. Does the RF clearly identify the project's critical activities?	.38	.80
8. Are the causal connections between activities and outputs traceable and plausible?	.16	.33

3.18. In the case of item 5, 17 additional projects were assessed as meeting the criteria. In the case of item 8, an additional eight projects met the criteria.

3.19. The changes in the scores did not affect the score distribution for the reduced evaluation template because that template omitted criteria 5 and 8. However, they do affect the original evaluation template score distribution. Although the scores for items 5 and 8 individually improved drastically, this rescoring based on the implementation plans resulted in only a slight improvement in the average overall score. The average score in the original evaluation template was 35.4%; when accounting for activities being listed in the implementation plans, the average score increases to 39.7% (see Figure 6). The lowest score obtained was 7% by two projects; the highest score was 79% obtained by two projects as well.

Figure 6: Overall Quality of RFs – Original Evaluation Framework with Revised Items #5 and #8



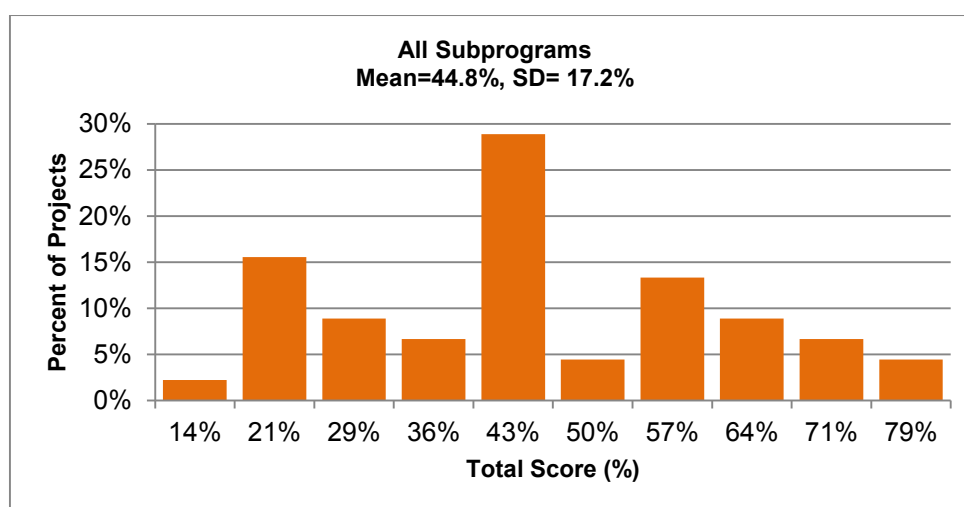
3.20. Annual reports and implementation plans. When taking into consideration the annual reports and the implementation plans, the scores for items 5, 8, 11 and 12 improve (see Table 8).

Table 8: Evaluation Template (Yes/No) with Percentage of Projects Complying for Original and Revised Scoring

Criteria	% of Projects scored Yes	Revised scoring: % of projects scored Yes
1. Does the Results Framework (RF) clearly state the problem or challenge that the project is trying to solve?	84%	84%
2. Are the stated assumptions required to reach the outcomes reasonably within the span of control of the project?	36%	36%
3. Do project activities realistically address project risks?	0%	0%
4. Are the outcomes that the project wants to achieve clearly stated and linked to the problem?	31%	31%
5. Does the RF clearly identify the project's critical activities?	38%	80%
6. Are the projects' outputs measurable?	31%	31%
7. Is the overall structure of the results framework clear and coherent?	27%	27%
8. Are the causal connections between activities and outputs traceable and plausible?	16%	33%
9. Are the causal connections between outputs and outcomes traceable and plausible?	42%	42%
10. Does the RF include activity indicators to measure progress?	0%	0%
11. Does the RF include output indicators with numerical baseline values (if applicable)?	31%	67%
12. Does the RF include outcome indicators with numerical baseline values (if applicable)?	27%	62%
13. Does the RF include <u>final-year</u> targets for all output and outcome indicators?	60%	60%
14. Does the results framework include any arrangements for measuring the project's output and outcome indicators once the project starts to implement?	73%	73%

3.21. When combining the information found in the results framework, the annual report, and the implementation plan in order to re-assess the projects against the **original evaluation template**, the average score improves from 35.4% to 44.8% (see Figure 7 below). The average RF meets 6.3 of the 14 essential criteria.

Figure 7: Overall Quality of RFs – Original Evaluation Template with Revised Items #5 #8, #11, #12¹⁸



¹⁸ For items 11 and 12, the updated score includes the revised score for all 20 projects that included baseline values for indicators in their annual reports *regardless of the annual report's approval status*.

3.22. The effect of reassessing items 5, 8, 11 and 12 of the original evaluation template is limited. Even when supplementing the RF with the baseline data in the annual report and the activities in the implementation plan, the average RF remains of low quality. The effect is more substantial when the more limited criteria of the reduced template are used (see Table 9). However, because some of criteria that were indicated by Norad as not corresponding to their requirements are actually found in the RF application form and in the revised RFs, the reduced template is not effective in evaluating the quality of RFs. It excludes criteria that should have been met in completing the results framework matrix.

Table 9: Mean and Standard Deviation of Relative Score

	Original Evaluation Template	Original Evaluation Template (with Revised Scores, incl. data from annual reports and implementation plans)	Reduced Evaluation Template (with items 11 and 12 updated)
Average Score	35.4%	44.8%	51.7%
Standard Deviation	18.3%	17.2%	21.8%

C. Conclusions

3.23. The analysis of the projects' results frameworks yielded some good news and some less good news with regards to their usefulness for evaluability purposes.

3.24. Unclear criteria for judging project results frameworks. The criteria against which grantees' results frameworks should be judged are not clear. The results matrix (section 3) in NORHED's main application form asks for a specification of activities, outputs, and outcomes, the assumptions and risks associated with each, and baselines for the "present situation as a basis for assessing the project's outcomes". This template works with the entire results chain except for inputs/resources, such as financing, staff, equipment, and facilities. When grantees revised their results frameworks early in their contracts, they used various templates but primarily the template from the application form.

3.25. On the other hand, Norad indicated that they held grantees accountable for only outputs and outcomes and that their frameworks should not be judged against criteria associated with the fuller results chain. Thus, for example, project results frameworks should not be judged on whether the causal connections between activities and outputs were traceable and plausible.

3.26. Good news from the evaluation of grantees' results frameworks. As table 8 showed, more than 8 out of every 10 grantees had a clear statement of the problem or challenge that the project was trying to solve. This result is to the credit of the grantees, FORSK, and Norad's Results Management Section. Clarity about the projects' objectives is critical to evaluating the projects, and constructing such statements is not easy. Once the implementation plans were reviewed, 8 out of every 10 grantees had clearly identified activities or interventions. Seven out of every 10 grantees had at least some arrangements for measuring the output and outcome indicators.

3.27. Less good news from the evaluation of grantees' results frameworks. Table 8 shows a pattern of weak links between the links in the results chain. Outcomes are clearly stated and linked to the problem in only 3 out of every 10 projects. In only 4 out of every 10 projects are outputs clearly linked to outcomes. Although reviewing the implementation plans improved the rating for the links between activities and outputs, again only about 3 out of every 10 projects showed acceptable linkages.

3.28. Reviewing the first year's annual reports for baselines for outputs and outcomes markedly improved the ratings of projects on these dimensions. However, despite significant efforts by FORSK and the Results Management Section to obtain baseline data so critical to the evaluation of the projects, a third of the projects still lacked complete baselines for outputs. Two out of every 5 projects lack complete baselines for outcomes. Only 2 out of every 5 projects had final year targets for its outputs and outcomes.

D. Recommendations

3.29. To improve evaluability, Norad should require results frameworks that reflect the full results chain. Norad has properly been trying to shift the focus from inputs, activities, and process to outputs and outcomes, as reflected in the *Grant Scheme Rules for Support for Capacity Building within Research and Higher Education* (2011). As the NORHED guidelines for applicants on their results frameworks states, "A results-based approach involves shifting management attention away from a focus on inputs, activities and processes to a focus on benefits – From what you have done to what you have achieved."(Norad, 2013, p.5)

3.30. The reality is that many Norad aid programs will be able to sustain only a theory-based evaluation model. This model requires attention to the entire results chain in order to draw even remotely credible conclusions about attribution/contribution of the aid program/project. It assumes that if the causal path is plausible and the links in the chain expected to produce certain results can be demonstrated to have materialized, the project/program can tentatively be concluded to have caused the observed changes. The greater the fidelity of how the program/project unfolds to the theory of change, the greater the confidence that results by the program's end can be attributed to the program.

3.31. A program should use a consistent result framework. It is important that a program be able to identify a consistent set of criteria against which results frameworks should be judged.

IV. Analysis of Quality of the Standard Indicators

A. Prior Analysis

4.1. The Inception Report sought to address two main questions with regards to the indicators to measure the effectiveness or NORHED: a) Are the standard proposed indicators developed by NORHED appropriate for determining the effectiveness of the program? and b) Is the operationalization of indicators sound?

4.2. Our analyses yielded two main conclusions:

- **The proposed set of standard NORHED indicators, while overall relevant, is not entirely appropriate for determining the effectiveness of the program.** As a whole, the rationale for the proposed set of standard indicators captures the most relevant dimensions of NORHED effectiveness. Involving grantees in their ultimate selection probably contributed to their relevance. The Addis workshop likely also prompted discussions about how to interpret various indicators. These discussions may have alerted FORSK early to potential misinterpretations that could arise down the road. They may have also helped to build a shared understanding of the meaning of each indicator, thus increasing the likelihood that the indicators would yield comparable data.
- **Despite the overall relevance and success of the stakeholder participation in selected the final standard indicators, grantees expressed concern about ambiguity in the operationalization of the standard indicators.** During a brief mission Dr. Jamil Salmi of DPMG queried project coordinators, students affected by the NORHED project, and university administrators at six universities on several aspects of the indicators. For projects that had submitted their annual reports, the responses of the project coordinators were not particularly patterned, but suggested that it had been challenging to prepare the first annual report because the indicators were abstract and not defined in concrete terms. DPMG suspects that converting as many indicators as possible into quantitative indicators, with clear instructions on what to count and what not to count, may help to concretize the indicators.
- **The proposed set of indicators misses some important dimensions of program effectiveness or require data not easily provided by grantees.** Dr. Salmi's mission flagged that the NORHED standard indicators did not measure some dimensions germane to the program. These were: student perceptions of the quality (or change in quality) of instructional practice; the quality of research; and system-level and institution-level processes known to enable or impede capacity development at the university level, especially the processes and standards for retaining and promoting faculty and staff. Item 4.4 in the annual report on labor market results requires measures of the relevance of NORHED-related educational programs and new graduates to labor markets. However, Dr. Salmi found that none of the universities/departments visited had a system in place to follow up accurately the labor market results of their graduates.
- **Few of the standard NORHED indicators are operational.** The NORHED indicators were assessed against the "CREAM" criteria for good performance indicators (Kusek and Rist 2004; Schiavo-Campo 1999), where CREAM is defined as Clear, Relevant, Economic, Adequate and Monitorable.¹⁹ Other criteria, such as those known as SMART, could have used. CREAM and SMART are very similar. However, the CREAM criteria have been found to be somewhat more flexible.

4.3. We assessed the operationalization of NORHED's definition of each standard indicator. Some indicators implied the possible utility of more than one indicator. In practice, this implied the evaluation of 22 indicators instead of 14. The assessment indicated that only seven of the 22 standard NORHED indicators²⁰ were operational, as determined by whether they met all of the five CREAM criteria.²¹ Two

¹⁹ The SMART acronym stands for indicators that are: Specific, Measurable, Achievable, Realistic and Time-Related. Doran's SMART approach to measuring progress towards goals and objectives (Doran 1981) is similar to CREAM.

²⁰ There are 14 standard indicators, but some of these have sub-indicators that bring the total to 22.

²¹ According to the "CREAM" criteria, performance indicators should be: Clear, Relevant, Economic, Adequate and Monitorable (Kusek and Rist 2004; Schiavo-Campo 1999).

of these seven referred to student enrollment and graduation (“number of enrolled students at program, by gender” and “number of graduated students at program, by gender”). One indicator referred to faculty credentials: “number of staff with strengthened (master/PhD) academic qualifications, by level and gender.” One indicator referred to staff retention, defined as the “fraction of NORHED sponsored staff that stayed in their job at least one year after their year(s) of duty agreement was terminated.”

4.4. The other three indicators meeting the CREAM criteria referred to the production of scientific knowledge (“number of peer-reviewed publications, total and with female coauthors”), “number of books, reports or other scientific publications, total and with female coauthors”, and “number and type of other dissemination activities by type.”

B. Relevance of NORHED's Standard Indicators Revisited

4.5. DPMG continues to find NORHED's standard indicators substantively relevant, given NORHED's purposes. This achievement might be credited to the decision by FORSK and Norad's Results Management section to hold a workshop for grantees in Addis Ababa in March 2014 to discuss results management, results frameworks, and the final selection of the NORHED standard indicators. This was good aid management practice. Involving grantees in the ultimate selection of the shared indicators almost certainly contributed to their relevance. The workshop also prompted discussions about how to interpret various indicators. These discussions probably alerted FORSK to potential misinterpretations and helped to build a shared understanding of the meaning of each indicator, thus increasing the likelihood that the indicators would yield comparable data.

4.6. However, the standard indicators are not complete because they do not measure some domains germane to NORHED's goals and the program's evaluability. These include:

- Quality of research;
- Quality of educational programs as a check on the quality of graduates;
- Sustainability of NORHED-funded capacity developments, and
- System and institution-level conditions that enable or impede capacity building at the university level.

4.7. The *NORHED Program Presentation* states that "The purpose of the programme is to strengthen capacity in higher education institutions in lower middle-income countries... By increased capacity it is meant strengthened capacity for institutions in developing countries to educate more and better qualified candidates, and to increased quality and quantity of research conducted by the countries' own researchers." (Norad, 2012, p.4)

4.8. This quotation flags the *quality of research* and the *quality of graduates* as immediate goals. Indicator 8 measures scientific publications as an indicator of the amount of research, but not its quality. In section C of this chapter DPMG proposes changes to this indicator to measure quality.

4.9. Indicator 4 measures the relevance of educational programs and new graduates to labor market needs. However, NORHED seeks not just relevant skills, but also "better qualified" graduates. Standards for selecting students into NORHED-funded programs and new/revised NORHED-funded education programs are the primary means for improving the quality of graduates. At the least, there should be a check on the quality of new/revised educational programs. Student evaluations of programs and courses provide a valid way to measure course/program quality. FORSK could create a student evaluation template similar to that used in many universities to get feedback on the beneficiaries' views of NORHED-funded programs and courses.

4.10. In terms of the *sustainability* of the capacities developed with NORHED funding, the *NORHED Programme Presentation* makes clear that sustainability is a program objective.²² Indicator 7 (retention

²² Under section 4 on Guiding Principles, section 4.1, entitled "Systematically Addressing Sustainability Challenges", states that: "Systematic actions and plans for sustaining activities shall be identified and developed from the out-set of the project, including issues such as staff retention and plans for the continuation of study programmes, research activities and supporting facilities (e.g. laboratories, libraries) after the end of the project period. Project applications must include strategies for sustainability of the achievements of the project, for

rates of qualified academic staff at relevant unit...supported by NORHED) touches on sustainability. Grantees might also identify sustainability challenges in their discussion of risk factors (items 3.3.2 and 3.3.5). However, systematic data on sustainability are not collected.

4.11. The theory of change for NORHED, relative to its goals of more/better-qualified graduates and more/higher quality research,²³ identified several *system-level and institution-level factors* that international research shows enable or impede capacity building at the university level. Data on these contextual factors are especially important relative to Norad's learning objectives for NORHED because they can help interpret NORHED projects' progress or their lack thereof. For example, the institution's processes and standards for retaining and promoting faculty and staff are relevant to indicator 7. However, FORSK did not solicit data on these factors at the time of the grant applications, and the annual template also does not collect data on these conditions. Although grantees might identify one or more conditions in their discussions of risks in the annual template (items 3.3.2 and 3.3.5), FORSK lacks systematic data on them at the start of the projects or should these factors change, during their implementation.

4.12. At the *system level*, governance and funding are the factors that most systematically influence the situation of individual universities. These are: a) the degree of institutional autonomy and academic freedom enjoyed by higher education institutions; b) the accountability instruments (quality assurance and accreditation, assessment of learning outcomes, labor market results, research assessment); c) student support and financial aid for students from disadvantaged groups; and d) the funding available for research and teaching, especially for public universities that are heavily dependent on state resources.

4.13. At the *institution level* important factors are: a) the degree of alignment of the project's objectives with the overall strategic plan of the university;²⁴ b) good institutional governance, especially transparent and objective decision-making criteria and processes--for example, selecting faculty and students on merit grounds; and c) a university leadership that helps to integrate donor-funded projects into the standard processes of the university. These processes include the university's internal quality assurance mechanisms; its other on-going research activities; its performance rewards systems; and student support mechanisms (student services, financial aid, psychological and academic support).

C. NORHED's Standard Indicators Revisited: Findings

4.14. FORSK staff and staff of Norad's Results Management Section disagreed with some of the conclusions in the Inception Report. They highlighted, for instance, that some of the indicators were explicitly intended to be measured in a qualitative way: the inclusion of gender perspectives in the different NORHED funded programs; the influence of research on public policies; and the influence of research on private sector and civil society and changes in the broader institutional environment.

4.15. As stated in the terms of reference for this Evaluability Study, we were asked to revisit the question of whether the standard indicators could be used to assess the effectiveness of NORHED. In particular, we were asked to take into account three additional considerations: a) the fact that some indicators were explicitly designed to measure qualitative aspects of the program; b) the fact that grantees were supposed to report baseline values for all indicators in the annual reports, not necessarily in the results frameworks; and c) the institutional constraints that grantees face with regards to the availability and logistical capacity to collect information to measure progress.

4.16. We are sympathetic to these concerns and revisited our analysis of the standard indicators, bearing in mind these considerations. Since almost all grantees had submitted their annual progress reports by August, we were able to extend our analysis by reviewing how grantees seemed to interpret the standard indicator items in the annual report template.

reporting and adjustments throughout the life of the project." (*NORHED Programme Presentation*, Norad, 2012, p.7)

²³ *Op. cit.* Evaluation Series of NORHED

²⁴ Projects work best when the planned activities are consistent with on-going institutional efforts to improve the quality of teaching and learning

4.17. Below we describe our findings separately for each standard NORHED indicator:

Indicator 1: Number of new and revised Bachelor/Master/PHD programs/modules supported by NORHED

4.18. The Inception Report found that this indicator meets the CREAM criteria. Our review of the annual progress reports yields no further updates to the prior assessment.

4.19. **Recommended changes:** None. If the distinction between new and revised programs and/or if the nature of the revision is important to FORSK, this indicator could be sub-divided into two indicators, possibly distinguished by level (e.g., Master's versus Ph.D.):

4.20. 1.1 Number of new Bachelor/Master/PhD programs/modules established: counts of number of new Bachelor/Master/PhD programs/modules established.

4.21. 1.2 Number of revised Bachelor/Master/PhD programs: Number of existing programs with a revised curriculum; Number of new courses in existing programs; Number of existing programs with new course materials

Indicator 2: Number of Bachelor/Master/PHD programs/modules supported by NORHED with a gender perspective

4.22. This indicator continues to be problematic. In reviewing the annual reports, we encountered two additional reasons of concern:

- There is very little variation across projects in baseline values. With one possible exception, all of the annual progress reports indicate that *at baseline*, projects included a gender perspective. If all projects report that at baseline they included a gender perspective, there is no scope for improvement. The most desirable outcome that could be achieved—to include a gender perspective—will have already been achieved by the time that the projects start. In survey design terminology this is typically known as a “ceiling effect.” Since there is no variation across projects or over time, this indicator cannot explain any performance variation between projects.
- The narratives about how gender perspectives are to be included refer to multiple possible dimensions in which a gender perspective can be included. For example, some projects refer to aspects of the curriculum. Others refer to the composition of the student body; still others, to the composition of the faculty. Since there is no clarity ex-ante on what projects are exactly supposed to report, this narrative lends itself to selective reporting. Given that the framing of the report creates an expectation for projects to include a gender perspective, all grantees will tend to find a way to highlight the most relevant aspect of their projects. While this is not necessarily bad, it is heavily subjective and selective.

4.23. **Recommended changes:** FORSK needs to identify its specific concerns in order to reformulate the indicator to CREAM standards. Candidate alternatives at this stage may include: Number of core courses that explicitly deal with gender issues; Number of programs that explicitly use gender-based rules for recruitment and admission.

Indicator 3: Capacity to enroll and graduate students in NORHED-supported programs (Bachelor/Master/PHD)

4.24. This indicator is designed to capture the contribution of NORHED projects to the increased enrolment and graduation capacity of the supported units. It also aims at assessing the progression of enrolled students (internal efficiency) and the proportion of students who actually graduate, and presumably the duration to completion. It meets the CREAM criteria, and reviews of the first annual progress reports implied no changes in our prior assessment.

4.25. **Recommended changes:** None.

Indicator 4: Relevance of educational programs and new graduates to local, national and regional needs and labor markets

4.26. This indicator continues to be problematic for reasons similar to those identified for indicator 2 (gender perspectives).

- There is little variation across projects in baseline values. The majority of grantees report that their projects are relevant, leaving little room for improvement attributable to NORHED funding. In other words, again there is a “ceiling effect.” Consequently, this indicator cannot be used to explain variation in project success.
- The narratives about the relevance of programs at baseline are heterogeneous and appear to be highly selective. Since grantees are expected to report that their projects are relevant, they are trying to demonstrate relevance in the best possible light.
- In general, it does not seem plausible that grantees can judge the relevance to local, national, and regional needs and labor markets early in the project.

4.27. Recommended changes: This indicator could be improved by relying on more objective data. For outputs, these might include: a) number of new programs that prior to their creation received input from industry or government stakeholders; b) number of existing programs that received input for a new curriculum from industry or government stakeholders; or c) number of scholarships provided by industry.

4.28. For outcomes, possible indicators might be: a) number of graduates finding jobs in relevant local labor markets (total and by gender); b) Average number of job offers a graduate receives before accepting a job (total and by gender); c) Number of graduates starting up businesses in relevant local labor markets (total and by gender); or d) survey results of employers about the relevance of the training of the program's graduates to them.

Indicator 5: Number of academic staff with strengthened qualifications (Master/PhD) by relevant institutional level (institute/department/faculty) supported by NORHED

4.29. This is a good indicator to measure the capacity building efforts of the project. Our original assessment was that this indicator met the CREAM criteria, and we maintain this conclusion after reviewing the first annual progress reports.

4.30. Recommended changes: None. FORSK might want to assess the quality of additional qualifications, not just the number. In this case, FORSK could complement this indicator with one that weights the absolute number of academic staff by ratings of the quality of the institutions or programs where beneficiaries attained their “strengthened qualifications”. However, such an indicator requires some semi-credible rating or ranking of institutional or program quality (i.e. ranking, reputation). These ratings may be available for Norwegian universities, but probably not for LMIC universities. Although independent committees could be assembled to assign ratings, this action does not seem warranted for relatively small grants.

Indicator 6: Ratio of qualified academic staff (Master/PHD) to students by relevant unit (institute/department/faculty) supported by NORHED

4.31. Our original assessment noted that this indicator is a function of both qualified staff (numerator) and enrollment (denominator). Both of these are separate outputs that NORHED aims to increase. Programs may increase qualified staff and enrollment, both of which are positive outputs. If staff and enrollment increase in the same proportion, the indicator will register no change. For these reasons, at the time of the Inception Report we were cautious about using it to evaluate the performance of NORHED.

4.32. After reviewing the narrative reports, we identified an additional challenge for this indicator. This indicator requires that grantees calculate a ratio without specifying *exactly* what should be included in the numerator and in the denominator. To be clear, the indicator specifies that grantees are to calculate the “Ratio of qualified staff (Master and PhD holders) to student numbers at relevant unit at project startup.” However, this leaves considerable ambiguity for the grantees. For example, should *all* faculty be included or only full-time faculty? Should the number of faculty be adjusted to Full Time Equivalent (FTE) units?

4.33. Similar potential confusions arise with regards to the student body: how should part-time students be accounted for? How should dual program students be accounted for? One project, for example, reports on two different values of the student body since there is no guidance in the reporting template as to how the student body number for this indicator should be reported. Grantees realize this. One grantee in their narrative report states with regards to this indicator that they “Need to agree on which student group to use.” The fact that some grantees do not know ex-ante which student group they should be referring to in the calculation highlights the ambiguity in the question.

4.34. As a result, it is unclear how comparable this indicator is across projects. Grantees are likely to make different implicit assumptions about which reference group for teachers or students to use. This comparability problem is likely to severely hamper the usefulness of this indicator in explaining end line variation in project performance. There is simply too much scope for measurement error. As a general rule, it is always preferable to ask respondents to report on absolute, unambiguous values. The analyst can then compute the desired value in a uniform and comparable way.

4.35. Recommended changes: Rather than just using a single indicator that measures variations in two different groups, both potentially affected by NORHED (faculty and students), it would be useful to complement it with an additional indicator that focuses solely on faculty. A candidate additional indicator is the fraction of total staff with strengthened qualifications. By including this additional indicator, one will be able to disentangle whether changes in the ratio of qualified staff to students is primarily driven by changes in qualified staff or by changes in student enrollment.

4.36. The problem and solution for this indicator are analogous to that of measuring unemployment. The unemployment rate alone is problematic because it depends not only on the number of unemployed people (numerator), but also on the size of the labor force (denominator), both of which are affected by economic cycles. For this reason measures of the unemployment rate are usually complemented by measures of the participation rate.

Indicator 7: Retention rate of qualified academic staff at relevant unit (institute/ department/faculty) supported by NORHED

4.37. This indicator asks from grantees to report “Number of qualified staff (Master/PhD holders) who left the unit out of total number of qualified staff employed at the unit during the year prior to project startup” The Inception Report noted that this indicator is very important as a proxy measure of the capacity building impact of NORHED projects. If the academic staff members with Master’s degrees and/or PhDs do not stay on, the training investment will have been wasted. We were concerned about this indicator being potentially difficult to capture because of the long duration of training involved, especially in the case of new PhDs.

4.38. After reviewing the grantees’ narrative reports, however, we note three additional concerns with this indicator. The first concern is the actual wording of the indicator. The title of the indicator is “retention rate”, but the indicator actually asks grantees about the “attrition rate”—the staff that leaves. This kind of “mirror-wording” could be confusing to some grantees.

4.39. The second concern is that nowhere for this item does NORHED ask grantees to report the denominator, namely, the number of qualified staff employed at the unit in the year prior to project startup. As a result, all grantees are reporting absolute numbers, which makes explicit the fact that grantees are not viewing this indicator as a rate. It is unclear, then, how one could convert this information into a rate.

4.40. The third problem is that for projects that seek to strengthen the qualifications of existing staff, the number of qualified staff employed at the unit in the year prior to project startup may simply be zero. This would imply that the “attrition rate” is infinite because any number of staff that leave divided by zero is infinite.

4.41. Recommended changes: The first problem can be easily overcome by using consistent language. The second problem could also be easily overcome by ensuring that somewhere in the report grantees list all of the staff at baseline (i.e., at the start of the project or at the time of the first annual progress report) that receive NORHED support, total, according to qualifications and by whether they are full or part-time. Then, for this question (and perhaps also for the question about faculty-student

ratios) the template could simply ask grantees to always refer to that initial staff roster. Doing so minimizes interpretation and recall bias and ensures comparability across projects and over time.

4.42. This way of asking these kinds of questions is standard in panel survey methodology where, for instance, in the baseline survey enumerators create a household roster that is time-invariant and always refer to it in later rounds. It may also be advisable to ask grantees to list the staff funded by NORHED, state whether they are full or part-time and state the number of years they have been at the relevant unit. As with the indicator on student-faculty ratio, this would help reduce measurement error because: a) it reduces the calculation burden on the part of grantees; and b) makes explicit what is being asked for. NORHED staff can always calculate later the average in various alternative but comparable ways.

4.43. The third problem can be resolved if, instead of using the number of staff with strengthened qualifications the year before the program began as the denominator, the denominator is simply the number of staff in the relevant unit. Presumably this number will never be zero.

Indicator 8: Number of scientific publications (peer reviewed and others) by NORHED project to date

4.44. This indicator contains two sub-indicators: a) number of peer-reviewed publications; and b) Number of books, reports or other scientific publications, total and with female coauthors. The Inception report stated that this indicator (both sub-indicators) meets the CREAM criteria and is therefore a useful indicator for evaluating the performance of NORHED. We maintain this assessment after reviewing the projects' first annual progress reports. However, as noted earlier, this indicator does not measure the quality of the publications as a check on the quality of NORHED-funded research.

4.45. Recommended changes. Include two additional indicators to capture the quality of research publications: a) whether the publication was in a peer-reviewed journal and b) the impact factor of the peer-reviewed journal.

Indicator 9: Number and type of other dissemination activities to date

4.46. The Inception report concluded that this indicator met the CREAM criteria and thus was useful for evaluating the performance of NORHED. However, reviews of the first annual progress reports raise two issues.

4.47. The first concern is that grantees are allowed to select the dissemination activities that they have accomplished. While this is desirable from the perspective of acknowledging all efforts in the field equally, it complicates comparability because grantees can potentially selectively report which activities they want to highlight initially.

4.48. Selective reporting in this context is problematic because—as is clear from reviewing the first annual progress reports—grantees are only listing baselines and end line targets for the activities that are salient/relevant to them in the first year. It is reasonable to expect different types of dissemination activities to become relevant later in the projects' lifecycle. However, the identification of dissemination activities salient/relevant to them in year one and targets for these types of activities by year 5 may inadvertently "anchor" their thinking in this type of activity, discouraging them from devoting effort to other dissemination activities. This is analogous to a common problem in education research called "teaching to the test."

4.49. Best practice evaluation dictates that baseline values are, in fact, measured at baseline, not retrospectively at a later date. For this indicator, it will be hard for grantees to pre-specify baselines for all of the dissemination activities that might occur during the lifecycle of the project simply because they find it difficult to anticipate the kinds of activities that might take place in the future.

4.50. Recommended changes. It is advisable to explicitly ask grantees to report on a pre-specified list of dissemination activities. This list could include many items, most of which could be drawn from the first year annual reports. For many grantees the baseline values for some of these activities will be zero. But this way, grantees are encouraged to consider a wide array of dissemination activities going forward and the baseline values for those would be establish, in effect, at baseline and not later, retrospectively.

Indicator 10: Uptake/influence of NORHED-supported research in public policies

4.51. This indicator contains two sub-indicators: a) Has NORHED-supported research influenced national/regional laws, regulations, standards, services or practices? And b) How has NORHED-supported research influenced national/regional laws, regulations, standards, services or practices?

4.52. During our meetings with Norad staff, we agreed on the importance of measuring this indicator qualitatively (i.e. as a narrative). As is expected from the first-year narrative reports, projects have not made enough progress to yield quantifiable progress in this indicator, which will likely only occur, if at all, at more advanced project stages. Therefore, beyond our initial comments in the Inception report, at this point we do not have further reactions with regards to this indicator.

4.53. Recommended changes. In order to make this a numerical indicator, the first sub-indicator can be revised to monitor 1) the number of Parliamentary testimonies by NORHED-supported researchers; 2) the number of presentations to government officials by NORHED-supported researchers; and 3) the number of contracts of NORHED-supported units with government agencies.

4.54. The second sub-indicator could be eliminated as part of the indicator set, even though it should be qualitatively documented in evaluation and in projects' progress reports.

Indicator 11: Uptake/Influence of NORHED-supported research findings/new technologies/innovations /solutions by local communities/civil society/ private sector

4.55. This indicator has two sub-indicators: a) Have research findings/technologies /solutions/innovations been taken up by local communities/civil society/private sector? And b) How have research findings/technologies/solutions/innovations been taken up by local communities/civil society/private sector? After reviewing the first annual progress reports, we highlight two aspects of concern.

- From the progress reports it is unclear what the baseline values for this indicator are. One might be inclined to assume that for all grantees the baseline value is zero (i.e. no influence). However, some grantees might have an established pipeline of influence with one or more of the intended beneficiaries. It would be helpful to determine baselines as explicitly as possible.
- Perhaps because of the way the indicator is framed (it is desirable to have projects influence the private sector and civil society), grantees are steered to conform to that expectation. The framing of indicators needs to be neutral. Value judgments are normative and should only come as a result of interpreting the evidence.
- Grantees may be misinterpreting what "uptake" of research/findings/ technologies/solutions/ innovations means. For example, one project reports: "Senior professionals from public sectors, NGOs and civil society leaders were engaged in different courses of MPPG as guest lecturers to make interactive sessions with students and professionals." While this is clearly valuable engagement between academia and the civil society, this does not respond directly to what the indicator is requesting from grantees. It does not refer specifically to research, and classroom engagement does not imply uptake of findings/ technologies/solutions/ innovations by civil society or the private sector.

4.56. Recommended changes: Similar to the recommendations for indicator 10, this indicator can be revised to include more objective data. The first sub-indicator could be revised to: Number of patents; number of media citations; amount of royalties paid to institutions/units/researchers; number of contracts with private sector companies; number of collaboration agreements with industries; and number of events directed to reach industry and community members. Each of these counts must be attributable to NORHED-supported research.

4.57. The second sub-indicator can be eliminated as part of the indicator set, even though should be qualitatively documented in evaluation and in projects' progress reports.

Indicator 12: Knowledge transfers within South-South and South-North networks and partnerships as a result of the NORHED project

4.58. This indicator has two sub-indicators: a) Have knowledge transfers within South-South and South-North networks and partnerships taken place? And b) How have knowledge transfers within South-South and South-North networks and partnerships taken place?

4.59. After reviewing the narrative reports, we note that rarely do projects distinguish between South-South and South-North networks and partnerships for this indicator. It may, therefore, be desirable to ask grantees to specify whether the knowledge transfer is South-South or South-North. Also, from the perspective of aggregation and the evaluation for performance of NORHED, it may be advisable to review the current narrative reports and create separate categories for the most common types of knowledge transfers that occur in the first year. Then, in future annual narrative reports, Norad can more explicitly ask grantees to report progress in each of these categories.

4.60. Recommended changes: Similar to the previous two indicators, this indicator can be revised to be more quantitative. Sub-indicator 12.1 can be revised to look at: Number of events directed at increased collaboration with other South institutions and networks; number of events directed at increased collaboration with North institutions and networks; number of presentation in academic conferences with South-South and South-North co-authors; number of grant proposals with South-South and South-North collaborators; number of publications by type (peer reviewed, books etc.) coauthored among South-South collaborators; number of publications coauthored among South-North collaborators; quality index of publications, when available; number of patents with South-South and South-North collaborators.

4.61. Sub indicator 12.2 can be eliminated as part of the indicator set, even though it should be qualitatively documented in evaluation and in the projects' progress reports.

Indicator 13: Changes in the broader institutional environment at NORHED-supported institute/faculty/department that strengthened the unit's capacity in education and research

4.62. Since this is a qualitative indicator, it would be desirable to elicit more detail from grantees about how the changes (if any) came into effect. The narrative reports respond to this item, but often do not show how NORHED funding connects to changes in the broader institutional environment. For example, one grantee reports these changes: "Software (Tally) based financial system in Bangladesh and Nepal" and "Financial policy at PPG program of NSU, Bangladesh." It is unclear what to make of statements such as these.

4.63. Recommended change. This indicator is best assessed qualitatively, but bullet-point statements, like the ones submitted in the first annual progress reports, are not a good qualitative benchmark. NORHED may wish to probe on specific types of changes of interest to them, such as gender. For example, instead of "Financial policy at PPG program of NSU, Bangladesh" a more qualitatively substantive, logically and causally connected statement could be as follows: "The way that NORHED requests financial reports from us is new to our institution. While adjusting to this way of financial reporting has been challenging, after working with this financial reporting template for some months its advantages over our prior way of keeping track of financial resources became evident. As a result, the entire department has adopted a similar financial reporting template, and our engineers have created specific software to systematize the process moving forward."

Indicator 14: Access to libraries, laboratories and ICT for staff and students in NORHED-supported institutes/departments/faculties

4.64. The inception report identified this indicator as unnecessarily subjective. Reviews of the first annual progress reports reveal additional problems with this indicator.

- The indicator is not framed neutrally. It conveys the expectation for the funded project to increase access to laboratories, libraries and ICT. The increase should be determined from the change from an objectively quantifiable baseline value to an objectively quantifiable end line value.
- As with other indicators with similarly slanted framing (e.g. relevance, uptake), many projects report that in the first year, access to libraries, laboratories and ICT for staff and students has increased. Therefore, there is very little variation at baseline across projects in this indicator.

- The framing of the indicator artificially creates a “ceiling effect.” If access is increased in the first year, then projects have already achieved the desirable outcome. It will therefore be impossible to gauge any further progress in the indicator in the remaining years of the projects. As a result, projects are unlikely to vary much at the end line, rendering this indicator not useful for explaining variation in project performance.

4.65. Recommended changes: This indicator should be revised to be quantitative (not yes/no). Norad staff raised reasonable concerns about measuring this indicator in more systematic ways due to technological barriers among grantees. Perhaps more objectively quantifiable versions of the indicator such as the ones we proposed in our inception report can be piloted and discussed with a sample of grantees: Number of staff using the library in the last year (or month); number of library books borrowed by staff in last month; number of staff using laboratories in the last month; number of hours staff used laboratories in the last month; number of staff who used ICT in the last month; number of hours staff used ICT in the last month. All of these possibilities have two problems: a) the administrative burden of obtaining counts, and b) potential reliability and validity of the responses. Responses will be self-reports and thus subject to bias. They could be inflated to convey a good impression--perhaps more to one's colleagues than to NORHED. They could be deflated because of forgetting. Asking about recent experience should reduce this source of bias. One possibility that has been repeatedly discussed to improve on the measurement of this dimension is the use of internet-based survey instruments such as SurveyMonkey.

D. Conclusions

4.66. NORHED's standard indicators are substantively relevant, although incomplete relative to NORHED's objectives. They do not measure four important variables: quality, not just amount, of research; quality of educational programs as a check on the quality of graduates; sustainability of NORHED-funded capacity developments, and system and institution-level conditions that enable or impede capacity building at the university level.

4.67. However, the main problems with the indicators lie with their operationalization. CREAM criteria were used to evaluate the indicators as originally specified. The first annual progress reports were reviewed to check how grantees were interpreting (or misinterpreting) requests for updated M&E data on each indicator. Against these standards, only four met CREAM standards (indicators 1, 3, 5, and 8), although three others can probably be quite easily modified to eliminate measurement problems.

4.68. The indicators had problems that commonly arise in survey design, the most problematic being:

- **Ambiguity in what NORHED is requesting from grantees.** Ambiguity creates considerable room for interpretation, which potentially compromises comparability across projects and over time. For example, indicator 6 involves "Ratio of qualified academic staff (Master/PHD) to students". However, it is not clear whether *all* faculty should be included, only fulltime faculty, or whether the number of faculty should be adjusted to Full Time Equivalent (FTE) units.
- **Ceiling effects.** Many projects have achieved the most desirable outcomes at baseline, leaving no room for improvement. Indicators that exhibit ceiling effects cannot be used to evaluate project performance because there is simply no variation across projects or within projects over time.
- **Measurement error.** Indicators that require grantees to calculate “rates” or “ratios” are always problematic. In general, one of the maxims of best practice survey methodology is to minimize the burden of calculation on the part of grantees. In other words, even if what NORHED is interested in measuring is a “rate”, grantees should never be asked to calculate that rate themselves. They should, instead, be asked to report the most basic information that would then allow NORHED to calculate the relevant rate. This is why, for example, one never asks survey respondents to estimate the fraction of their monthly consumption spent on food. One instead elicits how much respondents spent on meat, potatoes, clothes, transportation, TV, etc. Analysts then compute the relevant statistics.
- **Selective reporting.** The framing of many questions creates an expectation that grantees should be showing progress towards that goal. Since grantees are simultaneously undertaking those activities, reporting on those activities, and being judged on them, they have an incentive to present facts in the most favorable light. This makes it very hard to identify problems early

in the lifecycle of projects and to provide real-time feedback for improvement. As a general rule, all indicators should be framed as neutrally as possible to prevent conveying expectations that the funded project should increase access to, for example, laboratories, libraries and ICT. The increase should be determined from changes between an objectively quantifiable baseline value and an objectively quantifiable end line value.

- **Extremely general reporting on qualitative items.** Qualitative items are not always asked in ways that specify the change process. If the change process is poorly detailed, the causal connections between inputs/activities and output/outcomes cannot be credibly established.

E. Recommendations

4.69. Add indicators to measure the quality of NORHED-funded new/revised educational programs, sustainability of NORHED-funded capacity developments, and system and institution-level conditions that enable or impede capacity building at the university level. Indicator 8 can be revised to measure quality, not just amount, of research.

4.70. Revise the standard indicators in the annual progress report template to maximize the reliability and validity of M&E data. The report proposes many possible ways in which FORSK might modify the standard indicators in the annual progress report template to resolve the measurement problems identified. Without pretests or field work to develop a nuanced understanding of grantees' M&E contexts, it is difficult to judge the realism of some of these suggestions. However, FORSK is encouraged to implement as many of these suggestions as are feasible in order to yield sound, objective and empirically-grounded conclusions about NORHED's performance. The revision process should be alert to the types of problems with the current indicators that were identified above:

- a. Ambiguity in what NORHED is requesting from grantees;
- b. Ceiling effects;
- c. Measurement error when grantees are asked to make calculations;
- d. Questions framed in ways that encourage selective reporting; and
- e. Lack of guidance to grantees on how to respond to qualitative items.

V. NORHED Supervision Processes

A. Introduction

5.1. To arrive at a proper conclusion about the evaluability of NORHED, FORSK's supervision processes should be assessed. They constitute a major source of information about how the theory of change is unfolding at the program and project levels.

5.2. FORSK reviews each grant and meets with each grantee approximately annually. As indicated in Table 10, this annual supervision process for each project yields four main supervision products, in addition to considerable correspondence: the Mandate, Agreed Minutes of the annual meeting, a back-to-office report, and the finalized annual narrative report.

Table 10: Steps and products of FORSK's annual supervision process

Step #	Action by whom	Supervision product
1	Grantee submits annual financial report and annual narrative report, using annual report template	
2	FORSK project officer reviews the annual financial and annual narrative report. If time permits, the officer sends comments to the grantee in advance of the annual meeting, asking, as needed, for corrections or revisions.	
3	On the basis of this review, any revised financial and narrative reports, and the annual work plan and budget for the current year, FORSK prepares what is called a "mandate" that is used to guide its annual meeting with the grantee. The mandate states whether Norad has received sufficient information to approve the narrative and financial reports at the time of the annual meeting.	Mandate
4	FORSK holds annual meeting with the grantee, usually 2-3 months after submission of the annual reports. Grantee calls and chairs the meeting. All partner institutions are invited to attend. FORSK encourages the inclusion of university administrators, financial departments, students, and Ph.D. candidates.	
5	Grantee prepares Agreed Minutes of the meeting with FORSK, submitting them no later than two weeks after the meeting.	
6	FORSK staff member completes notes on the annual meeting with each grantee-- something like a "back-to-office" report.	"Back to office" report
7	FORSK staff member reviews grantee's Minutes for annual meeting and, perhaps after requesting revision of the minutes, approves them	Approved minutes of annual meeting between FORSK and grantee
8	Grantee revises annual financial and narrative reports on the basis of meeting with FORSK and submits it to FORSK	
9	FORSK approves grantee's annual reports (financial and narrative) as final, sometimes after additional iteration with the grantee.	Approved annual report

Source: Norad's Section for Research, Innovation and Higher Education (FORSK).

5.3. As noted in chapter I, at a minimum, an adequate assessment of the supervision process requires desk reviews of the supervision documents for a sample of projects to determine the nature and quality of the information generated during the supervision process. Optimally, desk reviews should be complemented by field visits to a small sample of projects.

5.4. Budget limitations for this exercise preclude both desk reviews and field visits. This reality leaves a question mark about the evaluability of NORHED since the nature and quality of this source of information for a theory-based evaluation of the program are only sketchily known. As a consequence:

- Norad's Evaluation Department will have an incomplete picture of the evaluability of NORHED.
- FORSK will lack the feedback that it needs to affirm the quality of its supervision process and/or to modify particular aspects of the process to increase its validity and reliability.

B. Findings

5.5. Within the scope of this exercise, five findings were identified.

- 1) The annual narrative report template collects data on an appropriate set of topics. It does not collect data on a few topics important to its goals or for interpreting results. In some cases the questions as phrased should yield valid and reliable data. In a number of cases, especially for items that request updated data on the standard indicators, they will not.
- 2) The M&E data collection schedule is defined and makes sense in most cases. The data collection schedule for only two standard indicators is undefined.
- 3) FORSK has no dashboard that summarizes "at a glance" each project's implementation status and likelihood of reaching its development objectives. It has commissioned annual summaries of the results of project implementation relative to the standard indicators and other Norad priorities. These summaries will yield important information for the NORHED management, but they aggregate data across the projects for the program as a whole. This is not the same as the proposed dashboard idea, which would discriminate between projects that seem on track to succeeding and those that, for whatever reason, are in trouble and for that reason warrant particular FORSK attention.
- 4) Grantees self-report indicator data. The literature identifies potential validity problems with self-ratings. Although FORSK staff review these data, these reviews have limits. How FORSK writes the TORs for independent mid-term and/or final reviews of the program will determine if they will obtain independent checks on the validity of grantee-provided data.
- 5) FORSK is considering commissioning independent tracking of whether aid-funded interventions actually benefit intended beneficiaries. Commissioning such studies is good aid management.

Annual Progress Report Template

5.6. Table 10 shows that the grantee's submission of a completed annual progress report is the starting point for FORSK's annual supervision of each project. FORSK's template for this report (see Annex C) was reviewed to determine if any modifications of the template could enhance the relevance and quality of information provided by grantees each year.

Adequacy of data collection categories

5.7. The template collects data in several categories.

- Implementation status of the project (items 3.3.1, 3.3.5, and 3.3.6)
- Risk factors (items 3.3.2 and 3.3.5)
- Performance against Norway's broad objectives, as reflected in the *NORHED Programme Presentation* (resource efficiency--item 3.3.4; gender equality, environmental issues, human rights, and conflict sensitivity--item 3.3.3)
- Performance on NORHED standard indicators (items 4.1-4.14)
- Performance on project-specific indicators (item 5)
- Financial overview (item 6)
- Case/success story (description of a positive result at any level of the results chain that project experienced in the reporting period--item 7.1)

5.8. These data collection categories make sense. However, as section B of chapter IV points out, the template does not request data for four domains germane to NORHED's goals and the program's evaluability: quality of research; quality of educational programs as a check on the quality of graduates; sustainability of NORHED-funded capacity developments; and system and institution-level conditions that enable or impede capacity building at the university level.

5.9. Despite missing data on these items, Table 11 shows that the annual progress report template meets several data requirements for evaluating projects during implementation.

Table 11: Crosswalk between data requirements during implementation and template items in NORHED annual progress report

Data required	Data collected in annual report
1. Updated disbursement data	Data from Norad's Plan, Tilskudd, Avtale (PTA) system; Section 3, item 3.3.6; Section 6; Section 8 (attachment of certified annual financial statement)
2. Updated data on standard indicators	Section 4, items 4.1-4.14
3. Status updates on project's causal path or theory of change, including updated data on project-specific indicators	Section 5, which requests data on the status of project-specific activities, outputs, and outcomes. Section 3, item 3.3.1, also provides information on changes to the path.
4. Update on system and institution level conditions that may enable or impede results achievement	Not available
5. Updated information on risks, anticipated and newly emergent	Section 3, items 3.3.2 and 3.3.5
6. Update on sustainability of NORHED-funded capacity developments	Not available
7. Assessment of implementation status: issues, solutions, successes	Section 3, items 3.3.1, 3.3.5, and 3.3.6; section 6; section 7
8. Independent verification of performance data	Annual report not structured to provide independent verification, although FORSK project officers will be alert to obvious problems with grantee-provided data.
9. Independent tracking of whether inputs are benefitting intended beneficiaries	Requires an instrument different from the annual report
10. Data on performance of overall NORHED portfolio, based on summaries of project-specific performances	Requires an instrument different from the annual report

Quality of data collection items

5.10. Rephrasing some items in the annual template should improve the validity and reliability of the data collected, **especially the standard indicators for which chapter IV identified reliability and validity problems.** In reformulating these in the annual progress report template, FORSK and Norad's Results Management section can consider the changes suggested in chapter IV.

5.11. However, there are two other items whose current phrasing is likely to generate data that are neither especially reliable nor valid. These are items 3.3.3 and 3.3.4 in Section 3 of the annual report template. The first of these items (3.3.3) asks for a "Brief description of the project's achievements on gender equality, environmental issues, human rights and conflict sensitivity (if relevant)." The second (3.3.4) asks for a "Brief description of the project's efficiency (how efficiently resources/inputs are converted into outputs, i.e. value for money)."

5.12. Both of these items are very general. If FORSK's intent is simply to use these items to start a discussion about them during the annual meetings, they will suffice. However, if there is any intent to aggregate achievements across projects or to compare projects on these dimensions, they are not suitable. They are both vulnerable to multiple interpretations, thus creating error. They also allow grantees to "pick and choose", potentially creating a positive bias.

M&E Data Collection Schedule

5.13. The annual progress report template specifies the schedule for M&E data collection schedule by item. **In almost all cases the intended schedule makes sense.** Most indicators (indicators 1, 2 3, 5, 7, 8, 9, 10, 11. 12, 13) are tracked annually, with only indicator 6 explicitly being tracked in year 1 and in year 5. For indicators 10 and 11, the template asks for updates only when events that constitute instances of the indicator occur. The data collection schedule for two standard indicators (indicators 4 and 14) is undefined.

Utility of a Dashboard to Monitor the Health of the NORHED Portfolio

5.14. At present FORSK has no dashboard that summarizes "at a glance" each project's implementation status and likelihood of reaching its development objectives. It has commissioned annual summaries of the results of project implementation relative to the standard indicators and other

Norad priorities. These summaries aggregate data across the projects for the program as a whole--e.g., gender mainstreaming activities implemented or disbursements of funding per sub-program and the NORHED program overall. For quantitative indicators, the summary can show the average or distributional performance of the portfolio for each indicator--e.g., the number of scientific publications.

5.15. The dashboard idea is different. It requires that FORSK staff conclude their annual review of each project by using their judgment to assign a summary rating (using a pre-determined rating scale) of that project's implementation progress and of the likelihood that it will attain its intended outcomes. Displaying these ratings in an "at-a-glance" dashboard display gives FORSK management a good way to monitor the overall health of the NORHED portfolio. It quickly shows management whether something is wrong or something is right.

5.16. A good dashboard has certain characteristics. It is simple and communicates easily. It has minimum distractions. It focuses on the users' particular information needs. It applies human visual perception to the visual presentation of information.

5.17. The dimensions rated should be the project performance indicators of particular concern to FORSK and Norad. A dashboard is most useful when it concentrates on particularly important variables. Variables such as these could be considered, with the first two seemingly the most important: the project's implementation status; its progress toward the project's objectives; the quality of the M&E data generated by the project; the quality of project management; likelihood of sustainability; or progress toward gender equality. The dashboard could also show the percent of each grant disbursed at any given point in time, using data from Norad's PTA system. Such data will reveal unusually slow disbursements or unusually fast disbursements, thus flagging outliers that may portend implementation problems.

5.18. Each performance indicator should be rated on a scale, with disbursement data perhaps being represented as a percent of the total grant. An extensive research literature tests the reliability, validity, and discriminating power of scales with different numbers of response categories (e.g., Preston and Colman, 2000). In general, scales with 2, 3, or 4 response categories are the least reliable, valid, and discriminating; scales with 7 or more response categories are, overall, the most reliable, valid, and discriminating.

5.19. Ultimately, the choice of response scales must be driven by the rating objectives. It is critical that each point in the scale be carefully defined to minimize subjective interpretations of the scale point by the rater--in this case, the FORSK project officer. The effort to anchor each point on the scale in a meaningful distinction in the project's performance will reveal the number of scale points that work for NORHED.

5.20. There is no right or wrong between the dashboard idea and the annual summary data that FORSK is now obtaining. They are different, and they could complement each other. The critical issue is what information FORSK management finds most helpful in monitoring the health of the NORHED portfolio of projects.

Independent verification

5.21. Independent verification is required to protect against (often unconscious) positive bias. The potential for positive bias is well documented in the development literature and in the economics literature on the theory of the firm and principal-agent problems (e.g., Burger and Owens, 2010; Clements, Chianc, and Sasaki, 2008; Clements, n.d.; Michaelowa and Bormann, 2006; World Bank, 1992²⁵).

5.22. Incentives to positively bias assessments of aid performance affect those implementing the projects (grantees in the NORHED case), those supervising them (FORSK in this case), officials of the

²⁵ The World Bank's Quality Assurance Group (QAG) specifically assessed the candor and realism of task teams' ratings of projects during their implementation. These ratings included ratings of the projects' implementation progress and progress toward development objectives. QAG consistently found that the percent of projects that should have been rated as failing projects was double the rate identified by the task teams' ratings. Thus, if 15 percent of the projects were rated as failing projects, the actual percent was closer to 30 percent.

aid agency (Norad), and contractors employed by the aid agency to evaluate the projects. The most obvious source of bias is the possibility that one's future access to resources may be diminished if one's work is seen as unsuccessful or if, as an evaluator, one shows that the agency's aid has been unsuccessful.

5.23. At the same time, the further removed the evaluator is from direct connection to the project, the lower the chances of bias. The earlier in the project that negative findings are surfaced, the lower the chances of bias. Earlier findings can be framed as "learning" because grantees and those supervising them have a chance to correct them. These conditions reduce incentives to hide defects.

5.24. Table 1 indicates that independent verification enhances the evaluability of a program in two cases: a) checks on the accuracy of performance data provided by those implementing a project; and b) in-depth tracking of whether and how aid-funded activities actually benefit intended beneficiaries.

5.25. Independent verification of performance indicator data. NORHED grantees self-report the status of their projects on the standard and the project-specific indicators. FORSK's project officers undoubtedly probe the accuracy of these self-reports in their annual meetings with grantees. However, they may lack the time or technical expertise required to identify questionable data. They are also subject to some of the same incentives that can bias grantees' self-reports.

5.26. Article X of Norad's legal agreement with each grantee specifies that a final and /or mid-term review of the Project will take place during the Support Period at a time agreed between the Parties. Depending on how the TORs for these reviews are written, they may protect the program against positive bias. The TORs for these independent final and/or mid-term reviews may require the evaluators to collect their own data, verify the accuracy of the data provided by grantees, rely on the data provided by grantees, or use some combination of these approaches.

5.27. If evaluators collect their own data or verify the accuracy of the data provided by grantees, these assessments will check for potential positive bias in the data that grantees provide. However, for reasons of time, budget, or feasibility, the TORs may have to ask evaluators to rely on the data generated by the projects. In this case the potential for positive bias remains unchecked, and the need for independent checks remains. To address this issue, FORSK could commission probes of a random (or purposive) sample of indicator data for a random sample of projects. The findings can reassure FORSK about the accuracy of self-report data or give them opportunities to work with grantees to correct them.

5.28. Independent tracking of whether aid-funded interventions actually benefit intended beneficiaries. FORSK is considering commissioning this type of study, often called public expenditure tracking surveys and quantitative service delivery surveys. They are "follow the money" or "follow the output" studies that help to explain disappointing outcome results or yield a more in-depth measure of outcomes. Did the intended money in fact reach the school (Reinikka and Svensson, 2004)?²⁶ Did the milk reach the school child (World Bank/Inter-American Development Bank, 2002)?²⁷

²⁶ According to official statistics, 20 percent of Uganda's total public expenditure was spent on education in the mid-1990s, most of it on primary education. One of the large public programs was a capitation grant to cover schools' non-wage expenditures. During 1991-1995, the schools, on average, received only 13 percent of the grants. Most schools received nothing. The bulk of the school grant was captured by local officials (and politicians). The data also reveal considerable variation in grants received across schools, suggesting that rather than being passive recipients of flows from the government, schools used their bargaining power to secure greater shares of funding. Schools in better-off communities managed to claim a higher share of their entitlements, making actual education spending, in contrast to budget allocations, regressive.

²⁷ This was the question asked about the *Vaso de Leche* ("glass of milk") program, the most important supplementary food program in Peru that included a number of foodstuffs, not just milk. The analysis showed that leakages in Peru are significant and far more pervasive and extensive at the bottom of the chain than at the top. From the entire amount of public funds intended for the *Vaso de Leche* program, barely 29 percent get to their intended beneficiaries. This did not mean that 71 cents from each dollar was fully lost in corruption costs. Rather, the diverted resources got leaked away through a combination of off-budget administrative costs; expenditure on non-eligible products; in-kind deliveries to non-beneficiaries; fees for overpriced items; and, last but not least, sheer corruption. These results challenge the predominant view that organizations that are closer to the people necessarily perform better in service delivery. This is not the case if the local organizations are not transparent and do not practice accountability, which was the case for the *Vaso de Leche* Committees in Peru.

5.29. Such studies typically "unpack" the delivery chain to see if the intended path was followed. As noted earlier, NORHED funds a range of "typical" interventions. One of these is "Scientific Equipment and Small Scale Infrastructure". A check on the delivery chain for this type of intervention might be:

1. Did the grantee purchase the intended equipment?
2. Did the grantee receive the intended equipment?
3. Is the equipment available for use, or is it still in its original packing boxes?
4. Is it available for use by the intended beneficiaries (e.g., graduate students or faculty for research), or is access highly restricted in order to "safeguard" the equipment?
5. Do the intended beneficiaries know how to use the new equipment properly?
6. Can the intended beneficiaries identify specific benefits from using the equipment for their education or research?

C. Conclusions

5.30. The analysis of FORSK's supervision of the NORHED projects was limited, but found the following.

5.31. The annual progress report template is relevant to the program's objectives, although it fails to collect data on some variables germane to NORHED's objectives. Several items are stated in ways that should generate relatively valid and reliable data. However, some items are not, especially selected standard indicators identified in the last chapter. The M&E data collection schedule makes sense in most cases, with the data collection schedule for only two standard indicators being undefined.

5.32. FORSK does not now have something like a dashboard that summarizes each project's implementation status and its likelihood of attaining its intended outcomes. It does obtain summary achievements across the portfolio for specified indicators. There is no right or wrong between the dashboard idea and the annual summary data that FORSK is now obtaining. They are different, and they could complement each other. The critical issue is what information FORSK management finds most helpful in monitoring the health of the NORHED portfolio of projects.

5.33. Grantees self-report indicator data. The literature identifies the likelihood of validity problems with self-ratings. Although FORSK staff review these data, these reviews have limits. Norad's legal agreement with each grantee specifies that a final and /or mid-term review of the Project will take place during the Support Period. Depending on how the TORs for these reviews are written, they may or may not protect the program against positive bias. If the potential for positive bias remains unchecked, and the need for independent checks remains, FORSK could commission probes of a random (or purposive) sample of indicator data for a random sample of projects.

D. Recommendations

5.34. These recommendations are directed to FORSK and Norad's Results Management Section.

5.35. Revise the NORHED annual progress report template in order to increase the evaluability of NORHED as much as possible:

- As recommended in chapter IV, add data collection items for the quality of NORHED-funded educational programs and courses; sustainability and enabling/impeding condition.
- Revise selected standard indicators to increase the quality of the data that they will yield. Refer to recommended ways to improve the indicators specified in chapter IV.
- Define the M&E schedule for indicators 4.4 and 4.14

5.36. If regarded as useful, annually rate each project on its implementation status and likelihood of attaining its intended outcomes. Create a dashboard that summarizes the health of the NORHED portfolio. If FORSK wishes to explore this possibility, the staff should identify a limited number of rating dimensions that they consider especially revelatory of the health of the portfolio. Scale

options should be carefully thought through, with each point on the scale anchored in a distinction that it is possible to make for a project. Consider adding data on disbursement rates from Norad's PTA.

5.37. If the TORs for the planned midterm and/or end of project evaluations do not check the accuracy of the data provided by grantees, protect the program against the inevitable positive bias associated with self-reported data by commissioning independent checks of output and outcome data for a random (or purposive) sample of projects.

5.38. FORSK is now considering commissioning independent tracking of whether NORHED interventions actually benefit intended beneficiaries. Pursue this possibility. The delivery chain from resources to benefits for intended beneficiaries can often be a long one, with multiple points along the chain that can cause benefits to fail to materialize.

VI. Can NORHED be adequately monitored and evaluated?

6.1. Conclusions about whether NORHED can be adequately monitored and evaluated are based on the data required to evaluate NORHED. Table 1 of chapter II on the framework for the evaluability assessment specifies these data requirements. The requirements are revisited in light of the findings for chapters III, IV, and V. As noted, desk reviews and field studies of Norad's supervision processes for this program were outside of the scope of this assignment.

A. Specific Conclusions

6.2. Theory of change for NORHED. With some caveats, the implicit theory of change for NORHED has been judged to be adequate, relative to the international literature on theories of change for capacity development at the tertiary level (Norad, 2014). This literature identifies conditions that importantly affect the development of capacities at the university level in LMICs. Although the NORHED documents discuss enabling/impeding conditions, they were not systematically surfaced. Thus, the collection of data from the projects that a systematic treatment might have triggered did not occur.

6.3. Comparable data on these conditions are important for learning about the limits of what NORHED can hope to achieve and for analyzing the reasons for success or failure.

6.4. Theories of change for the NORHED projects. Operationally, the results frameworks for the projects constitute their theories of change. The evaluation of these frameworks found a pattern of weak links along the results chain. The only evaluation model available for NORHED, a theory-based model, requires attention to the entire results chain in order to draw even remotely credible conclusions about attribution/contribution of the aid program/project. Since clarity about the entire results chain for each project is crucial for evaluating the overall NORHED program, the weak results frameworks of many of the projects undermine program evaluability.

6.5. Performance indicators common to all projects. The indicators are substantively relevant, but did not measure four variables germane to NORHED's objectives and evaluability. Only four of the 14 standard indicators were well operationalized, meaning that the data yielded by the remaining 10 indicators will not be interpretable or will not register the range of change that NORHED intends to allow.

6.6. Baseline data for output and outcome indicators. Baselines are the *sine qua non* for evaluating projects' achievements--and ultimately those of NORHED. The damage that gaps in baselines will inflict depends on which are missing. Some of the indicators are more central to judging the performance of NORHED than others. If projects have baselines for core indicators, the damage will be more limited.

6.7. Data to monitor and evaluate project implementation. Since a primary purpose of the annual report template is to collect updated data on the standard indicators, problems with these indicators undercut the utility of FORSK's annual progress report template. Although FORSK obtains annual summaries of data on the standard and other indicators, they cut across the projects. The lack of a dashboard to display annual ratings and disbursement data for each individual project means that FORSK has no easy way to monitor the health of each grant within the NORHED portfolio. Depending on how it writes the TORs for independent midterm/final evaluations of the program, FORSK may or may not have ways to protect against self-report bias on the part of grantees. It does not yet have in place, but is considering independent tracking of whether NORHED interventions actually benefit intended beneficiaries.

B. Overall Conclusion

6.8. Using a binary scale to judge whether NORHED can or cannot be evaluated is inappropriate. This review shows that some aspects of NORHED can be properly evaluated. It also shows that several standard indicators for the program do not meet criteria for well-designed indicators; that baselines for output and outcome indicators for the standard indicators are often missing; and that grantee-generated data on outputs and outcomes may or may not be independently verified. Revising the standard indicators to meet criteria for valid and reliable measurement would have the single biggest effect on improving NORHED's evaluability.

References

- Altbach, P., and Salmi, J. (2011) *The Road to Academic Excellence: the Making of World-Class Research Universities*. Washington, D.C.: World Bank.
- Burger, R. and T. Owens. 2010. "Promoting Transparency in the NGO Sector: Examining the Availability and Reliability of Self-Reported Data." *World Development* Vol. 38, No. 9, pp. 1263–1277.
- Clements, P. "Monitoring and Evaluation for Cost-Effectiveness in Development Management." Draft. Unpublished document.
- Clements, P., T. Chianca, and R. Sasaki. 2008. "Reducing World Poverty by Improving Evaluation of Development Aid." *American Journal of Evaluation*. Vol. 29, No. 2, pp. 195-214.
- Doran, George. 1981. "There's a S.M.A.R.T. way to write management's goals and objectives." *Management Review (AMA FORUM)* 70 (11): 35–36.
- Independent Evaluation Group. 2012. *Designing a results framework for achieving results: a how-to guide*. World Bank: Washington, D.C.
- International Initiative for Impact Evaluation (3ie.) n.d. "Principles of Impact Evaluation." http://www.3ieimpact.org/media/filer_public/2014/01/15/principles_for_impact_evaluation.pdf
- Kusek, Jody Zall, and Ray C. Rist. 2004. *Ten steps to a results-based monitoring and evaluation system: a handbook for development practitioners*. Washington DC: World Bank.
- Michaelowa, K. and A. Bormann. 2006. "Evaluation Bias and Incentive Structures in Bi- and Multilateral Aid Agencies." *Review of Development Economics*, Vol.10, No. 2, pp. 313–329.
- Norad and Norwegian Ministry of Foreign Affairs. 2008. *Results Management in Norwegian Development Cooperation: A practical guide*. <http://www.norad.no/globalassets/import-2162015-80434-am/www.norad.no-ny/filarkiv/vedlegg-til-publikasjoner/results-management-in-norwegian-development-cooperation.pdf>
- Norad. 2012. *A Presentation of NORHED: The Norwegian Programme for Capacity Development in Higher Education and Research for Development*. Oslo.
- Norad. 2013a. *Guidelines for NORHED applications*. Oslo. <http://www.norad.no/globalassets/import-2162015-80434-am/www.norad.no-ny/filarkiv/4.-huf-2012---/2013-3101-norhed-guidelines-for-applications.docx>
- Norad. 2013b. *NORHED Main Application Form*. Oslo.
- Norad. 2014a. *Evaluation Series of NORHED: Theory of Change and Evaluation Methods*. Oslo. <http://norad.no/en/toolspublications/publications/2014/evaluation-series-of-norhed-theory-of-change-and-evaluation-methods/>
- Norad. 2014b. *NORHED Seminar - Managing for Results Capacity in Higher Education Institutions*. Presentation. Addis Ababa.
- Norad. 2015a. *Tender Document: Procurement of Consultancy Services for Annual Reporting for the NORHED Programme: Status and Aggregation of Project Results 2014*. Oslo.
- Norad. 2015b. *Terms of Reference: Consultancy Services for Performance of NORHED Annual Reporting 2014. (Annex 5 to Tender Document)*. Oslo.
- Norad. 2015c. *Template for NORHED Annual Progress Reporting. (Annex 7 to Tender Document)*. Oslo.

- Norwegian Ministry of Finance. 2003 (amended 2010). *Regulations on Financial Management in Central Government. Provisions on Financial Management in Central Government*. Oslo.
- Norwegian Ministry of Foreign Affairs. (2011). *Grant Scheme Rules for Support for Capacity Building within Research and Higher Education. Chapter/Item 165.70*. Oslo.
- Norwegian Ministry of Foreign Affairs. (2013). *Grant Scheme Rules for Support to International Organizations and Networks. Chapter post 160.75*. Oslo.
- OECD Development Assistance Committee (DAC). n.d. "Outline of Principles of Impact Evaluation." <http://www.oecd.org/dac/evaluation/dcdndep/37671602.pdf>
- Preston, Carolyn C. and Andrew M. Colman. 2000. "Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences." *Acta Psychologica*, Vol. 104, pp.1-15.
- Reinikka, R. and J. Svensson. 2004. "Local Capture: Evidence from a Central Government Transfer Program in Uganda." *The Quarterly Journal of Economics*. Vol. 119, No. 2, pp. 679-705.
- Rossi, Peter H., M. W. Lipsey, and H. E. Freeman. 2004. *Evaluation: A Systematic Approach*. 7th Edition. Sage Publications.
- Sasaki, R. September 2006. "A Review of the History and the Current Practice of Aid Evaluation." *Journal of Multi-Disciplinary Evaluation*, No. 5.
- Schiavo-Campo, Salvatore. 1999. "'Performance' in the Public Sector." *Asian Journal of Political Science* 7(2): 75–87.
- White, H. and Phillips, D. (2012). *Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework*. (Working Paper 15). London, U.K.: International Initiative for Impact Evaluation (3ie).
- White, Howard. (2009). *Theory-based impact evaluation: principles and practice*. (Working Paper 3). Washington, D.C.: International Initiative for Impact Evaluation (3ie).
- World Bank. 1992. *Effective Implementation: Key to Development Impact*. Report of Portfolio Management Task Force.
- World Bank. 2013. *Results Framework and M&E Guidance Note*.
- World Bank/ Inter-American Development Bank. 2002. *Peru Restoring Fiscal Discipline for Poverty Reduction: Public Expenditure Review*, Report No.24286-PE, Chapter 4, pp. 57-85.

Baseline study of NORHED Call-off A2

Terms of References (30th of September 2014)

1 Background

The Norwegian Programme for Capacity Development in Higher Education and Research for Development (NORHED) is operated by the Norwegian Agency for Development Cooperation (Norad) and has an annual budget of approximately 130 million NOK. The objective of NORHED is to contribute to development in the recipient countries by strengthening capacity in higher education institutions. In the programme document, this is operationalized as follows: “The purpose of the programme is to strengthen capacity in higher education institutions in LMIC [low and middle income countries] to contribute to a) a more and better qualified workforce, b) increased knowledge, c) evidence-based policy and decision-making and d) enhanced gender equality. The future NORHED impact is sustainable economic, social and environmental development in low and middle-income countries. (...) Strengthening of higher education institutions refers to: 1. Producing more and better research relevant to the identified areas/sub-programmes [and] 2. Producing more and better qualified graduates, men and women, relevant to the identified areas/sub-programmes.” (Norad, 2013:4).

In the first call-off, NORHED received 173 applications²⁸ for project support, out of which 46 will receive funding. The projects are owned by 26 higher education institutions from 16 low- and middle income countries, and 10 higher education institutions from Norway (Norad, 2013). The higher education institutions have been awarded project funding for 750 million NOK over five years. The first disbursements were made in December 2013 and the first projects started in January 2014. Most projects are situated in Uganda, Ethiopia, Malawi and South Sudan. NORHED has developed an overall results framework and all project applicants have been asked to develop independent results frameworks in accordance with this overarching framework. To aid this process, all recipients of NORHED funding were invited to participate in a workshop in Addis Ababa in March 2014. The workshop resulted in a set of standard indicators (see reporting template for standard indicators (Norad, 2014a). The majority of institutions were required to submit a revised version of this framework by the 31st of March 2014, a deadline that was later extended to the 11th of April. After this, the NORHED administration has been in dialogue with the project owners to ensure that the results frameworks adhere to NORHED’s overall results and monitoring framework. About half of the projects will hand in their first annual report in September 2014, and the first yearly project meetings will take place between October and November 2014. Many of these meetings are likely to take place at the South-led institutions.

Norad’s Evaluation Department has commissioned a real-time evaluation of NORHED. The evaluation consists of a series of studies to be conducted over the period 2014-2017 by the University of Southern California’s Development Portfolio Management Group. The purpose of the real-time evaluation is to enable learning within the field of higher education and development, so that future investments can become more effective in building capacity in higher education institutions as a contribution to California’s Development Portfolio Management Group.²⁹ The purpose of the real-time evaluation is to enable learning within the field of higher education and development, so that future investments can become more effective in building capacity in higher education institutions as a contribution to development. The evaluation programme will increase knowledge about the extent

²⁸ As stated in the TOR for the framework agreement: all applicants should have accepted to be part of an evaluation, including those who did not receive funding.

²⁹ Norad’s Evaluation Department has entered into a framework agreement with the University of Southern California’s Development Portfolio Management Group. The agreement is for 2 years, and Norad’s Evaluation Department has an option to renew the contract twice, for one year at a time.

to which NORHED builds capacity in higher education institutions, how capacity development of higher education institutions can be conducted most effectively and how higher education institutions affect development. This is the second call-off under the real-time evaluation.

2 Purpose and scope

The purpose of the assignment is dual: first to ensure that changes in institutional capacity (positive and negative) caused by the NORHED programme can be evaluated and plausibly attributed to the programme; and second to allow for learning from choices made early on in the programme period. This will be done through assessments of the existing monitoring and results framework and the selection mechanism for awarding funds, and by undertaking a baseline study of the current level of institutional capacity in institutions.

3 Objectives

3.1 PHASE 1: Inception

3.1.1 Assess the existing monitoring and results framework

In general, the evaluation team shall be specific about whether indicators and other data are at the level of input, output, outcomes and impact. The evaluation team shall respond to, but need not limit itself to, the following questions and topics:

a) Relevance of standard indicators: Using the general theory of change developed during call-off A1 as reference and taking into account costs associated with monitoring and evaluation; are the chosen standard indicators and the individual programme indicators developed by NORHED and the projects, appropriate for determining the effectiveness of the programme? Are any important indicators left out? If so, please be specific.

b) Operationalization of standard indicators: Is the operationalization of indicators sound? Is the definition of indicators specific enough to ensure that the same phenomena is measured across institutions? Are the operationalization of indicators likely to be biased?

c) Data quality: How will project managers go about to collect these data? Is it realistic that they can be collected as part of the routine monitoring/reporting procedures? Are they likely to be accurate?

d) Project-specific indicators: For project-specific indicators proposed by the individual NORHED projects, the usefulness of the indicators as such (not the operationalization and monitoring mechanisms) shall be assessed.

e) Need for additional data: Assess to what extent existing data (archives, registers, programme documents, etc.) can be used to undertake a descriptive analysis of the baseline situation and will allow for a future study of the effectiveness, or whether additional data need to be collected. The team shall clearly state if additional information is required and propose a method for data collection.

f) Update the evaluation team's database of NORHED-projects to take into account recent changes of the projects.

3.1.2 Describe the award mechanism

This part of the assignment shall include a description of the award mechanism³⁰ for the first NORHED call for proposals, including identification of award criteria, and mapping the process including relevant documents. The aim is to prepare for an assessment of the award mechanisms as such (3.2), and to make a recommendation regarding the use of non-grantees as a comparison group (3.1.3 b).

3.1.3 Propose detailed methods to determine effects

a) The evaluation team shall propose a detailed case-study design, including specifying methods, and discuss whether the units of analysis should be institutions or projects. See recommendation in the

³⁰ The process of selecting 46 grantees out of 173 applicants.

report “Theory of Change and Capacity Development in Higher Education Institutions” (call-off A-1) (p.38), to evaluate “a sample of projects that have been implementing for a sufficient number of months to allow progress and problems to surface”.

b) The evaluation team shall discuss which of the indicators that lend themselves to analysis using quasi-experimental design, and propose methods, including sample size strategy, data collection if required and costs. The sampling strategy should specify the control group, the number and level of cluster, units per cluster, the calculation of the intra-cluster correlation coefficient, and the minimum detectable effect of the outcome variables of interest for different powers. Please note that equations should be specified in an appendix.

c) Based on the assessment of available data, the evaluation team shall propose a detailed data collection plan for both qualitative and quantitative data, to ensure that effects can be documented and explained. In the report: *Theory of Change and Methods for Evaluating NORHED* (2014b), the evaluation team points out that the success of NORHED depends in part on system and university level enabling factors. The data collection strategy shall include a plan to detect whether such factors are likely to aid or impede NORHED funded projects from reaching their development objectives. In addition, the team shall discuss strategies for detecting adverse effects.

3.2 PHASE 2: Baseline

a) Collect additional data through open sources, interviews, surveys etc., as agreed during the inception phase.

b) Undertake a descriptive analysis of the baseline situation of the NORHED programme, taking the institutional context within which the institutions are operating into account. Unless otherwise is agreed during the inception phase, this analysis shall include the baseline situation for both institutions that applied for and were awarded funding and for institutions that applied for, but did not receive funding. The baseline analysis should include a discussion of factors that are likely to aid or impede NORHED funded projects from reaching their development objectives. The analysis may also highlight areas of particular interest for future call-offs.

c) Assess the award mechanism. This assessment shall include a discussion of the choices of projects and institutions selected to build capacity in light of the proposed general theory of change (Norad, 2014b) and other relevant literature on capacity building of higher education institutions. The assessment could include a comparison of types of projects and their contexts, available indicators and/or other factors for institutions that received and did not receive funding.

4 Fieldwork

During the inception phase the evaluation team should if possible, organize visit(s) to one or two of the annual meetings organised by the NORHED administration (likely to take place between mid-October and mid-November). Such visit(s) could combine observation of meetings, interviews, and on-site demonstration of procedures to collect indicator data in the institution hosting the meeting(s), to support the assessment of accuracy. Please note that any observation of closed meetings is subject to the consent of the organisers.

The inception phase will determine how later fieldwork should be conducted.

5 Team competencies

It is expected that methods such as interviews with the NORHED administration and other stakeholders, observation, literature and document reviews are the most important methods employed in this preparatory study. In addition, the team should have knowledge of quasi-experimental methods, as the preparatory study shall include a discussion of these methods, even though analysis, besides descriptive statistics, is not expected at this stage.

The team should at least cover the following competencies:

- Capacity building of higher education institutions

- Proficiency in qualitative methods, including observation, interviews and focus group discussions
- Proficiency in quasi-experimental methods
- Experience with conducting fieldwork in a low-income setting.

6 Deliverables

Deliverables will be organised in two phases corresponding to the organisation of objectives above.

6.1.1 PHASE 1: Inception

- a) A technical proposal, a work plan specifying who should do what and when (responding to these terms of references) and two separate budgets: one for the inception phase and one for the baseline phase. The technical proposal, work plan and budgets should be submitted by e-mail, at the latest by the 10th of October 2014, and are subject to the approval by the Evaluation Department.
- b) A draft inception report by the end of January 2015. The Evaluation Department will invite stakeholders to comment on this draft. The inception report shall include an assessment of the existing monitoring and results framework, a description of the award mechanism, and a detailed proposal for methods that can be utilised to assess the effects of NORHED, with respect to institutional strengthening.
- c) A final inception report including a work plan and budget for phase two. The inception report is subject to the approval by the Evaluation Department.

6.1.2 PHASE 2: Baseline (dates are tentative and may be revised during the inception phase)

- a) A draft report delivered no later than the end of May 2015. The draft report will be distributed to stakeholders for comments.
- b) Based on comments from the Evaluation Department and stakeholders, the evaluation team shall revise the draft and submit a final report no later than the end of August 2015. The report will be made publically available [format TBA] after approval by the Evaluation Department.
- c) Presentation of the baseline study at a seminar. Venue and date will be determined later. Two researchers who have been central in writing the baseline study shall attend.

Unless otherwise is agreed, the Evaluation Department's guidelines for report writing applies.

References

- NORAD 2013. A presentation of NORHED: The Norwegian Programme for Capacity Development in Higher Education and Research for Development. *In: DEPARTMENT FOR GLOBAL HEALTH, E. A. R. (ed.)*. Oslo Norad
- NORAD. 2013 *46 projects on higher education and research to receive funding* [Online]. www.norad.no. Available: <http://www.norad.no/en/support/norhed/46-projects-on-higher-education-and-research-to-receive-funding> [Accessed 11.12.2013 2013].
- NORAD 2014a. Narrative reporting template, P360 no. 1300798-25. Oslo Department for Global Health, Education and Research.
- NORAD 2014b. Theory of Change and Methods for Evaluating NORHED. *In: (DPMG), D. P. M. G. (ed.) Real-time evaluation of NORHED* Oslo: Norad's Evaluation Department

Amendment to the June 28, 2015, DPMG Technical Proposal for Completing the Evaluability Study of NORHED

A. Tasks

1. *Results framework*: How does the overall assessment of results frameworks change when we account for the choices and priorities established by Norad's results management practice?

1. We will analyze: a) implementation plans for activities and their causal connections to outputs; and b) the first year's annual narrative reports to determine the quality of data on baselines. This exercise will let us: a) revise the information status of grantees' results frameworks; and b) reassess the quality of the projects' "theories of change" and measurement arrangements when these additional sources of information are examined. Items 5, 11, and 12 in table 13 on page 24 of DPMG's Inception Report will be affected.

2. **Limitations.** The following limitations will have to be noted in the draft and final reports:

- Task 1 assumes that a complete set of *finalized and approved* annual narrative reports for the first year is available. If several annual reports have not been finalized, judgments about the evaluability of the NORHED program that require information from the annual reports will be based on partial information. The bias may or may not be serious, depending on the reasons that finalized documents are not available.
- Task 1.2 in the original proposal is deleted. This task would have assessed the extent to which the quality of the results frameworks predicts project implementation progress. Omitting it misses an opportunity to shed light on what seems to be a vigorous debate within Norad and within the broader Norwegian development community about the utility of results frameworks themselves--or the utility of standard results framework forms.

2. *Status of indicators*

3. Based on work conducted for the inception report and discussions with staff members of FORSK and the Results Management Section, DPMG will propose alternative ways of collecting data for the standard indicators. These proposals will be based on prior data collection experiences and survey best-practice for in-house computation of the final indicators that potentially will minimize respondent bias and error.

4. **Limitation.** The proposed indicators should better meet standards for homogenous reporting and aggregation. However, they will not have benefitted from pretests that usually improve the measurement quality of indicators. This issue will need to be noted in the draft and final reports.

3. *How well does FORSK's annual narrative report template support evaluations of the performance of projects and of the NORHED program?*

5. Attention to FORSK's supervision processes, including its annual narrative report template, was not included in the TOR for call-off A2. However, DPMG feels strongly that it must be added in order to arrive at a proper conclusion about the evaluability of the NORHED program. *In a small n evaluation situation, FORSK's supervision processes are the major source of information about performance and explanations of performance of NORHED*

projects and, ultimately, of the program itself. Given budget limitations for the exercise, DPMG will only evaluate the annual report template to determine if feasible revisions of the template can yield more complete and valid data for each project and thus for the overall program.

6. The template is already useful--it can be used to surface grantee's *de facto* redesign of the project (section 3.3) and to measure how well the project is progressing toward its intended outcomes (sections 3.3, 4, 5, and 7.1). However, the results of task 2 should be reflected in a redesign of the standard indicators being measured by the template, and other changes might enhance the quality and information yield of the annual report. For example, it might be possible to modify the template to enhance FORSK's and Norad's understanding of why a project is progressing or not progressing by adding measures (possibly one-time) of the system-level and institution-level enabling conditions affecting the project.

7. Limitations. At the Evaluation Department's request, we have omitted task 3.2. Using a sample of projects, this task would have assessed FORSK's supervision processes, as measured by the products listed in table 1 of the June 28, 2015, proposal. Omitting the analysis of FORSK's full supervision process will have to be noted in the draft and final reports as it has two consequences:

- First, since the supervision process bears the evidentiary burden for any small *n* impact evaluation of the NORHED program, Norad's Evaluation Department will have an incomplete picture of the evaluability of NORHED.
- Second, the Evaluation Department is engaged in a formative evaluation of NORHED. FORSK will lack the feedback that it needs to affirm the quality of its supervision process and/or to modify particular aspects of the process to increase its validity and reliability.

4. *Operationally practical ways to improve the quality of effectiveness and management evaluations*

8. Based on the results of tasks 1-3, the report will recommend ways to improve the quality of the performance monitoring and impact evaluation of the NORHED program. These recommendations will aim to be operationally practical, balancing scientific rigor with practical stakeholders and country constraints. Wherever possible, they will suggest when and how to collect data, what types of data to collect, standards for project progress reporting, and construction of performance monitoring indicators.

5. *With feasible changes, can the effectiveness and management of the NORHED program be adequately monitored and evaluated?*

9. Task 5 assumes that changes recommended in task 4 for improving NORHED's evaluability have been made. The report will conclude by assessing the evaluability of the NORHED program. Distinctions between levels of evaluation rigor may be introduced to show which evaluation questions probably can be answered and which probably cannot. For example, given the available data, will it be possible to determine if NORHED reached its intended objectives or to determine if NORHED affected the research productivity of its recipients?



Annual progress report for NORHED projects

Norwegian Agency for Development Cooperation
 Pb. 8034 Dep.
 NO-0030 OSLO
 Norway
 postmottak@norad.no

The report should be sent by e-mail to postmottak@norad.no, with copy to your desk officer. Please make reference to the agreement number in the subject field.

1. Project information	
Agreement title	
Agreement number	Total amount granted from Norad (NOK)

2. About the grant recipient		
2.1 Contact information of institution		
Name of institution and unit		
Telephone	Email	
Project coordinator	Email	Telephone

3. About the project		
3.1 Cooperating partner (s) please add rows if needed		
Name of institution and unit		
Project coordinator	Email	Telephone
3.2 Grant recipient's and cooperating partner's/partners' respective roles in the project		
Brief description of partnership to date including any challenges encountered, and how these challenges have been addressed and will be managed during the next period.		
3.3 Project implementation		
3.3.1. Brief description of any deviation between the approved application, including the project implementation plan, and the actual implementation of the project, and how these may have affected the project's results. Which measures have been undertaken in order to counteract possible delays or manage other deviations?		

3.3.2. Brief description of the management of identified risk factors, internal and external, including financial irregularities, so far in the project. Describe any new risk factors which have been identified, and how these will be managed during the next period.

3.3.3. Brief description of the project's achievements on gender equality, environmental issues, human rights and conflict sensitivity (if relevant).

3.3.4. Brief description of the project's efficiency (how efficiently resources/inputs are converted into outputs, i.e. value for money).

3.3.5 Assessment of the need for adjustments to work plans and/or inputs and outputs, including actions for risk mitigation.

3.3.6 Brief summary of the use of funds compared to the approved budget.

4. Reporting on standard indicators for the NORHED programme

Please note the importance of registering data including for the first reporting year, even if it may be too early to expect any direct results from the NORHED project for some indicators. Even reporting “none”/“0” or a low number is important since this provides a baseline for measuring project progress later on.

Standard indicators 1 and 2 should be reported together in the following table and below text box, and include achievements for all years to date:

4.1 Number of new/number of revised Bachelor/Master/PhD programmes/modules supported by NORHED

4.2 Number of Bachelor/Master/PhD programmes/modules supported by NORHED with gender perspectives included

Name of university hosting the programme/module	Degree type and name of programme/module (<i>please indicate whether programme or module and include level: Bachelor/Master/PhD</i>)	Please indicate if new or revised	Are gender perspectives included?	Year programme/module established/revised	Targets final year
			Yes/No		
			Yes/No		
			Yes/No		

Please add rows if needed.

Note: Please include all years to date. Please enter final year targets for each relevant partner institution in order to measure progress.

1. Is number of new/revised programmes/modules in line with the project plan and timeline? Please confirm or explain the deviations:

2. Please describe briefly for each new or updated programme or module:

- The new/revised items (including aspects not fully captured above, such as a programme becoming research-based or the mainstreaming of particular issues or pedagogical approaches at unit or at whole institution):

- How gender perspectives have been included:

4.3 Capacity to enroll and graduate students in NORHED supported programmes (Bachelor/Master/PhD)

Please note: The first year's reporting will be the baseline. If a new programme is being established or a current programme is being revised and has not resulted in enrollment at the time of reporting, you should report "0". The enrollment will be captured in next year's report or later.

Programme name <i>Please include level: Bachelor/Master/PhD</i>	Achievements reporting year (<i>first year's reporting = baseline</i>)												Total achievements including reporting year (<i>to be completed from second year's reporting onwards</i>)						Targets final year								
	Number of students enrolled during reporting year			Number of students enrolled with NORHED scholarships during reporting year			Number of students graduated during reporting year			Number of students graduated with NORHED scholarships during reporting year			Number of students graduated to date			Number of students graduated to date with NORHED scholarships			Number of students to graduate by year 5			Number of students to graduate with NORHED scholarships by year 5					
	F	M	Total	F	M	Total	F	M	Total	F	M	Total	F	M	Total	F	M	Total	F	M	Total	F	M	Total			

Please add rows if needed.

Note: From year 2 onwards, please also fill in the column on total achievements including reporting year

Please comment on:

1. How the NORHED project has supported the programme's capacity to enroll and graduate students (including other ways than through scholarships):
2. Student progression (enrolment vs. graduation rates), and whether progress is in line with project plan and timing, indicating any challenges and possible mitigating strategies:

4.4 Relevance of educational programmes and new graduates to local, national and regional needs and labor markets:

Are educational programmes and new graduates more relevant to local, national and regional needs and labor markets as a result of NORHED support? (Yes/No)

Please describe how relevance has been changed/increased in your project during the reporting year, with reference to sources where priorities and needs are defined (e.g. national development plans; university strategic plans; labour market needs) and if possible, evidence of increased uptake of graduates in the labor market:

4.5 Number of academic staff with strengthened qualifications (Master/PhD) by relevant institutional level (institute/department/faculty) supported by NORHED

Please note: You are asked to report these numbers for programmes supported by NORHED even if the project may not have had any effects on these numbers as of yet. If that is the case, please note it as a comment in the text box below the tables. The figures for year 1 will then be used as baseline for measuring project progress.

Name of university and unit (institute/department/faculty or other unit)	Baseline						Achievements reporting year									Targets final year											
	Number of staff at unit with Master degree prior to project startup			Number of staff at unit with PhD prior to project startup			Number of staff obtained Master degree during reporting year			Of which, number of NORHED scholarships			Number of staff obtained PhD degree during reporting year			Of which, number of NORHED fellowships			Number of staff to obtain Master degree by year 5			Number of staff to obtain PhD by year 5					
	F	M	Total	F	M	Total	F	M	Total	F	M	Total	F	M	Total	F	M	Total	F	M	Total	F	M	Total			

Please add rows if needed

Note: From year 2 onwards, a column will be added on total achievements including reporting year

Please comment on the above table (e.g. whether targets are being met, any challenges and how they are being addressed):

4.6 Ratio of qualified academic staff (Master/PhD) to students by relevant unit (institute/faculty/department) supported by NORHED
 (To be reported first and last year of project period only, as baseline and end-line).
 Please report all qualified staff in the NORHED supported unit, whether these have received direct NORHED support or not. The first year's reporting will constitute a baseline, and the purpose is to measure change by year 5 (and possibly beyond, if the project continues).

Name of university and unit (institute/department/faculty or other unit)	Baseline: Ratio of qualified staff (Master and PhD holders) to student numbers at relevant unit at project startup - <i>to be completed in the year 1 reporting</i>					Targets final year: Project target number for qualified staff (Master and PhD holders) to student numbers at relevant unit by year 5 - <i>to be completed in the year 1 reporting</i>				
	MA	PhD	Total (MA + PhD)	Student number at unit	Ratio: Total number (MA & PhD) to student number	MA	PhD	Total (MA + PhD)	Student number at unit	Ratio: Total number (MA & PhD) to student number
					:					:
					:					:

Please add rows if needed

Please note: In the final year's reporting, a column will be added on achievements for the final project year.

Please comment on the above figures, including whether any external changes have influenced staff/student ratio (e.g. large increases in student numbers, national regulations) :

4.7 Retention rates of qualified academic staff at relevant unit (institute/department/faculty) supported by NORHED

The retention rate of academic staff is influenced by many factors, also outside of the NORHED project’s control. This indicator is recorded in order to say something about retention of invested staff capacity building versus brain drain, which is a common critical comment to capacity building in general. You are encouraged to add an analysis of this issue in the text box below the table.

Name of university and unit (institute/department/faculty or other unit)	Baseline - Retention rate at project startup								Achievements reporting year - Retention rate of academic staff								Project target for retention rate						
	Number of qualified staff (Master/PhD holders) who left unit out of total number of qualified staff employed at unit during the year prior to project startup								Number of qualified staff (Master/PhD holders) who left unit out of total number of qualified staff employed at unit during reporting year								Target for final year retention rate of qualified staff (Master/PhD holders)– to be completed in the year 1 reporting						
	MA hired		MA left		PhD hired		PhD left		MA hired		MA left		PhD hired		PhD left		MA hired		MA left		PhD hired		PhD left
F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M

Please add rows if needed

Note: From year 2 onwards, a column will be added on total achievements including reporting year.

Please keep records of staff who receive direct support from NORHED to upgrade their academic qualifications, but leave the unit during the project period. This will enable you in the final year of NORHED project support to report on the average retention rate of all academic staff compared with the average retention rate of academic staff who received their direct support from NORHED to upgrade their degrees.

Please comment on:

1. Staff retention rates, including whether reporting period was representative or exceptional regarding staff retention, and suggested mitigating measures:
2. Why staff left the institution, and where they were subsequently employed. Please specify as much as possible, using numbers where relevant:

Please note: This indicator also requires you to establish tracer mechanisms for staff who have received NORHED scholarships/fellowships. Please complete the below table for NORHED scholarship/fellowship holders from project onset and keep updated so records are available for future NORHED evaluations.

TRACING NORHED SCHOLARSHIPS/FELLOWSHIP HOLDERS:								
Institution and unit (institute/department/faculty)	Name of scholarship/fellowship holder	Gender (F/M)	Degree type and title	Graduation year	Employment year at institution	If left institution, when?	New employer type (public/private university/NGO/UN etc) and country	Comments

Please add rows if needed

4.8 Number of scientific publications (peer reviewed and others) by NORHED project to date										
Publication year (please also note if "in press" and include all publications to date).	No. of publications in peer reviewed journals	How many male/female authors		Of these, how many joint publications between LMIC and Norwegian researchers	Target number of peer reviewed publications by year 5	No of other scientific publications (including all publications to date)	Of these, how many male/female authors		Of these, how many joint publications between LMIC and Norwegian researchers	Target number of other scientific publications by year 5
		F	M				F	M		

Please add rows if needed

4.9 Number and type of other dissemination activities to date				
Year of activity (please include all to date)	Type and topic of dissemination activity related to NORHED projects (e.g. hosting academic conferences, media coverage, stakeholder workshops, project website...)	Number	Targets final year	Briefly describe the activity, e.g. the results, contributing institutions, the gender balance of the main contributors, and the number of participants if relevant

Please add rows if needed

4.10 Uptake/influence of NORHED-supported research in public policies

To be reported the year when uptake happens. Please also note if it is too early to expect such uptake or if uptake has not happened. In the final year report, please add a summary of all achievements during project period.

Please describe if and how NORHED supported research in your project has been included in or influenced public policies (laws, regulations, standards, services, practices etc.)

- Please list examples of public policies and how they were influenced by project research:
- Please comment on whether this in line with project plan and timeline:

4.11 Uptake/influence of NORHED-supported research findings/new technologies/innovations/solutions by local communities/civil society/private sector

To be reported the year when uptake happens. Please also note if it is too early to expect such uptake or if uptake has not happened. In the final year report, please add a summary of all achievements during project period.

Please describe if and how NORHED supported research findings/new technologies/solutions in your project have been taken up by local communities, civil society, and/or the private sector.

- Please list examples of findings and how they have been taken up by relevant stakeholder:
- Please comment on whether this in line with project plan and timeline:

4.12 Knowledge transfers within South-South and South-North networks and partnerships as a result of the NORHED project

In the final year report, please add a summary of all achievements during project period.

Please describe whether and how knowledge transfers within South-South and South-North networks and partnerships have taken place, and if so, how they have contributed to strengthened capacity for education and research (e.g. updated curricula; new teaching methods; international examination committees...):

4.13 Changes in the broader institutional environment at NORHED supported institute/faculty/department which strengthened the capacity education and research

In the final year report, please add a summary of all achievements during project period.

Please describe changes in the broader institutional environment as a result of NORHED support, and how it has affected the capacity for education and research (e.g. institutional policies and systems at the university):

4.14 Access to libraries, laboratories and ICT for staff and students in NORHED supported institutes/departments/faculties

Has your NORHED project led to increased staff and/or student access to:

For staff		For students	
Yes	No	Yes	No

Libraries (please specify):

Laboratories (please specify):

ICT equipment (please specify):

Other (please specify):

Please add rows if needed

Please describe whether and how upgraded ICT, library and/or laboratory infrastructure supported by NORHED has influenced education and research capacity and quality at the institution, referring to relevant evidence where available (e.g. staff and student satisfaction and retention, institutional reputation, pedagogical innovation):
(Outputs should be reported under section 5).

5. Results report

Intended impact on society:

Please report on the remaining indicators specific to your project which have not been covered by the above reporting, according to the approved results framework for your project. Please use gender disaggregated data where relevant. If more than one LMIC is involved in the project, please indicate at which institution the achievement has happened.

	Indicator	Targets reporting year (if relevant)	Achievements reporting year	Total achievements including reporting year	Targets final year	Comments

Outcome 1:						
Output 1.1:						
Activity 1.1.1.:						
Outcome 2:						
Output 2.1. :						
Activity 2.1.1.:						
Outcome 3:						
Output 3.1:						
Activity 3.1.1.:						

6. Financial overview

The detailed financial report must be provided in a separate attachment. The financial report should include project accounts with related explanations, and must be presented according to the same structure and elements as in the approved, detailed budget. The financial report must be confirmed by the person being responsible for financial matters in the grant recipient's organization.

	1	2	3	4	5	5
	Approved total budget	Total expenditure until the reporting period	Approved budget for the financial year	Total expenditure for the financial year	Variance (3) – (4)	Variance between (3) and (4) in %
	(2013-2018)	(state period)	(state period)	(state period)		
Partner 1						
Partner 2						
Partner 3 (if relevant)						
Total						

6.1 Financial contribution from other sources (if relevant)

--

7. Additional information
7.1. Case/success story
Please give a short description of a positive result (on any level of the results chain) which the project has experienced in the reporting period. The case should include a short description of the activity, a description of what was achieved and how this relates to the planned outcome(s).
7.2. Any other information of relevance for the report

8. Attachments		
Tick the boxes below if attached, and give each attachment a number. Any other attachment should also be listed		
Attached	Number	Attachment
<input type="checkbox"/>		Certified annual financial statement for the project covering the previous financial year, using the NORHED financial reporting template and including all attachments as described in Article VI, Clause 4 of Agreement (mandatory)
<input type="checkbox"/>		Protocols for procurements and/or disposals effectuated during the reporting period (if relevant)
<input type="checkbox"/>		List of PhD/ Post Doc candidates with brief information on topic, research question and name of supervisor and institution (if relevant)
<input type="checkbox"/>		List of students/staff receiving scholarships funded by the project (if relevant)
<input type="checkbox"/>		Revised annual project implementation plan (work plan) and budget for current financial year (if relevant)
<input type="checkbox"/>		

9. Date and confirmation	
I am authorized to sign legally binding agreements on behalf of the grant recipient, and confirm that the information in this report is correct to the best of my knowledge.	
Place and date	Name and signature