"Coaching School Leadership to achieve high level learning outcomes"

Outcomes Evaluation Report

March 2015

"Coaching School Leadership to achieve high level learning outcomes"

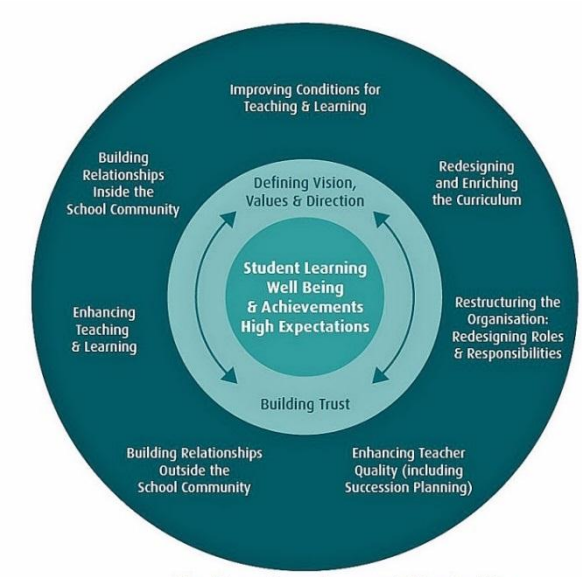
Outcomes Evaluation Report

Contents

§ 1 Introduction

There is robust empirical evidence that learning achievement of pupils is better in schools led by a well performing school leader. This evidence however is stemming from descriptive research in western countries. Research findings also indicate that school leaders are performing better if they are being coached. But again, these findings come from research in western society.

Does coaching contribute to the improvement of school leaders' performance in Rwanda? Is improved school leaders' performance contributing to pupils' learning achievement in Rwanda? If so, is it possible to design cost-effective approaches for implementation in Rwanda?



The dimensions of successful leadership

In the "**Coaching School Leadership to achieve high level learning outcomes**" project action research was conducted in order to come with answers to these questions. Samples of Head Teachers and their Sector Education Officers were familiarized with theoretical knowledge on successful school leadership[1]. Using a contextualized self-assessment tool, based on the 'eight dimensions of successful school leadership', Head Teachers identified their own strengths and weaknesses. They then formulated 'good intentions' and a plan of action to actually achieve them. In the meantime Sector Education Officers were trained in coaching and moderating intervision processes, and were facilitated to practice these with the Head Teachers of their Sector. The objective was of course to effectively support this Head Teachers in their pursuit of achieving their 'good intentions', leading towards improved performance in school leadership and increased learning achievement of pupils. All of this was embedded in a semi-experimental design with conditions for generating 'evidence-based' knowledge. Working with four samples in which different types of interventions were used also gave the opportunity to compare the level of impact of the distinct approaches and the related costs, thus making it possible to come up with statements on cost-effectiveness.

This report gives an overview of the project's outcomes evaluation. Paragraph 2 provides an executive summary. Paragraph 3 describes the design and underlying theory of the action research, the questions formulated and the indicators used. Paragraph 4 describes the methodology used: the way groups were sampled, instruments were developed, data were collected and analysis was set up. In paragraph 5 different levels of results are discussed. In the sixth and last paragraph conclusions and recommendations are formulated.

§ 2 Executive summary

In this study, we followed a group of approximately 140 Head Teachers of schools for Basic Education in Rwanda. The Head teachers participated in a professional development trajectory on school leadership, and were randomly assigned to four experimental groups that received (A) training and coaching; (B) coaching only (C) training only or (D) no training or coaching at all. Next to the research question if professional development trajectories like these have impact the other question was which of the above described approaches is most cost-effective.

[1] The theory offered was based on the research from Day, C., Sammons, P., Leithwood, K., Hopkins, D., Harris, A., Gu, Q. & Brown, E. (2010) as it was summarized in *Ten Strong Claims about Successful School Leadership.* Nottingham, UK: University of Nottingham/National College for School Leadership.

Head teachers have a crucial impact on the quality of education. Through influence on their staff, effective school leaders improve teaching and learning, and thus indirectly learning outcomes of pupils. The project is based on this 'domino-effect': the 'first stones' are being pushed in the sense that Sector Education Officers (SEO's) and Head Teachers (HT's) are familiarized with the latest theory on effective school leadership and SEO's are trained to coach HT's to become more effective School Leaders. The ultimate objective is to improve pupils' learning achievement. Since the available time in this project was too short to expect results at the level of the last stone in the domino (at the level of pupils' learning achievement), a number of indicators were formulated at the level of the other levels in the domino. Summarizing intermediate effects at the level of SEO's, HT's, Teachers and pupils were described and affiliated indicators were formulated. Furthermore per level the research methodologies are described and the outcomes of the distinct research were discussed and interpreted.

As was expected at the level of primary school pupils no evidence was found that the exam results of schools led by Head Teachers from experimental groups differ from the control group.

For pupils' learning conceptions, although there were some significant differences found between the different experimental groups, those were difficult to interpret since there was little consistency in these differences. Pupils in schools with HT from group A are more in agreement with a development orientation and with external regulation than the other experimental groups and the control group D. Further research is needed to find out whether this combination of subscales is actually predicting higher learning achievements of pupils.

At the level of teachers significant differences were found in the use of data between the different experimental groups: teachers with HT's of group C use data slightly more often than teachers of HT of group A and B and especially more often than the teachers with HT in the control group D. The differences between group C and A are very small.

At the level of Head Teachers their use of data in 2015 differs significantly over the different experimental groups, with head teachers of group A using data more often than head teachers of group B and C, and especially more often than head teachers of the control group D.

When looking at the assessment of Head Teachers measured against the indicators of the eight dimensions of effective school leadership significant differences were found between the baseline and the end line and between the different groups. Remarkable is that while the differences between the end line overall scores of the groups only are very small, the difference between baseline and end line for group A and B is much bigger than for the other groups. Although there is no good explanation for the differences between the groups in the baseline, at least it could be assumed the measured growth of performance of Head Teachers could be attributed to the approach used in A and in a lesser extent in B.

Taking into account the level of SEO's, knowing both SEO's of group A and B were trained together and no differences were observed in their coaching performance while conducting their intervision sessions at Sector level, it is not plausible that differences between Head Teacher performance can be attributed to differences between SEO's.

Bringing findings together at an overall level it is stated that, apart from the learning outcomes in exams, the experimental groups differ significant from the control group. There are however no clear, unambiguous and consistent findings yet that could legitimize statements about the ultimate best approach. At the level of Head Teachers' performance, the growth in performance of Head Teachers in group A is best, followed by those in group B. The differences in measured performance at the end line between group A and B however are significant, but small.

Knowing this turning to the level of costs became important for answering the cost-effectiveness question. As was found the costs per pupil of the distinct approaches were respectively £2,07 for group A, £1,25 for group B and £1,35 for group C. These cost-calculations made clear that the approach in which only Sector Educational Officers were trained to guide and coach intervision

meetings with their Head Teachers and supported to organize these meeting on a monthly basis was the least expensive. Since we found at result level no decisive evidence legitimizing other considerations for deciding for group A or B, it was concluded approach B is most cost effective.

In accordance with this conclusion two different sets of recommendations are given. The first set of recommendations are recommendations for further research. The other set of recommendations are policy recommendations related to the existing policies and strategies on school leadership in Rwanda.

Summarizing the following research recommendations are given:

1. Knowing the project is based on the assumption that the impact of effective school leadership works via a 'domino'-effect, it is recommended to extend the project over a number of years, so that in due course the effect on pupils' learning achievement can be measured;
2. In order to be able to come up with more reliable data it is recommended to further elaborate some of the indicators and develop more adequate tools for measuring;
3. In order to be able to come up with more underpinned interpretation of data it is recommended to also systematically gather more qualitative data that could make this possible.

In addition to the recommendations for further research the following policy-recommendations are given:

4. Given the policies as formulated in the ESSP 13/14 – 17/18 and the TDM policy document, in which the importance of an initial training for Head Teachers, complemented by a system for Continuous Professional Development is emphasized, it is recommended to use the approach in which SEO's are trained to coach Head Teachers (approach B) as the most cost-effective way of working;
5. In case up scaling of the project is considered it is recommended to replace the '8 dimensions' and its affiliated tools with the '5 standards for effective school leadership' and its tools;
6. In case up scaling of the project is considered it is recommended to incorporate REB's-VVOB pilot project[2] and to further elaborate it as a system for CPD for Head Teachers in line with TDMs policy target 5.1;
7. In case up scaling of the project is considered it is recommended to build on existing structures and to involve all stakeholders from the outset in the design, implementation and maintenance of this system of continuous professional development for School Leaders.

§ 3 Purposes

As mentioned in the introduction there is descriptive empirical research stating that the quality of school leadership is (after the quality of teaching) the second important school related factor having impact on pupils' learning achievements. Also was mentioned that there are research findings indicating that coaching improves the quality of school leadership. In the "**Coaching School Leadership to achieve high level learning outcomes**" project action research was conducted to find out if and how different approaches, in which different combinations of training and coaching were used, worked out in the daily reality of primary education[3] in Rwanda. The design of the project is represented in the undergoing table.

[2] As can be read in §6 REB in collaboration with VVOB started a pilot project in which SEO's are trained to coach HT's in Professional Learning Networks.

[3] School leaders that participated in this project all were at least responsible for primary education. Since the reality of 9 years and 12 years basic education many of them however also were responsible for lower and often also for upper secondary education at their school.

Project design "**Coaching School Leadership to achieve high level learning outcomes**"		
	Head Teacher is trained	Head Teacher is not trained
Sector Education Officer is trained and coaches the Head Teachers of her/his Sector	Group A: Both Head Teacher and SEO's are familiarized with theory of effective school leadership. SEO's are trained to coach her/his (group) of Head Teachers. Based on a self-assessment Head Teachers formulated resolutions to improve their performance and a plan of action to realize their resolutions.	Group B: SEO's are familiarized with theory of effective school leadership. SEO's are trained to coach her/his (group) of Head Teachers.
Sector Education Officer is not trained and does not provide coaching to the Head Teachers of her/his Sector	Group C: Head Teachers are familiarized with theory of effective school leadership. Based on a self-assessment Head Teachers formulated resolutions to improve their performance and a plan of action to realize their resolutions.	Group D: 8 Sectors being control group

Table 1 project design

By using this design the questions that were sought to be answered were about the effectiveness of the different approaches and their respective cost-effectiveness. Underlying assumption is that the quality of School Leadership is improved by training and/or coaching and is leading to changes in the school culture, impacting the quality of teachers which has impact on learning conceptions of pupils and ultimately on pupils' learning achievement. And this was actually what we aimed for. By pushing the first and/or second 'domino stone' we want to cause this domino effect on the last stone in the row, the learning achievement stone.

Knowing that the project time of 28 months (including starting up and phasing out time) would be too short to expect impact at the level of pupils' learning outcomes (if all stones of the domino would fall), indicators have been formulated at the different levels that are involved.

The Specific objective of the project was to generate evidence-based information on cost-effective and scalable capacity development approaches for improving School Leaders' competencies in educational leadership, leading to a

significant improvement in primary school pupils' learning outcomes.

The expected results were:

1. Three different capacity development approaches for improving School Leaders' competencies in educational leadership have been implemented.
2. The effects of the three capacity development approaches have been documented, monitored and evaluated, at the different levels in the Theory of Change (ToC).
3. The costs of the three capacity development approaches have been measured, allowing an estimation of cost-effectiveness.

The major focus was on the effectiveness of the innovation in terms of (learning) outcomes. Since three different capacity development approaches were actually applied and tested, several options for the innovation exist. The innovation option with the biggest impact on (learning) outcomes will not necessarily be the most cost effective as well, as this will also depend on the level of inputs (in terms of time and resources) required to achieve that impact. The most cost effective capacity development approach in terms of learning outcomes will then be proposed as the innovation.

The table below gives an overview of the indicators at the different levels.

	Indicators	Baselines	Targets	Sources of evidence
At the level of primary school pupils				
1	Results in the Primary Leaving National Examination	Results per group of 8 sectors in 2012	Significant statistical differences to reject[4] the hypothesis within (or more probable after) the project period	Rwanda National Examinations Council (RNEC) results per group of 8 sectors in 2012, 2013, 2014
2	Attendance figures	Figures 2012 per 8 sectors	Significant statistical differences to falsify the hypothesis	Figures 2012, 2013, 2014 per 8 sectors; figures coming from the school records
3	Dropout rate	Figures 2012 per 8 sectors	Significant statistical differences to falsify the hypothesis	Figures 2012, 2013, 2014 per 8 sectors figures from the Rwanda Education Statistics report provided by MINEDUC
4	Motivation and learning responsibility	Baseline results January 2013	Observable improvement of motivation and learning responsibility in terms of significant statistical differences to falsify the hypothesis	Baseline study and End-of-project evaluation. Measuring tool coming from the Research program "Impact School leadership on Pupil Achievement"
At the level of primary school teachers				
5	Academic standards and expectations	Baseline survey January	Observable improvement of standards and expectations in terms of	Baseline study, Progress reviews, End-of-project evaluation;

[4] The hypothesis is that there will be no significant statistical differences between the 'experimental' and 'control' group.

	Indicators	Baselines	Targets	Sources of evidence
		2013	significant statistical differences to falsify the hypothesis	Measuring tool coming from the Research program "Impact School leadership on Pupil Achievement"
6	Use of assessment data in the teaching and learning process	Baseline survey January 2013	Observable improvement of teaching and learning process in terms of significant statistical differences to falsify the hypothesis	Baseline study, Progress reviews, End-of-project evaluation; Measuring tool coming from the Research program "Impact School leadership on Pupil Achievement"
7	Attendance & turnover rate	Figures 2012 per 8 sectors	Significant statistical differences to falsify the hypothesis	Figures 2012, 2013, 2014 per 8 sectors figures coming from the school records,
At the level of primary school Head Teachers (HTs)				
8	Progress indicators on the 8 key dimensions[5] of successful school leadership	Baseline values 2013	Relevant progress on at least 1 dimension after 1 year; and on at least 3 dimensions after 2 years Significant statistical differences to falsify the hypothesis.	Baseline study: 360° feedback form for Head Teachers, completed by Teachers, PTC member, SEO to measure. A form in which the 8 key competencies are made measurable in a reliable way in terms of practices to be used. In the process of on-going self-evaluation a.o. during intervision, this tool will also be used as developmental tool;
At the level of the Sector Education Officers (SEOs)				
9	SEO's demonstrate responsiveness to the needs of the learning group and are able to adjust interventions to these needs.	SEO's have insufficient intervision coaching skills	SEO's are applying at least 50% of relevant coaching competencies	Reports on these indicators from their own (trainer-) coaches on the basis of visiting coaching sessions and providing feedback on that. A standardized list of key-competencies for coaching will be used to score and report.

[5] This is the level where we expect the most direct impact of the project and monitoring and evaluation will be done at a higher level of detail at this level. Sub-indicators, in terms of observable practices (behaviour) will therefore be developed for each of the key dimensions of school leadership.

	Indicators	Baselines	Targets	Sources of evidence
At the level of the school leadership facilitators				
10	Capability of school leadership facilitators in training HTs and SEO's on school leadership.	School leadership facilitators have limited knowledge on the concept of 'educational leadership'	At least 10 school leadership facilitators are able to train HTs and SEO's in educational leadership	Report on basis of observations of trainers of trainers, who visit and participate in the training sessions. This will be complemented with evaluation reports completed by the attendants.
At the level of the Rwanda Education Board (REB)				
11	Availability of evidence-based information on capacity development approaches for school leadership	No evidence-based information is available	A clear set of recommendations is available, shared and owned by REB by the end of the project period.	Report with set of recommendations. Whether these recommendations are owned and shared by REB can only appear from plans of action for up scaling the project from their side. The latter depends on a lot of other factors like available budget, intra- and inter-ministerial priorities, etc.
At the level of costs				
12	Availability of costs related to the distinct capacity development approaches for school leadership	No information is available	A clear body of knowledge is available on costs and cost effectiveness related to the distinct capacity development approaches for school leadership	Report on the costs of the distinct capacity development approaches for school leadership
At the level of the innovation process				
13	Availability of information on the innovation process	No information on the innovation process is available	Recommendations on implementing the innovation are available.	Process reports coming from the monitoring/QA process as described here under.

Table 2 Indicators, targets and sources of evidence

§ 4 Methodology

In order to give a thorough and systematic overview of the methodologies used this will be described at the same levels as used in table 2. Each group of Sectors (Groups A to D) was sampled at random from the total population of 416 Sectors in Rwanda. From this list then 3 Sectors in Gasano District, where Wellspring Foundation is working on a school development project had to be replaced for

attribution reasons, with three newly generated random numbers. Further details can be found in annex 1.

	Indicators	Baselines	Targets	Sources of evidence
At the level of primary school pupils				
1	Results in the Primary Leaving National Examination	Results per group of 8 sectors in 2012	Significant statistical differences to falsify the hypothesis within (or more probable after) the project period	Rwanda National Examinations Council (RNEC) results per group of 8 sectors in 2012, 2013, 2014
➢ Data on examination results are collected at sector level and allows us to identify trends in examination results and compare over sectors (and therefore over experimental groups); ➢ Available data are based on examination results and indicate per sector the distribution of pupils in divisions (from 1 to 5, where 5 is "not qualified"). The score per group is calculated as the percentage of the maximal possible score. ➢ If there are significant differences between groups, the improvement of the distinct experimental groups over 2012-2013-2014 is measured and compared;				
2	Attendance figures	Figures 2012 per 8 sectors	Significant statistical differences to falsify the hypothesis	Figures 2012, 2013, 2014 per 8 sectors; figures coming from the school records (checks on sample basis by SEO's).
➢ Data on attendance figures turned out to be difficult to get our hands on and available data are questionable in terms of reliability. Therefore this indicator was dropped.				
3	Dropout rate	Figures 2012 per 8 sectors	Significant statistical differences to falsify the hypothesis	Figures 2012, 2013, 2014 per 8 sectors figures from the Rwanda Education Statistics report provided by MINEDUC
➢ Data on dropout rate turned out to be difficult to get our hands on and available data are questionable in terms of reliability. Therefore this indicator was dropped.				
4	Motivation and learning responsibility	Baseline results January 2013	Observable improvement of motivation and learning responsibility in terms of significant statistical differences to falsify the hypothesis	Baseline study and End-of-project evaluation. Measuring tool coming from the Research program "Impact School leadership on Pupil Achievement"
➢ The tool coming from the original research program wasn't availed by the National College for School Leadership. Alternatively a self-report questionnaire, as developed by Prof. Dr. P. Van				

	Indicators	Baselines	Targets	Sources of evidence
	Petegem on learning conceptions was used. So data on learning conceptions of pupils in primary education are collected with this self-report questionnaire. Even though some of the scales in the questionnaire are less reliable, for some subscales that are more reliable, relevant analysis can be made[6]; ➢ Mean scores will be measured on 5 sub-scales: 'Development orientation', 'Negative orientation', 'External regulation', 'Individual regulation' and 'Serious learning'. Two sub-scales were skipped since they turned out not to be reliable;			
At the level of primary school teachers				
5	Academic standards and expectations	Baseline survey January 2013	Observable improvement of standards and expectations in terms of significant statistical differences to falsify the hypothesis	Baseline study, Progress reviews, End-of-project evaluation; Measuring tool coming from the Research program "Impact School leadership on Pupil Achievement"
➢ The tool coming from the original research program wasn't availed by the National College for School Leadership. Alternatively a self-report questionnaire, as developed by Prof. Dr. P. Van Petegem on school teachers' academic standards and expectations was used. So data on primary school teachers' academic standards and expectations are collected with this self-report questionnaire on their perceptions on their lessons, their school and themselves as a teacher (towards their pupils' capacities). Unfortunately, none of these subscales has an internal consistency of .70 or above (Cronbach's alpha).				
6	Use of assessment data in the teaching and learning process	Baseline survey January 2013	Observable improvement of teaching and learning process in terms of significant statistical differences to falsify the hypothesis	Baseline study, Progress reviews, End-of-project evaluation; Measuring tool coming from the Research program "Impact School leadership on Pupil Achievement"
➢ The tool coming from the original research program wasn't availed by the National College for School Leadership. Alternatively a self-report questionnaire, as developed by Prof. Dr. P. Van Petegem on primary school teachers' use of assessment data was used. The scale on perceived use of assessment data is a reliable measurement instrument with a Cronbach's alpha of .892.				
7	Attendance & turnover rate	Figures 2012 per 8 sectors	Significant statistical differences to falsify the hypothesis	Figures 2012, 2013, 2014 per 8 sectors figures coming from the school records, (checks on sample basis by SEO's)

[6] More details in annex 2

	Indicators	Baselines	Targets	Sources of evidence
➢ Data on attendance and turnover rates are difficult to get our hands on and available data are questionable in terms of reliability. Therefore this indicator will be dropped.				
At the level of primary school Head Teachers (HTs)				
8	Progress indicators on the 8 key dimensions of successful school leadership	Baseline values	Relevant progress on at least 1 dimension after 1 year; and on at least 3 dimensions after 2 years Significant statistical differences to falsify the hypothesis.	Baseline study: 360° feedback form for Head Teachers, completed by Teachers, PTC member, SEO to measure. A form in which the 8 key competencies are made measurable in a reliable way in terms of practices to be used. In the process of on-going self-evaluation a.o. during intervision, this tool will also be used as developmental tool;
➢ Data on 8 key dimensions of successful school leadership is collected through the 360 tool in which head teachers (self-assessment), teachers, PTC members and SEOs evaluate 8 key competencies of head teachers. Analysis of head teachers' self-assessment data showed that all 8 subscales are reliable, with Cronbach's Alpha values in between .823 and .906. ➢ Head Teachers are assessed on a scale from 1 (totally disagree) to 9 (totally agree) on the way they: o define their values and vision to raise expectations, set direction and build trust o reshape the conditions for teaching and learning o restructure parts of the organisation and redesign leadership roles and responsibilities o enrich the curriculum o enhance teacher quality o enhance the quality of teaching and learning o build collaboration internally o build strong relationships outside the school community Assessment per dimension was done based on a number of indicators as developed by a forum of Rwandan education experts. The average of the scores on the distinct dimensions is the overall score.				
8a	Use of data by Head Teachers in the process of informed decision making	Baseline survey 2013	Observable improvement in terms of significant statistical differences to falsify the hypothesis	Baseline study, End-of-project evaluation;
A self-report questionnaire, as developed by Dr. P. Van Petegem on primary Head Teachers' use of data was used. The scale on perceived use of data is a reliable measurement instrument with a Cronbach's alpha of .927.				
At the level of the Sector Education Officers (SEOs)				
9	SEO's demonstrate responsiveness to the	SEO's have insufficient	SEO's are applying at least 50% of	Reports on these indicators from their own (trainer-) coaches on the

	Indicators	Baselines	Targets	Sources of evidence
	needs of the learning group and are able to adjust interventions to these needs.	intervision coaching skills	relevant coaching competencies	basis of visiting coaching sessions and providing feedback on that. A standardized list of key-competencies for coaching will be used to score and report.
➢ Data on SEOs responsiveness to the needs of the learning groups are collected through reports of feedback visits of intervision sessions. Qualitative analysis focuses on the coaching techniques of SEOs and the dialogue between participants of the sessions, as well as the participation of participants in general and the atmosphere of the sessions. ➢ For end evaluation all reports of the five rounds of feedback visits to intervision sessions will be analysed comparing individual performance in different rounds and comparing rounds as a whole.				
At the level of the school leadership facilitators				
10	Capability of school leadership facilitators in training HTs and SEO's on school leadership.	School leadership facilitators have limited knowledge on the concept of 'educational leadership'	At least 10 school leadership facilitators are able to train HTs and SEO's in educational leadership	Report on basis of observations of trainers of trainers, who visit and participate in the training sessions. This will be complemented with evaluation reports completed by the attendants.
Data on capability of school leadership facilitator in training HTs and SEOs is collected through workshop evaluation sheets of reflection workshops.				
At the level of the Rwanda Education Board (REB)				
11	Availability of evidence-based information on capacity development approaches for school leadership	No evidence-based information is available	A clear set of recommendations is available, shared and owned by REB by the end of the project period.	Report with set of recommendations. Whether these recommendations are owned and shared by REB can only appear from plans of action for up scaling the project from their side. The latter depends on a lot of other factors like available budget, intra- and inter-ministerial priorities, etc.
Since the source of evidence can only be accessed after finalizing the project this will not be part of the evaluation of the project.				
At the level of costs				
12	Availability of costs related to the distinct capacity development approaches for school leadership	No information is available	A clear body of knowledge is available on costs and cost effectiveness related to the distinct capacity development approaches for	Report on the costs of the distinct capacity development approaches for school leadership

	Indicators	Baselines	Targets	Sources of evidence
			school leadership	
Data will be gathered about costs of the three distinct capacity development approaches and measured in terms of staff time and budget. For reasons of scale-advantages[7] however, several activities related to different approaches were conducted in combination with each other. In order to come to a realistic comparison of the costs of the different approaches, calculated costs will be extrapolated to a situation in which activities are conducted for 2* 8 Sectors. Included will be all costs (HR, logistics, activities, capital investments, and office costs, overhead) that are related to the implementation of the project. Cost related to doing research will be excluded. Since the ultimate objective of the project is improving pupil learning achievement, the costs will be expressed in costs per pupil.				
At the level of the innovation process				
13	Availability of information on the innovation process	No information on the innovation process is available	Recommendations on implementing the innovation are available.	Process reports coming from the monitoring/QA process as described here under.
At this level a separate report on the Process of Innovation will be submitted.				

§ 5 Results

At the level of the different indicators the results from baseline and end line have been compared and related to expectations as formulated at the start of the project.

	Indicators	Baselines	Targets	Sources of evidence
At the level of primary school pupils				
1	Results in the Primary Leaving National Examination	Results per group of 8 sectors in 2012	Significant statistical differences to reject the hypothesis within (or more probable after) the project period	Rwanda National Examinations Council (RNEC) results per group of 8 sectors in 2012, 2013, 2014

[7] Training of SEO's was done with the SEO's of group A and B together. It is obvious that training 16 SEO's together has financial advantages above training them apart.

<table>
<tr><th></th><th>Indicators</th><th>Baselines</th><th>Targets</th><th>Sources of evidence</th></tr>
<tr><td colspan="5">All exam results of 2012, 2013 and 2014 were collected per sector. Those results indicate per sector the distribution of pupils in divisions (from 1 to 5, where 5 is “not qualified”). The score per group is calculated as the percentage of the maximal possible score, whereby
<table>
<tr><td>division</td><td>I</td><td>II</td><td>III</td><td>IV</td><td>U</td><td>score</td></tr>
<tr><td>Nbr of pupils</td><td>a</td><td>b</td><td>c</td><td>d</td><td>e</td><td>x%</td></tr>
</table>
x = (a*4+b*3+c*2+d*1)/((a+b+c+d+e)*4) *100. As was expected no significant differences (p> 0.05) were found between groups.</td></tr>
<tr><td>4</td><td>Motivation and learning responsibility</td><td>Baseline results January 2013</td><td>Observable improvement of motivation and learning responsibility in terms of significant statistical differences to reject the hypothesis</td><td>Baseline study and End-of-project evaluation.
Measuring tool coming from the Research program “Impact School leadership on Pupil Achievement”</td></tr>
<tr><td colspan="5">➢ By measuring mean scores on different sub-scales on learning conceptions that relate to motivation and learning responsibility (from 1 – totally agree to 5 – totally disagree), we concluded that in 2013 (baseline) pupils adhere stronger to a development orientation (M = 1.62) and external regulation (M = 1.61), and less to a negative orientation (M = 3.07) and individual regulation (M = 3.04). On average they are moreover into serious learning (M = 1.78), as they mostly agree it is important to spend a lot of time and effort on a task and that repetition and memorization are relevant learning activities for them. At the same time they are also convinced about learning activities as elaboration, understanding of learning content and meta-cognitive activities.
➢ In 2015 (end line), pupils still adhere stronger to development orientation (M = 1.57) and external regulation (M = 1.56), and still less to a negative orientation (M = 2.87) and individual regulation (M = 2.79). On average they are moreover still into serious learning (M = 1.68), and they even agree (slightly) more with the statements on profound learning than in 2013.
➢ There are significant differences between the different experimental groups, but it is difficult to find consistency in these differences[8]. Pupils in schools with HT from group A are significantly more in agreement with a development orientation and with external regulation than the other experimental groups and the control group D.</td></tr>
<tr><td colspan="5">At the level of primary school teachers</td></tr>
</table>

[8] For more details see annex 2

	Indicators	Baselines	Targets	Sources of evidence
6	Use of assessment data in the teaching and learning process	Baseline survey January 2013	Observable improvement of teaching and learning process in terms of significant statistical differences to reject the hypothesis	Baseline study, Progress reviews, End-of-project evaluation; Measuring tool coming from the Research program "Impact School leadership on Pupil Achievement"
➢ In 2013 (baseline), primary school teachers on average perceive to use data (see annex 4) for a large range of activities on a monthly or weekly basis (mean scores in between 3.80 and 4.96, on a scale from 1 – hardly to never, to 6 – a few times a week). Almost half of the teachers (47.53 %) use data to investigate why pupils make certain mistakes, or to adapt instruction to the needs of pupils (41.60 %), and to divide their class in groups to give specific instruction (43.07 %). However, 22.02 % never uses data to send pupils to a remedial teacher. ➢ In 2015, there is a significant difference in the use of data (mean scores) between the different experimental groups (at $p < 0.05$): teachers with HT's in group C ($M = 4.95$) use data slightly more often than teachers of HT's in group A (4.81) and especially more often than the teachers with HT in Group B (4.44) and the control group D (4.38). ➢ In 2013 (baseline), teachers of HT's of group C only had a mean value of 4.28 (i.e. significantly less than in 2015 at $p < 0.05$).				

At the level of primary school Head Teachers (HTs)				
8	Progress indicators on the 8 key dimensions of successful school leadership	Baseline values	Relevant progress on at least 1 dimension after 1 year; and on at least 3 dimensions after 2 years Significant statistical differences to reject the hypothesis.	Baseline study: 360° feedback form for Head Teachers, completed by Teachers, PTC member, SEO to measure. A form in which the 8 key competencies are made measurable in a reliable way in terms of practices to be used.

All Head Teachers were scored with the 360⁰FB tool in a baseline and an end line survey by parents, teachers and themselves (self-assessment). Since SEO's tended to fill only one form for what they perceived to be the average of their sector Head Teachers those scores were excluded. In 2013 (baseline) we already found significant differences ($p < 0,05$) between the groups:

2013	Baseline A	Baseline B	Baseline C	Baseline D
Overall score	6.564	6.693	7.220	6.986

In 2015 (end line) also significant differences ($p < 0,05$) were found:

2015	End line A	End line B	End line C	End line D
Overall score	7.743	7.344	7.510	7.279

Since there is no obvious explanation for the differences as we observed them, we have looked to the differences per group, and per teachers & parents (assessment) and Head Teachers (self-assessment) between base and end line:

		Teachers & Parents		Head Teachers		all included	
Group		Ov	difference	Ov	difference	Ov	difference
A-ba	Mean	6.993		4.874		6.564	
	N	138		35		173	
	Std. Deviation	1.257		0.771		1.451	
A-end	Mean	7.718	0.725	7.838	2.964	7.743	1.178
	N	109		27		133	
	Std. Deviation	0.934		0.643		0.891	
B-ba	Mean	7.181		5.124		6.693	
	N	151		47		198	
	Std. Deviation	0.994		1.376		1.401	
B-end	Mean	7.295	0.114	7.647	2.523	7.344	0.651
	N	150		29		175	
	Std. Deviation	0.879		0.743		0.873	
C-ba	Mean	7.209		7.222		7.220	
	N	144		44		184	
	Std. Deviation	1.077		0.991		1.051	
C-end	Mean	7.483	0.274	7.513	0.291	7.510	0.290
	N	208		51		251	
	Std. Deviation	0.768		0.686		0.726	
D-ba	Mean	6.952		7.179		6.986	
	N	106		29		132	
	Std. Deviation	0.876		1.026		0.905	
D-end	Mean	7.197	0.245	7.621	0.442	7.279	0.293
	N	128		33		157	
	Std. Deviation	0.898		0.647		0.875593	

What is remarkable is that while the differences between the end line overall scores of the groups are only very small, the differences between Base and End line for group A and B are much bigger than for the other groups. Especially in the self-assessment of Head Teachers the difference between base and end line of A and B and the other groups is remarkable.

To get a better idea of what was happening with the Head Teachers we had a closer look at the self-assessments that group A and C had provided us with during the reflection workshops in January and July 2014.

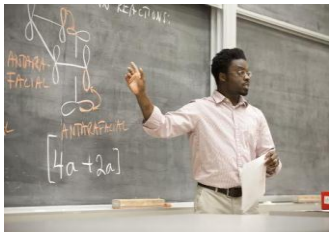
Group		Ov	difference
A-ba	Mean	4.874	
	N	35	
	Std. Deviation	0.771	
A-1-14	Mean	6.785	1.911
	N	30	
	Std. Deviation	0.837	
A-7-14	Mean	7.168	0.383
	N	34	
	Std. Deviation	0.672	
A-end	Mean	7.838	0.671
	N	27	
	Std. Deviation	0.643	
accumulated difference			2.964

Group		Ov	difference
C-ba	Mean	7.222	
	N	44	
	Std. Deviation	0.991	
C-1-14	Mean	7.231	0.009
	N	26	
	Std. Deviation	0.512	
C-7-14	Mean	7.606	0.375
	N	50	
	Std. Deviation	0.551	
C-end	Mean	7.513	-0.093
	N	51	
	Std. Deviation	0.686	
accumulated difference			0.291

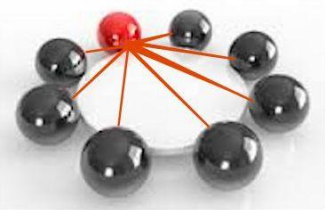
It is interesting to observe that HT's from group A probably were more critical towards themselves than those of group C. One of the assumptions that came to the mind of the project team was that this might be due to the circumstance that these Head Teachers of group A, whenever they completed the 360^0 feedback form this is always done in the presence of their SEO. In order to find more information about this, two Focus Group Interviews were conducted, each involving four head teachers from Group A and the second comprising four head teachers from Group C. Participants were selected at random from both groups. The interviews took place immediately after the participants completed the 360° self-assessment tool during the reflection workshop of July 2014. Although findings from these interviews (see annex 5) seem to give some confirmation of this assumption, it doesn't provided us with new insights about why the same phenomenon occurs with the teachers and parents of groups A. The growth comparing the baseline and end line figures of group A and B is substantial, whereas in the other groups growth seems to be negligible.

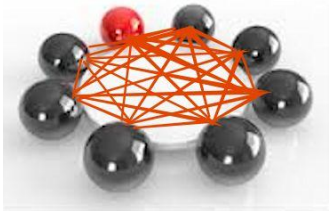
8a	Use of data by Head Teachers in the process of informed decision making	Baseline survey 2013	Observable improvement in terms of significant statistical differences to reject the hypothesis	Baseline study, End-of-project evaluation;
Head teachers' use of data in 2015 differs significantly over the different experimental groups (at $p < 0.05$), with head teachers of group A ($M = 4.09$) using data more often than head teachers of group B ($M = 3.37$) and C ($M = 3.74$), and especially more often than head teachers of the control group D ($M = 3.58$). In 2013 (baseline), head teachers in group A used data less than in in 2015 (in 2013, $M = 3.58$), but this difference is not significant (at $p < 0.05$).				
At the level of the Sector Education Officers (SEOs)				
9	SEO's demonstrate responsiveness to the needs of the learning group and are able to adjust interventions to these needs.	SEO's have insufficient intervision coaching skills	SEO's are applying at least 50% of relevant coaching competencies	Reports on these indicators from their own (trainer-) coaches on the basis of visiting coaching sessions and providing feedback on that. A standardized list of key-competencies for coaching was used to score and report.

In the first phase of the project most of the SEO's were 'telling' or 'selling'. They took the role of lecturer/teacher, the one who knows and is there to distribute their knowledge. Only few of them tried to take a coaching role, i.e. to listen to HT's 'needs' and mobilize peer experience to share in order to 'meet' that needs. Often it was a sequence of monologues, more than a dialogue. On average the SEO used 80 – 90% of the available time to speak. Only initial steps of the 'incident' methodology were followed by SEOs. After this first phase we've seen a tendency of going from lecturing towards a round table conversation. More and more the conversation takes the format of the round table where everybody is talking with each other and sharing experiences, knowledge, possible solutions etc.



At this round table the SEO is no longer the one who is giving turns to participants when to share with him/her. Most of the time this has turned into a situation in which everybody is reacting on everybody. This is going towards a 'richer' conversation guided by the SEO who is guiding and directing the process.





The role of the SEO has changed from the one who tells towards the role of the one who is coaching the process of sharing experience, knowledge, etc. etc.
It was also observed that the initial rather rigid advocated 'intervision'-'incident'-method is not the only method used anymore. Adjusted to the situation also other 'methods' are being used to make sure that benefiting from each other's experience is maximal exploited.

At the level of the school leadership facilitators				
10	Capability of school leadership facilitators in training HTs and SEO's on school leadership.	School leadership facilitators have limited knowledge on the concept of 'educational leadership'	At least 10 school leadership facilitators are able to train HTs and SEO's in educational leadership	Report on basis of observations of trainers of trainers, who visit and participate in the training sessions. This will be complemented with evaluation reports completed by the attendants.

The Educational Leadership Facilitators all were recruited from the 'pool' of trainers working for VVOBs' School Management program. In a first workshop with them they were familiarized with the theory of effective school leadership. They were involved in the formulation of the indicators for the eight dimensions of effective school leadership. After this initial workshop they trained the SEO's of group A and B and the Head Teachers of group A and C. They also facilitated the reflection workshops of the same groups. All of them in the evaluation reports were evaluated as very good trainers (>94%)				
At the level of the Rwanda Education Board (REB)				
11	Availability of evidence-based information on capacity development approaches for school leadership	No evidence-based information is available	A clear set of recommendations is available, shared and owned by REB by the end of the project period.	Report with set of recommendations. Whether these recommendations are owned and shared by REB can only appear from plans of action for up scaling the project from their side. The latter depends on a lot of other factors like available budget, intra- and inter-ministerial priorities, etc.
Since the source of evidence can only be accessed after finalizing the project this will not be part of the evaluation of the project. In the meantime REB's School Leaders & Management Unit together with VVOB started in 2014 a pilot project training 30 SEO's to guide and coach their Head Teachers. In this project the first experiences with this IfE project are being further elaborated and 'transformed' into working with PLN's for Head Teachers as a method for Continuous Professional Development.				
At the level of costs				
12	Availability of costs related to the distinct capacity development approaches for school leadership	No information is available	A clear body of knowledge is available on costs and cost effectiveness related to the distinct capacity development approaches for school leadership	Report on the costs of the distinct capacity development approaches for school leadership

All data were gathered about the costs of the three distinct capacity development approaches and were measured in terms of staff time and budget. Interesting situation however is that, for reasons of scale-advantages[9], several activities related to different approaches were conducted in combination with each other. In order to come to a realistic comparison of the costs of the different approaches, calculated costs were extrapolated to a situation in which activities were conducted for 2* 8 Sectors. Included were all costs (HR, logistics, activities, capital investments, office costs, overhead) that were related to the implementation of the project. Cost related to doing research were excluded. Since the ultimate objective of the project is improving pupil learning achievement, the costs were expressed in costs per pupil[10].

group	A	B	C
costs per pupil	£2,07	£1,25	£1,35

Table : extrapolated cost per pupil per approach

Those calculations make clear the approach in which only Sector Educational Officers were trained to guide and coach intervision meetings with their Head Teachers and supported to organize these meeting on a monthly basis was less expensive.

§ 6 Conclusions and recommendations

Conclusions are first given for the different levels of indicators. This is followed by overall conclusions and recommendations.

As was expected at the level of primary school pupils no evidence was found that the exam results of schools led by Head Teachers from experimental groups differ from the control group. The timeframe to already see impact at this level (to make this 'domino stone' fall) was simply too short.

For pupils' learning conceptions, although there were some significant differences found between the different experimental groups, those are difficult to interpret since it is difficult to find consistency in these differences. Pupils in schools with HT from group A are more in agreement with a development orientation and with external regulation than the other experimental groups and the control group D. There is still a lot of discussion needed to be sure if this combination of subscales is actually predicting higher learning achievements of pupils.

At the level of teachers significant differences are found in the use of data between the different experimental groups: teachers with HT of group C (M = 4.95) use data slightly more often than teachers of HT of group A (4.81) and B (4.44) and especially more often than the teachers with HT in the control group D (4.38). The differences between group C and A are very small and although significant one could say that the difference in effect (if attributable to the difference in intervention) is negligible.

At the level of Head Teachers their use of data in 2015 differs significantly over the different experimental groups (at $p < 0.05$), with head teachers of group A (M = 4.09) using data more often than head teachers of group B (M = 3.37) and C (M = 3.74), and especially more often than head teachers of the control group D (M = 3.58). In 2013 (baseline), head teachers in group A used data

[9] Training of SEO's was done with the SEO's of group A and B together. It is obvious that training 16 SEO's together has financial advantages above training them apart.

[10] For further calculation see annex costs per approach and staff salary costs per approach

less than in in 2015 (in 2013, M = 3.58), but since this difference is not significant (at $p < 0.05$) there is no 'evidence' this can be attributed to the intervention.

When looking at the assessment of Head Teachers measured against the indicators of the eight dimensions of effective school leadership significant differences ($p < 0,05$) were found between the baseline and the end line and between the different groups. Remarkable is that while the differences between the end line overall scores of the groups only are very small, the difference between baseline and end line for group A and B is much bigger than for the other groups. Especially in the self-assessment of Head Teachers the difference between base and end line of A and B and the other groups is remarkable. Although there is no good explanation for the differences between the groups in the baseline, at least it could be assumed the growth the performance of Head Teachers could be attributed to the approach used in A and in a lesser extend in B.

Taking into account the level of SEO's, knowing both SEO's of group A and B were trained together and no differences were observed in their coaching performance while conducting their intervision sessions at Sector level, it is not plausible that differences between Head Teacher performance can be attributed to differences between SEO's. A comparable statement is valid for possible differences between Education Leadership Facilitators who facilitated the Head Teachers of group A and C and SEO's of group A and B, since they were performing at the same high level and rotated over the groups.

Bringing findings together at an overall level we dare state that, apart from the learning outcomes in exams, the experimental groups differ significant from the control group. There are however no clear, unambiguous and consistent findings yet that could legitimize statements about the ultimate best approach. Only at the level of Head Teachers performance it could be stated the growth in performance of Head Teachers in group A is best, followed by those in group B. The differences in measured performance at the end line between group A and B however are significant, but relatively small.

As can be found in §5 the costs per pupil of the distinct approaches were calculated respectively £2,07 for group A, £1,25 for group B and £1,35 for group C.

Those calculations make clear the approach in which only Sector Educational Officers were trained to guide and coach intervision meetings with their Head Teachers and supported to organize these meeting on a monthly basis were less expensive. Since we found at result level no decisive evidence that legitimizes other considerations for deciding for group A or B, it can be stated approach B is most cost effective.

Recommendations

In accordance with this conclusion two different types of recommendations can be given. The first type gives recommendations about how to proceed the research in the ongoing pursuit of getting 'evidence'. The other type of recommendations concerns the way in which the conclusion above can be embedded in the existing Rwanda policy on school leadership and the way that is developing under the direction of the Ministry of Education and the Rwanda Education Board.

The following research recommendations can be given:

1. Knowing the project is based on the assumption that the impact of effective school leadership works via a 'domino'-effect, it is recommended to extend the project over a number of years, so that in due course the effect on pupils' learning achievement can be measured;
2. In order to be able to come up with more reliable data it is recommended to further elaborate some of the indicators and develop more adequate tools for measuring;

3. In order to be able to come up with more underpinned interpretation of data it is recommended to also systematically gather more qualitative data that could make this possible.

As said above In addition to the recommendations for further research also policy-recommendations can be given. Projects of action research like this do not occur in a social vacuum. There is a context of policies that is always present in the background and should be taken into account. Before formulating these recommendations a brief overview is given of this policy context in which these recommendations need to be placed.

All activities in education in Rwanda have to be in line with the Education Sector Strategic Plan 2013/14 – 2017/18. In this ESSP both in priority 2 and 10 the importance of training and performance assessment of Head Teachers is emphasized. In the formulated strategies is said that "A national training and development programme in school leadership and management will be rolled out to build head teachers' capacity for effective school management, which will be accompanied by a system of school leadership and management performance assessments". The expected outputs are an "Increased proportion of head-teachers and school managers demonstrating capabilities for effective school management" and "All education managers and governance bodies demonstrate requisite skills and competencies for effective leadership and governance".

The Teacher Development and Management (TDM) Policy is in line with that and describes its target 5 of this policy as "Enhance school leadership quality and training", specifying this under target 5.1 "Provide a major programme of CPD activities for head teachers" where it is emphasized that "It is essential that head teacher training focuses on 'leadership' (rather than only 'administration,' i.e. in financial management). School leader training must also include leadership in pedagogical best practices and the use of assessment data to improve teaching and learning. Another specification is given as target 5.2 "Require all newly appointed head teachers to obtain a School Leadership Certificate by the end of their first year of headship at the latest".
Also in line with the ESSP REBs' School Leadership & Management Unit developed in collaboration with the British Council and VVOB the '5 Standards for effective School Leadership". In fact the '8 dimensions of successful School Leadership' as they were used in the project, have been incorporated in these 5 standards and a comparable 360^{0}FB tool has been developed and validated.

Also in line with the ESSP as a first successor of the "Coaching School Leadership to achieve high level learning outcomes" project, a new pilot (for 30 Sectors) has been started in 2014 as a collaborative project of REB and VVOB. The lessons learnt were taken into account in the process of preparing and conducting this pilot as a part of the joint action plan REB-VVOB. One essential change in the design is the change from an strict individual approach in the 'coaching school leadership…' project towards a more collective approach in the new pilot project. It was found that it took a lot of effort to rigidly stick to individual learning needs (as formulated by individual Head Teachers) and use the intervision 'incident'-method to deal with them. In addition to this method other methods were used and often the reason for this was that there was a need for dealing with 'shared problems'. In doing this the group of Head Teachers of a Sector seemed to get the character of a Professional Learning Network (PLN) that was coached by the SEO. In the new pilot project this has been made official and the focus of the SEO's training has shifted towards 'coaching effective PLN's' in perspective of establishing a system for CPD for Head Teachers at Sector level.

In order to realize target 5.2 of the TDM policy another project is being executed as a collaborative project of REB, the University of Rwanda-College of Education (URCE) and VVOB concerns the development of the initial in-service training for School Leadership, that is competence based and leads to a URCE diploma.

Taken this context and evolving developments into account, from the findings of the 'coaching school leadership to achieve high level learning outcomes' project the following 'policy'-recommendations can be given:

4. Given the policies as formulated in ESSP 13/14 – 17/18 and the TDM policy document, in which the importance of an initial training for Head Teachers, complemented by a system for Continuous Professional Development is emphasized, it is recommended to use the approach in which SEO's are trained to coach Head Teachers (approach B) as the most cost-effective way of working;
5. In case up scaling of the 'Coaching School Leadership …' project is considered it is recommended to replace the '8 dimensions' and its affiliated tools with the '5 standards for effective school leadership' and its tools;
6. In case up scaling of the 'Coaching School Leadership …' project is considered it is recommended to incorporate REB's-VVOB pilot project and to further elaborate it as a system for CPD for Head Teachers in line with TDMs policy target 5.1;
7. In case up scaling of the 'Coaching School Leadership …' project is considered it is recommended to build on existing structures and to involve all stakeholders from the outset in the design, implementation and maintenance of this system of continuous professional development for School Leaders.

Annexes

Annex 1: Random sample of Sectors

Available was the list, to which the numbers 1 to 416 were attributed:

Provence	District	Sector	Nbr
North	in alphabetical order	in alphabetical order	1 to
South	in alphabetical order	in alphabetical order	
East	in alphabetical order	in alphabetical order	
West	in alphabetical order	in alphabetical order	
Kigali	in alphabetical order	in alphabetical order	to 416

With the excel function 'RANDBEWTEEN(top; bottom)' 32 numbers were sampled where the places 1, 5, 9, 13, 17, 21, 25 and 29 were attributed to box A; 2, 6, 10, 14, 18, 22, 26 and 30 to box B; 3, 7, 11, 15, 19, 23, 27 and 31 to box C; and 4, 8, 12, 16, 20, 24, 28 and 32 to box D.
Then 3 sectors in Gasano District, where Well Spring Foundation is working on a school development project had to be replaced for 'doubling' and/or attribution reasons, with three newly generated random numbers.

With the following result:

Province	**District**	**Sector**	**BOX**		**Province**	**District**	**Sector**	**BOX**
Eastern	Kayonza	Ndego	A		Eastern	Kayonza	Kabare	B
Eastern	Kirehe	Kirehe	A		Northern	Gakenke	Gashenyi	B
Eastern	Ngoma	Rukira	A		Northern	Gakenke	Nemba	B
Northern	Gicumbi	Nyamiyaga	A		Northern	Gicumbi	Miyove	B
Northern	Rulindo	BUREGA	A		Northern	Musanze	Musanze	B
Southern	Kamonyi	Ngamba	A		Western	Nyamasheke	Bushekeri	B
Southern	Nyanza	Kigoma	A		Western	Rubavu	Rugerero	B
Western	Rutsiro	Musasa	A		Western	Rutsiro	Mushubati	B
Province	**District**	**Sector**	**BOX**		**Province**	**District**	**Sector**	**BOX**
Eastern	Kirehe	Mpanga	C		Southern	Muhanga	Nyabinoni	D
Eastern	Rwamagana	Gahengeri	C		Southern	Nyaruguru	Mata	D
Northern	Gicumbi	Shangasha	C		Eastern	Ngoma	Mutenderi	D
Southern	Huye	Rusatira	C		Western	Ngororero	SOVU	D
Southern	Muhanga	Muhanga	C		Western	Nyabihu	Jenda	D
Southern	Nyanza	Busasamana	C		Western	Nyamasheke	Karengera	D
Southern	Nyanza	Cyabakamyi	C		Western	Rubavu	Cyanzarwe	D
Southern	Ruhango	Kabagali	C		Kigali	Kicukiro	Kicukiro	D

Annex 2: Learning conceptions of pupils in primary education in Rwanda

1. Theoretical background

1.1. Learning conceptions as mediator for learning outcomes

Learning conceptions can be considered as an aspect of meta-cognition regarding personal knowledge, attitudes, conceptions and views that persons have related to learning and education (Ten Dam et al., 1999; Verloop & Lowyck, 2003). Learning conceptions determine what pupils, in a specific context, understand and how they interpret learning tasks, learning goals and learning situations. Traditionally, different studies describe learning conceptions one dimensionally (Marton & Säljö, 1976; Marton, et al., 1993; Van Rossum & Schenk, 1984). However, further investigations of the content of learning conceptions lead to the insight that conceptions can be categorized in different aspects of learning (Klatter, 2001). According to Klatter (2003) everyone's belief system on learning consists of a set of beliefs concerning different aspects of learning, such as the goal, the results and the advantages of learning. Therefore, learning conceptions can be considered as a multi-dimensional concept (Klatter, 2004; Verloop & Lowyck, 2003). Verloop en Lowyck (2003) distinguish the following aspects as most relevant: conceptions of learning activities and learning strategies, conceptions of the learner on his/herself, conceptions of learning goals, conceptions of learning tasks, conceptions of learning and studying and conceptions of steering and regulation (Klatter, 2003; Van Petegem, Klatter, Dang & Donche, submitted).

Learning conceptions play an important role in how learning situations are approached. A learning conception functions as a filter for new information and guides the interpretation of different learning situations. Those interpretations form the foundation of the eventual learning activities that lead to a certain learning outcome. Therefore, learning conceptions play a mediating role between the learning environment and the actual learning behavior.

1.2. Measuring learning conceptions

Previous research has shown that learning conceptions of pupils can be measured with the learning conceptions survey scale (Klatter, Lodewijks & Aarnoutse, 2001). The survey allows to map the conceptions of pupils on a larger scale and to study the specific correlations between them. The original survey consists of 100 items that refer to three aspects of learning: motivation (the goal of schooling and learning orientation), steering or regulation, and mental processing activities (learning conditions and learning activities). Based on this survey, eight factors of learning conceptions can be identified: 1) development orientation, 2) negative motivation (avoidance orientation), 3) external steering, 4) individual steering, 5) steering by peers, 6) superficial learning, 7) undirected learning and 8) profound learning (Klatter, 2003). In table 1 (hereunder) we describe in more detail the different factors of learning conceptions as identified with the learning conceptions survey of Klatter (2003).

Orientation (motivation)	
Negative orientation	. These pupils have a negative orientation on schooling and learning and do not understand the use of school tasks
Development orientation	These pupils stress the interest they have in learning. They enjoy learning and aim for personal growth and independence.
Steering	
External steering	These pupils strongly depend on the support of the teacher. They are convinced it is the task of the teacher to explain in detail the tasks and about the evaluations.
Steering by peers	These pupils prefer to receive learning and study advice of their peers. They are also interested in how other pupils think and learn and compare this with their own learning habits.
Individual steering	These pupils prefer to work and learn on their own.
Learning (mental processing activities)	
Superficial learning	These pupils are characterized by a quantitative approach regarding learning. Spending a lot of time and effort on a task is seen as an important prerequisite for learning. Repetition and memorization are relevant learning activities for them.
Undirected learning	These pupils have difficulties with carrying out a task, as they don't know how to start. They consider their learning style inefficient and stress the factor of luck.
Profound learning	These pupils are convinced about learning activities as elaboration, understanding of learning content and meta-cognitive activities.

Table 1: Factors of learning conceptions (Klatter, 2003)

Correlations between the identified learning conception factors confirm the definition of Klatter (2003) in which learning conceptions are defined as a cluster of interrelated conceptions and in which learning conceptions can be considered as a multi-dimensional concept.

2. Research questions and methodology

The learning conceptions survey has been developed and validated for measurement of learning conceptions of primary education pupils in a Western context. The question for this study is whether the survey is a valid and reliable instrument to measure learning conceptions of primary education pupils in Rwanda as well. If that is the case, the instrument can be used to get a detailed insight in the current learning conceptions of Rwandese primary education pupils and the correlations between the different learning conception factors.

Therefore, a reduced learning conceptions survey, adjusted from the original learning conceptions survey, consisting of 44 items has been used to collect responses on learning conceptions of about 10,000 Rwandese pupils of grade 4, 5 and 6. In order to investigate whether the instrument can be used for longitudinal research on patterns in the development of learning conceptions, a series of validity and reliability tests are performed on the survey.

Factor analysis

To explore the number of scales in the learning conceptions survey, a factor analysis with Oblimin rotation is performed on the three aspects or components of learning conceptions as identified by Klatter (2003): motivation, steering and mental processing activities. As a criterion for item acceptance

is that each item should contribute to the homogeneity of the total scale, items with a factor loading below .30 are removed from the scale.

Internal consistency

For each sub-scale, internal consistency is calculated. The reliability of the different scales will be determined based on the homogeneity expressed in the reliability coefficient Chronbach's Alpha (Cronbach, 1951). Usually the acceptance norm for educational research is a reliability of .70 in which 50 % of the variance is explained (Nunnally & Bernstein, 1994). The squared Cronbach's Alpha (α) identifies the proportion of explained variance that is commonly explained by the different items of a scale.

Construct validity

In general, validity refers to the degree in which an instrument measures what it is supposed to measure. In particular, validity refers to whether the identified scales constitute a sound construct. The degree of validity defines the degree to which conclusions that are based on the measurements with the instrument are valid conclusions. To assess construct validity we map the coherence between the different subscales with a correlation matrix (through Pearson correlation coefficients).

The findings of the learning conceptions survey offer insight in the conception of pupils in grade 4, 5 and 6 of primary education in Rwanda, with regards to their learning orientation, steering and mental processing activities. Mean scores on the subscales are interpreted and differences between male and female pupils are explored.

3. Findings

3.1. Factor analysis

Before analyzing the factor loadings it has to be noted that one item (item 15) was formulated in a negative way. All factor loadings resulting from the factor analysis on the three components or aspects of learning as identified by Klatter (2003) can be found in annex 1.

In the factor analysis on the component of *motivation* two factors were retained: development orientation (explaining 16.81 % of variance) and negative orientation (explaining 15.34 % of variance). For both factors, one item has been removed from the subscale as they did not load as expected in the original survey design of Klatter 92003). This resulted in two subscales with 6 items in the subscale "development orientation" and 7 items in the subscale "avoidance orientation".

For the second component, *steering*, three factors were identified, as in the original study of Klatter (2003). The subscale on "external steering" consists of 7 items with each a factor loading between .55 and .68. For the subscale on "individual regulation", 4 items are retained. Item 15, which was formulated negatively, in fact loaded higher on the last factor, "regulation by peers". For this last subscale, all 5 items of the subscale are retained.

For the third component, *learning*, only two distinct factors are found through the factor analysis, while in the original study of Klatter (2003) 3 factors were retained. In this study, we see that items of both theoretical subscales load in fact on the same factor, which we call "serious learning". Pupils that adhere to "serious learning" are both in favour of profound learning as well as superficial learning activities. In total, 8 items have a factor loading on this factor above .30. The second factor that is identified through the factor analysis refers to "undirected learning". This subscale consists of 5 items with factor loadings above .30.

3.2. Reliability

In table 2 we provide an overview of Cronbach's α for the different subscales identified through the factor analysis as described above, as well as the mean score (M) and standard deviation for the combined subscale.

Component/scale	*# items*	*α*	*N*	*M*	*SD*
Orientation					
Development orientation	6	0.63	9129	1.62	0.56
Negative orientation	7	0.62	8664	3.07	0.76
Regulation					
External regulation	7	0.71	9180	1.61	0.55
Individual regulation	4	0.66	9136	3.04	0.99
Regulation by peers	6	0.56	8992	1.96	0.61
Learning					
Serious learning	8	0.67	8917	1.78	0.55
Undirected learning	5	0.50	9039	3.06	0.80

Table 2: Descriptive statistics for identified subscales on aspects of learning

Cronbach's α ranges between .50 to .71. In fact, only one subscale, on "External regulation" can be accepted according to the criteria as formulated by Nunnally & Bernstein (1994). All other scales have Cronbach's α below .70. However, if the reliability criterion is reduced to a Cronbach's α value of .60 or higher, five subscales are retained as reliable.

3.3. Construct validity

The factor analysis and reliability analysis make clear that the learning conception survey instrument consists of five homogeneous subscales, while in the original instrument, eight subscales are identified. In table 3 we present a correlation matrix (Pearson correlation coefficient) between the subscales.

Table 3: Correlation matrix between subscales learning conceptions

	S1_devori	S2_negori	S3_extreg	S4_indreg	S5_regpee	S6_serlea	S7_undlea
S1_devori	1.00						
S2_negori	-0.02	1.00					
S3_extreg	.613**	.024*	1.00				
S4_indreg	-.031**	.459**	0.00	1.00			
~~***S5_regpee***~~	~~*.465***~~	~~*.080***~~	~~*.419***~~	~~*.061***~~	~~*1.00*~~	-	-
S6_serlea	.623**	.048**	.603**	0.01	~~*.469***~~	1.00	
~~***S7_undlea***~~	~~*.045***~~	~~*.558***~~	~~*0.00*~~	~~*.408***~~	~~*.077***~~	~~*.028***~~	~~*1.00*~~

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Strikethrough: subscale is not reliable (Cronbach's α < .60)

Most interesting are the subscales on regulation. While "external regulation" correlates strongly with "development orientation" (r = .613) and "serious learning" (r = .603), "individual regulation" correlates negatively with "development orientation" (r = -.031), and strongly with "negative orientation" (r = .459), while there is no significant correlation with "serious learning". "Individual regulation" moreover

correlates positively with "undirected learning". However, as shown in the reliability analysis, the subscale "undirected learning" does not seem to be a reliable measurement. It is clear that pupils that strongly depend on the support of the teacher and are convinced it is the task of the teacher to explain in detail the tasks and evaluation, also enjoy learning and aim for personal growth and independence. This is the opposite for pupils that prefer to work and learn on their own. These individual learners also have a negative orientation towards schooling and learning and do not understand the use of school tasks. Moreover, they seem to have difficulties with carrying out a task, as they don't know how to start. They consider their learning style inefficient and stress the factor of luck.

PE Pupils' (G4-G6) conceptions of learning in Rwanda

In general, as is shown in table 4, Rwandese pupils adhere stronger to a development orientation (M = 1.62) and external regulation (M = 1.61), and less to a negative orientation (M = 3.07) and individual regulation (M = 3.04). On average they are moreover into serious learning (M = 1.78), as they mostly agree it is important to spend a lot of time and effort on a task and that repetition and memorization are relevant learning activities for them. At the same time they are also convinced about learning activities as elaboration, understanding of learning content and meta-cognitive activities.

Table 4: Comparison between male and female pupils' learning conceptions (Mean scores)

	Male	Female	Total
S1_devori	1.60	1.64	1.62
S2_negori	3.09	3.06	3.07
S3_extreg	1.61	1.61	1.61
S4_indreg	3.09	3.00	3.04
S6_serlea	1.77	1.79	1.78

When we compare boys and girls, there are a few small but significant differences in mean scores. While boys are slightly more oriented towards development than girls, on average, girls tend to be slightly more inclined towards individual regulation than boys.

If these scores are compared with how primary education pupils in a study in the Netherlands responded to the learning conceptions survey, it can be seen that Rwandese pupils seem to be more oriented towards development and have a less negative orientation towards schooling and education then pupils in the Netherlands and that they are also less inclined to learn and work on their own (individual regulation). Pupils in the Netherlands seem to be both profound and superficial learners, which is similar to Rwandese pupils, who can also be considered as serious learners, in the sense they combine both profound and superficial learning.

Annex 2a

Item factor loadings

Item	F1	F2	F3	F4	F5	F6	F7	F8
SCALE 01 Development orientation								
35. In Arithmetic/Maths I always want to find out exactly how everything works.	**0.68**	-0.03						
21. In Language/Dutch I always want to find out exactly how everything works.	**0.45**	0.02						
9. I often want to learn more about all the things we cover at school.	**0.54**	-0.03						
45. If I have to do something for Arithmetic/Maths, I always want to know	**0.62**	0.00						
31. If I have to learn something, I really enjoy finding out everything about it.	**0.48**	-0.02						
6. I often want to learn things for Arithmetic/Maths just because they are fun	0.01	**0.49**						
18. If I have to learn something, I always want to know exactly how everything	**0.58**	-0.07						
SCALE 02 Negative orientation								
39. I mainly go to school because I have to.	-0.08	**0.67**						
42. I mainly enjoy school if I don't have to do very much.	-0.21	**0.67**						
12. When I'm at school, I want time to go as quickly as possible.	-0.20	**0.48**						
3. I only go to school in order to be able to get a better job later.	0.32	**0.38**						
19. If I could choose some of my subjects, I would only choose subjects I like.	0.00	**0.48**						
24. I mainly go to school, so that I can earn a lot of money later.	0.28	**0.49**						
29. How I learn isn't important, so long as I get a good mark.	0.06	**0.58**						
16. Learning is only worthwhile, if you can use what you've learnt later.	**0.55**	0.04						
SCALE 03 External regulation								
27. I think that the teacher has to explain everything very precisely as that's what			**0.59**	-0.02	0.17			
46. In Arithmetic/Maths I prefer the teacher to tell me exactly what I have to			**0.52**	0.04	0.10			
30. When doing a task, I want to have as much explanation as possible as this makes			**0.64**	-0.05	0.29			
34. In Language/Dutch I prefer the teacher to tell me exactly what I have to do.			**0.68**	0.02	0.14			
25. I like to have as much explanation as possible for every task.			**0.57**	-0.10	0.22			

38. If I don't understand something in Language/Dutch, I prefer to ask the teacher			**0.55**	-0.03	0.13			
32. I like to have instructions from the teacher in order to know the best way to			**0.60**	-0.02	0.11			
SCALE 04 Individual regulation								
4. In Arithmetic/Maths I prefer to work on my own rather than in a team.			0.07	**0.63**	0.09	-		
15. In Arithmetic/Maths I prefer to work in a team with other pupils (reverse coded)			0.29	-0.35	**0.40**			
13. In Language/Dutch I prefer to work on my own rather than in a team.			-0.01	**0.71**	-0.04			
43. I prefer to do Arithmetic/Maths on my own because other pupils have a different			-0.02	**0.72**	-0.02			
47. I prefer to do Language/Dutch on my own because other pupils have a different			-0.06	**0.72**	-0.05			
SCALE 05 Regulation by peers								
36. I think it's important to know what other pupils think and how they learn.			0.15	-0.05	**0.63**			
33. I like to get tips from other pupils about the best way to learn.			0.47	-0.09	**0.45**		.	
48. I like to get tips from other pupils about the best way to do assignments for			0.45	-0.15	**0.41**			
23. I often compare my way of working in Language/Dutch with that of other pupils.			0.16	0.04	**0.66**			
2. I often compare my way of working with that of other pupils.			0.15	0.04	**0.64**			
SCALE 06 Superficial learning demands								
37. If you spend a very long time studying for a test, you can actually get a good						**0.44**	0.01	-0.41
26. If you work hard for Language/Dutch, you can always get good marks.						0.26	0.02	-0.50
41. Pupils get good marks if they learn everything by heart.						0.07	0.06	-0.76
8. If you work hard for a subject, you can always get good marks.						**0.58**	-0.03	-0.24
28. The best thing to do to get good marks is to learn as much as possible by heart.						**0.38**	0.01	-0.68
SCALE 07 Undirected learning demands								
5. You can only get a pass mark, if a test is easy.						0.12	**0.53**	-0.23
20. I often don't manage to learn the material because I don't know how to go						0.06	**0.56**	0.15
22. Getting a good mark in a test is just a question of luck.						-0.17	**0.61**	-0.13
17. Whether or not you get a good mark depends entirely on how difficult the test						0.02	**0.61**	-0.05
11. In my opinion, I often don't do assignments for Arithmetic/Maths in the						-0.08	**0.53**	0.16

SCALE 08 Profound learning demands and								
14. In my opinion, when you are learning something, you need to try to imagine						**0.61**	-0.12	-0.25
1. When I learn I always try to think of examples that relate to the topic I'm						0.21	0.00	-0.41
7. When you are learning something, you must always think carefully about what you						**0.55**	-0.01	-0.17
44. In my opinion, it is important when you are learning something to check that you						**0.63**	0.00	-0.22
10. If you're learning something, you must always think about which parts of the topic						**0.65**	0.11	-0.17
40. In order to understand a difficult passage in a test, the best thing to do is to						**0.39**	0.27	-0.25
Percentage of explained variance	16.81	15.34	19.25	12.20	7.69	17.66	10.84	7.22
Cumulative variance	16.81	**32.15**	19.25	31.45	**39.14**	17.66	28.50	**35.73**

Annex 3: exam results 2012, 2013, 2014 in the A, B, C and D group

Report exam results 2012. 2013. 2014 in the A. B. C and D group							
Group		# I	# II	# III	# IV	# U	SCORE
A12	Mean	4.50	97.75	95.75	58.62	34.38	46.1006%
	N	8	8	8	8	8	8
	Std. Deviation	3.891	63.529	36.507	29.223	26.785	12.70108%
A13	Mean	7.13	80.50	116.00	83.50	76.38	37.2339%
	N	8	8	8	8	8	8
	Std. Deviation	9.906	70.723	69.344	33.726	44.484	13.14964%
A14	Mean	6.12	76.50	92.88	66.75	59.12	39.4423%
	N	8	8	8	8	8	8
	Std. Deviation	13.747	70.959	52.133	34.358	61.821	18.75562%
B12	Mean	6.75	86.87	116.00	80.88	60.25	44.6877%
	N	8	8	8	8	8	8
	Std. Deviation	9.498	38.996	53.018	52.477	40.142	7.29218%
B13	Mean	5.38	82.75	120.88	95.12	82.62	40.2607%
	N	8	8	8	8	8	8
	Std. Deviation	2.875	29.485	39.966	60.239	42.352	7.16918%
B14	Mean	9.00	100.88	117.62	79.25	50.75	46.2899%
	N	8	8	8	8	8	8
	Std. Deviation	8.018	52.409	62.186	52.062	29.798	9.66803%

C12	Mean	6.75	157.88	124.25	63.25	29.38	53.1733%
	N	8	8	8	8	8	8
	Std. Deviation	7.421	95.492	30.686	36.276	23.910	9.64870%
C13	Mean	5.75	108.25	156.50	89.88	56.00	43.8881%
	N	8	8	8	8	8	8
	Std. Deviation	12.669	81.456	51.733	33.327	17.004	6.51319%
C14	Mean	9.75	133.62	131.00	63.25	30.12	50.7085%
	N	8	8	8	8	8	8
	Std. Deviation	18.752	87.985	45.635	23.255	19.497	6.16803%
D12	Mean	5.25	71.12	103.88	84.50	81.62	40.7395%
	N	8	8	8	8	8	8
	Std. Deviation	11.273	24.015	50.665	43.802	84.020	10.93784%
D13	Mean	6.62	46.75	106.75	88.88	87.38	37.4052%
	N	8	8	8	8	8	8
	Std. Deviation	16.353	25.516	45.361	44.700	56.556	15.13144%
D14	Mean	7.62	60.12	104.50	82.00	54.88	42.3577%
	N	8	8	8	8	8	8
	Std. Deviation	18.769	20.074	46.953	40.235	36.192	11.97134%

ANOVA Table exam results 2012. 2013. 2014 in the A. B. C and D group

			Sum of Squares	df	Mean Square	F	Sig.
# I * Group	Between Groups	(Combined)	204.531	11	18.594	.125	1.000
	Within Groups		12452.875	84	148.249		
	Total		12657.406	95			
# II * Group	Between Groups	(Combined)	83453.083	11	7586.644	2.053	.033
	Within Groups		310348.250	84	3694.622		
	Total		393801.333	95			
# III * Group	Between Groups	(Combined)	26131.000	11	2375.545	.961	.488
	Within Groups		207649.000	84	2472.012		
	Total		233780.000	95			

# IV * Group	Between Groups	(Combined)	12703.615	11	1154.874	.667	.765
	Within Groups		145375.375	84	1730.659		
	Total		158078.990	95			
# U * Group	Between Groups	(Combined)	36706.615	11	3336.965	1.695	.088
	Within Groups		165338.875	84	1968.320		
	Total		202045.490	95			
SCORE * Group	Between Groups	(Combined)	2191.447	11	199.222	1.543	.132
	Within Groups		10846.646	84	129.127		
	Total		13038.093	95			

Annex 4 - Comparison of mean values on teachers' use of data in 2015 (A-B-C-D)

Report

USEdata

Condition	Mean	N	Std. Deviation
A	4.8138	511	1.01032
B	4.4414	660	1.10764
C	4.9483	643	.71708
D	4.3799	558	1.14741
Total	4.6446	2372	1.03355

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
USEdata * Condition	Between Groups	(Combined)	140.261	3	46.754	46.275	.000
	Within Groups		2392.518	2368	1.010		
	Total		2532.779	2371			

Comparison of mean values of group C (2013-2015)

Report			
USEdata			
Year	Mean	N	Std. Deviation
2013	4.2758	832	1.22317
2015	4.9483	643	.71708
Total	4.5690	1475	1.08568

ANOVA Table							
			Sum of Squares	df	Mean Square	F	Sig.
USEdata * Year	Between Groups	(Combined)	164.006	1	164.006	153.539	.000
	Within Groups		1573.413	1473	1.068		
	Total		1737.418	1474			

Tlesson = (q_01 + q_02 + q_03)/3
Tschool = (q_05 + q_06)/2
Tbelief = (q_07 + q_09 + rq_04 + rq_08) / 4

Annex 5 - Comparison of mean values of HT use of data in 2015 (A-B-C-D)

Report			
USEdata			
Condition	Mean	N	Std. Deviation
A	4.0884	21	.83143
B	3.3714	30	.88271
C	3.7433	32	.71379
D	3.5750	40	.94567
Total	3.6568	123	.87757

ANOVA Table							
			Sum of Squares	df	Mean Square	F	Sig.
USEdata * Condition	Between Groups	(Combined)	6.863	3	2.288	3.126	.028
	Within Groups		87.093	119	.732		
	Total		93.956	122			

Comparison of mean values group A (2013-2015)

Report			
USEdata			
Year	Mean	N	Std. Deviation
2013	3.5844	22	.86662
2015	4.0884	21	.83143

Total	3.8306	43	.87732

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
USEdata * Year	Between Groups	(Combined)	2.729	1	2.729	3.781	.059
	Within Groups		29.597	41	.722		
	Total		32.327	42			

Annex 6: cost calculation per approach

In order to come to a realistic comparison of the costs of the different approaches, calculated costs were extrapolated to a situation in which activities were conducted for 2* 8 Sectors. Included were all costs (HR, logistics, activities, capital investments, office costs, overhead) that were related to the implementation of the project. Cost related to doing research were excluded. Since the ultimate objective of the project is improving pupil learning achievement, the costs were expressed in costs per pupil

Items	Unit cost	Unit name	No. of units	A	B	C
staff salary costs	see annex staff salary costs per approach			£92.246,53	£75.528,47	£54.398,31
logistics per activity						
Activity 1a						
travel costs trainers	£0,28	km	130	£36,99		£36,99
travel costs trainees	£10,29		20	£205,82		£205,82
Activity 1b						
travel costs trainers	£0,28	km	520	£147,97		£147,97
travel costs trainees	£10,29		40	£411,65		£411,65
activity 2a						
travel costs trainers	£0,28	km	1560	£443,90	£36,99	£443,90
travel costs trainees	£10,29		114	£2.346,40	£329,32	£2.017,08
activity 2b						
travel costs trainers	£0,28	km	6240	£1.775,61	£295,93	£1.775,61
travel costs trainees	£10,29		912	£9.385,61	£1.317,28	£8.068,33
activity 3a						
travel costs trainers	£0,28	km	520	£147,97	£147,97	
travel costs trainees	£10,29		128	£1.317,28	£1.317,28	
travel costs consultant	£1.180,00	return flight	4	£4.720,00	£4.720,00	
activity 3b						
travel costs coaches	£0,28	km	7200	£2.048,78	£2.048,78	
activity 4						
travel costs SEO's	£5,15	travel fee	368	£1.893,59	£1.893,59	

Items	Unit cost	Unit name	No. of units	A	B	C
travel costs HTs	£5,15	travel fee	2254	£11.598,23	£11.598,23	
activity 5						
travel costs trainers + support officers	£0,28	km	2080			
travel costs participants (SEOs & HTs	£10,29	travel fee	114			
activity 6						
travel costs participants	£10,29	travel fee	56	£576,31	£576,31	£576,31
accommodation, materials, etc.:						
Sub-total Activity 1				£4.072,82		£4.072,82
Sub-total Activity 2				£62.393,32	£9.119,02	£62.393,32
Sub-total Activity 3				£15.123,83	£15.123,83	
Sub-total Activity 6				£1.500,00	£1.500,00	£1.500,00
capital investments						
presentation materials	£706,74	p/unit	2	£1.413,48	£706,74	£1.413,48
Kigali compound rent	£224,31	per month	27	£6.056,25	£6.056,25	£6.056,25
Communication costs	£2,92	p/p p/w	3*118w	£1.033,68	£1.033,68	£1.033,68
Sub-total costs				£220.896,02	£133.349,67	£144.551,54
overhead 7%				£15.462,72	£9.334,48	£10.118,61
TOTAL COSTS				£236.358,75	£142.684,15	£154.670,14
			pupils	114.174	114.174	114.174
			costs per pupil	£2,07	£1,25	£1,35

Annex 6a Annex staff costs per approach

		budget explanations	A	B	C	salary costs A	salary costs B	salary costs C
		calculation in case we work with 2*8 Sectors for A, B or C. Training 2*8 SEO's is feasible; in that case we work with 2*51 HT's for A or C	days	days	days			
Act 1a 1b: train trainers out of our pool in theory of Educational Leadership and prepare together with them the first workshop for HM + SEOs (5 day WS for 10 educational leadership facilitators i.t.) + 4 preparation days for reflection days HT + SEOs	Preparation TA	5 days + 4*0,5 day	7		7	£2.368	£0	£2.368
	Preparation TL	5 days + 4*0,5 day	7		7	£724	£0	£724
	trainer TA	5+4 days workshop	9		9	£3.044	£0	£3.044
	trainer TL	5+4 days workshop	9		9	£931	£0	£931
	10 educational leadership facilitators	5+4 days * 10	90		90	£3.242	£0	£3.242
	support officer	5+4 days + 2 days invitations, copying, etc. (administrative preparations)	11		11	£447	£0	£447
Act. 2a: training HM + SEO's: Educational leadership, why, how, what competences, what behavior (in 4 groups parallel, each group with facilitator and co-facilitator) (5 day	preparation	preparation was done as part of act. 1	0	0	0			
	trainer TA	5 days workshop	5	5	5	£1.691	£1.691	£1.691
	trainer TL	5 days workshop	5	5	5	£517	£517	£517
	educational leadership facilitators	5 days for 10 ELF's	50		50	£1.801	£0	£1.801
	support officer	5 days at 4 places + 6 days administrative preparations	26	6,5	26	£1.057	£264	£1.057

		budget explanations	A	B	C	salary costs A	salary costs B	salary costs C
workshop)								
Act. 2b: Reflection days HT + SEO's every 6 months, what' progress, what needs to be adjusted/improved (2 days sessions)(4 times)	preparation	preparation was done as part of act. 1	0	0	0			
	trainer TA	4* 2 days	8	8	8	£2.706	£2.706	£2.706
	trainer TL	4* 2 days	8	8	8	£827	£827	£827
	10 educational leadership facilitators	4* 2 days * 10	80		80	£2.882	£0	£2.882
	support officer direct	4* 2 days + 4* 2days administrative preparations	8	2	8	£650	£162	£650
		we worked with two times 8 SEO's = 16 SEO's per group						
Act. 3a: training SEO's: purpose of intervision coaching, theory of coaching with strong focus on practicing interview techniques (1 group of 16) (4 day workshops)	preparation TA	4 times 4 days	16	16	0	£5.411	£5.411	£0
	preparation TL	4 times 4 days	16	16	0	£1.655	£1.655	£0
	preparation expert North	6 days	6	6	0	£3.191	£3.191	£0
	trainer expert North	4 times 4 days	16	16	0	£8.508	£8.508	£0
	trainer TA	4 times 4 days	16	16	0	£5.411	£5.411	£0
	trainer TL	4 times 4 days	16	16	0	£1.655	£1.655	£0
	support officer	4 times 4 days + 4 * 2 days administrative preparations	16	16	0	£975	£975	£0
Act. 3b: coaching SEO's on coaching intervision groups: providing follow up for SEO (5 * per SEO, including attending intervision	coach TA	5* 16 visits = 80; 2 per day makes 40 days. Half of the visits will be in remote places, where driving back can only be the next day → +20 * 0,5 day = 10 days, together 50 days, for TA and L 25 each	25	25	0	£8.455	£8.455	£0
	coach TL		25	25	0	£2.586	£2.586	£0

		budget explanations	A	B	C	salary costs A	salary costs B	salary costs C
coaching session: 1 day incl. traveling)								
Act. 6a, 6b: Conference to start to extend support relevant stakeholders and end conference to present results and recommendations with representatives of REB & Mineduc, Districts, Educational partners, Mott MacDonald, DFID, etc.	preparation TA	2 * 1day	2	2	2	£676	£676	£676
	preparation TL	2 * 1day	2	2	2	£207	£207	£207
	trainer/facilitator TA	2 * 1day	2	2	2	£676	£676	£676
	trainer/facilitator L	2 * 1day	2	2	2	£207	£207	£207
	support officer	2 day + 2* 1,5 days administrative preparations	5	5	5	£203	£203	£203
	programme manager direct	2 * 1day	2	2	2	£641	£641	£641
Act. 8: reporting finance and progress	accountant	42 days per 52 weeks = 96 days per 118 weeks	96	96	96	£6.337	£6.337	£6.337
	TL		9	9	9	£931	£931	£931
	TA direct progress reports	1 day per month	18	18	18	£6.088	£6.088	£6.088
Act. 10 indirect hours from the management	program manager	2 days per month + 10 extra days start efforts in the first half year, makes 64 (=38+26) days	38	38	38	£12.307	£12.307	£12.307
	deputy program manager		26	26	26	£3.239	£3.239	£3.239
	total					£92.247	£75.528	£54.398

Annex 7: Caveats in measuring school leadership

Caveats in measuring school leadership and its effects on quality of education in Rwanda

Table 1: Professional development and coaching trajectory on school leadership (2013-2014)

Time	Activity	Participants[11]
January 2013	Initial input on (eight dimensions of) school leadership + self-assessment	HT from group A and C; SEOs from group A and B Self-assessment by group A, B and C
January-June 2013	Monthly coaching by SEOs through intervision sessions	HT from group A and B, coached by their respective SEOs
July 2013	Reflection workshop on (eight dimensions of) school leadership Self-assessment	HT from group A and C; SEOs from group A and B Self-assessment by group D
July-December 2013	Monthly coaching by SEOs through intervision sessions	HT from group A and B, coached by their respective SEOs
January 2014	Reflection workshop and input on Most Significant Change technique + self-assessment	HT from group A and C; SEOs from group A and B Self-assessment by group A and C
January-June 2014	Monthly coaching by SEOs through intervision sessions	HT from group A and B, coached by their respective SEOs
July 2014	Reflection workshop on (eight dimensions of) school leadership + self-assessment	HT from group A and C; SEOs from group A and B Self-assessment by group A and C
July-December 2014	Monthly coaching by SEOs through intervision sessions	HT from group A and B, coached by their respective SEOs
January 2015	Reflection workshop on (eight dimensions of) school leadership + Stories of significant change + self-assessment	HT from group A and C; SEOs from group A and B Self-assessment by group A, B, C and D (in end line)
January – March 2015	Monthly coaching by SEOs through intervision sessions	HT from group A and B, coached by their respective SEOs

Research questions

- How do HTs assess themselves on different dimensions of school leadership, and which aspects for self-improvement do they identify?
- Is there a significant difference between different groups (according to support modalities – see further details on experimental design)?

[11] A detailed description of the different experimental groups A, B, C and control group D is given in the section 3.1. below.

- How do different experimental groups evolve in their self-assessment over time?

360 degrees feedback and self-assessment tool for head teachers

The 360 degrees feedback and self-assessment tool for HTs was developed in two steps. In a first step, national experts on school leadership in Rwanda, who were involved as training facilitators since the start of the program, were asked to develop indicators of successful school leadership for the 8 dimensions that were identified by Day et. al (2009). In a second step, these indicators were validated by a representation of HTs and SEOs participating in the project. After his validation, a total of 51 indicators remained and rephrased as items for the self-assessment questionnaire tool (see appendix).

The 360 degrees feedback and self-assessment tool has been completed at the start of the project, in January 2013, by 106 HTs as a self-assessment (response rate of 54.08 %), as well as by a selection of teachers of the involved schools and parents with children in these schools. As such, the tool provides a full picture on the capacities of the HTs from the full range (i.e. 360 degrees) of stakeholders involved in the education of the pupils in the involved schools.

Mean scores are used to measure self-assessment of all involved HTs on the different dimensions of school leadership. **Analysis of Variance (ANOVA)** are carried out to compare mean scores of the different experimental groups and to compare group scores over time.

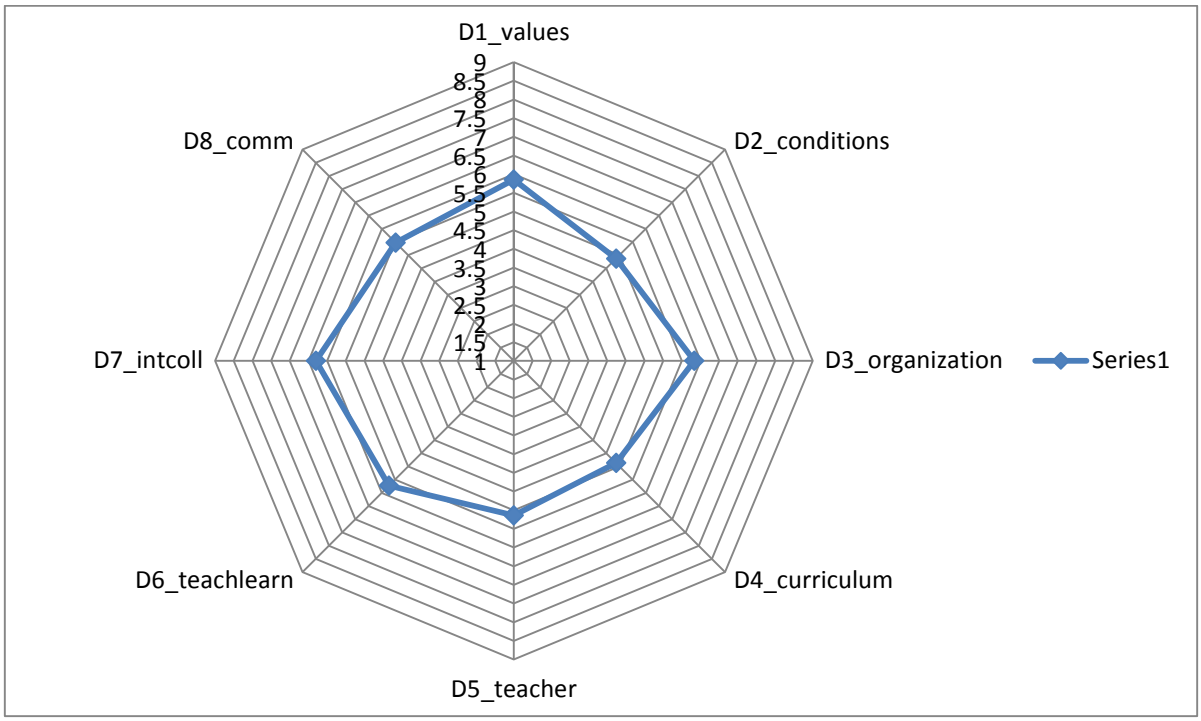
Table 2: Participation of HTs in 360 degrees feedback and self-assessments on eight dimensions of school leadership

	A	B	C	D	Total
jan-13	35	47	44	0	126
jul-13	0	0	0	29	29
jan-14	30	0	26	0	56
jul-14	34	0	50	0	84
feb-15	27	29	51	33	140

To assess the reliability of the instrument, we used the self-assessment data collected from 126 HTs in January 2013. A reliability analysis of the eight scales addressing the different dimensions of school leadership, shows that all subscales are reliable, with Cronbach's Alpha values above .70.

For all of these dimensions, the HTs see much room for personal improvement at the start of the project, especially when it comes to "reshaping conditions for teaching and learning (dimension 2)", "enriching the curriculum (dimension 4)" and "enriching teacher quality (dimension 5)" (see also fig. 2 below).

Fig. 2: Diagram of head teachers' mean scores on eight dimensions of school leadership (January 2013)



Comparing groups

A crucial finding is that already at the start of the project, in January 2013, analysis of variance (ANOVA) shows significant differences in the mean scores on these dimensions between the different experimental groups (see table 3), with group A (trained HTs, coached by trained SEOs) scoring significantly lower than group B and especially group C (trained HTs, no coaching).

Table 3: Comparison of mean scores on dimensions of school leadership (self-assessment) between experimental groups A, B and C (January 2013)

	Jan-2013			
	Group A	Group B	Group C	Sig. (p)
D1_values	5.03	5.73	7.45	<0.01
D2_conditions	3.99	4.60	6.40	<0.01
D3_organization	4.98	5.59	7.42	<0.01
D4_curriculum	4.26	4.49	6.79	<0.01
D5_teacher	4.91	4.54	6.90	<0.01
D6_teachlearn	5.41	5.17	7.46	<0.01
D7_intcoll	6.00	5.82	7.82	<0.01
D8_comm	4.41	5.06	7.53	<0.01
D_Overall	4.87	5.12	7.22	<0.01

While HTs from group A and also B go through a tremendous and significant change in how they assess themselves on the different dimensions, as also illustrated in table 4, this change is non-

existent in group C. However, in February 2015, there are no significant differences in self-assessment over the different experimental groups.

Table 4: Comparison of mean scores on dimensions of school leadership (self-assessment) over time

	Group A				Group B		Group C				Group D	
	A-1-2013	A-1-2014	A-7-2014	A-2-2015	B-1-2013	B-2-2015	C-1-2013	C-1-2014	C-7-2014	C-2-2015	D-7-2013	D-2-2015
D1_direction	5.03	6.93	7.30	8.01	5.73	7.71	7.45	7.28	7.90	7.78	7.59	7.99
D2_conditions	3.99	5.98	6.40	7.24	4.60	7.17	6.40	6.52	6.75	6.83	5.93	7.01
D3_restructuring	4.98	7.21	7.45	7.92	5.59	7.74	7.42	7.40	7.95	7.75	7.30	7.82
D4_curriculum	4.26	6.15	6.54	7.47	4.49	7.23	6.79	6.60	6.98	7.07	6.70	7.39
D5_teacher qua	4.91	6.88	7.23	7.99	4.54	7.60	6.90	7.16	7.46	7.51	7.12	7.46
D6_teach learn	5.41	6.77	7.30	7.88	5.17	7.83	7.46	7.47	7.87	7.71	7.57	7.64
D7_int. collabor	6.00	7.43	7.74	8.18	5.82	8.07	7.82	7.88	8.14	7.85	7.73	7.98
D8_community	4.41	6.93	7.38	8.02	5.06	7.83	7.53	7.55	7.81	7.59	7.48	7.66
D_overall	4.87	6.78	7.17	7.84	5.12	7.65	7.22	7.23	7.61	7.51	7.18	7.62

Better insight in self-assessment scores through focus group interviews

Several assumptions have been made by the project team to explain the disparities above, including:

- ✓ Lack of trust, making head teachers from Group A that is coached by trained SEOs more careful in the assessment of their personal school leadership capacities while on the other hand,
- ✓ Head teachers from group C are scoring significantly higher throughout the first year of the project which can be explained by the fact that they might be less critical towards themselves and show no evolution or reflective progress.

For head teachers of Group B, who have not been trained, but who are being coached by their SEOs, the mean scores on the self-assessment tool are somehow in between the mean scores for Group A and C.

Nevertheless, the above assumptions had to be tested carefully and additional, more qualitative data collected to find out exactly how involved head teachers complete the 360 self-assessment tool, how the tool is used as a capacity development tool and how the use of this tool as well as the relationship between trainees and coaches evolves over time.

To validate or reject the above hypotheses, there was a need to collect more insights from the 360 self-assessment users on how they fill the tool by collecting qualitative data, using Focus Group Interviews.

Two (2) Focus Group Interviews were conducted, each involving four (4) head teachers from Group A and the second comprising four (4) head teachers from Group C. Participants were selected at random from both groups. The interviews took place immediately after the participants completed

the 360 self-assessment tool during the reflection workshop of July 2014. The Focus Group Interviews lasted between 1 hour and 1 hour and 30 minutes and took place away from the training rooms and other participants to ensure quietness and create a climate of confidence.

A set of a few questions to guide the interviews was developed beforehand and we anticipated that more detailed questions would arise during the interviews. The interviews were recorded, after participants signed a consent form guarantying their anonymity and no further repercussions in the use of the information given. The interviews were conducted by a VVOB staff member who is not directly involved in the project and took place in Kinyarwanda. Notes were also taken by the interviewer during the interviews.

Findings

In what follows we highlight some answers from both focus group interviews on the use of the assessment tool and the relationship between the head teacher and their coaches (for group A only).

Question 1: How do you complete the self-assessment on dimensions of school leadership?
Group C: Trained, not coached
Concepts are understood although there might be some vocabulary difficulties (English is not yet mastered by all head teachers).
At the beginning of the self-assessment we thought that if we rate ourselves more critically this might backfire on us. We may appear incompetent and might not receive trainings.
I was a bit confused about the rating scale. There seems some ambiguity. E.g.: I strongly agree has two grades 8 and 9. I feel this gives room for feelings and this makes it more difficult to be totally honest on my performance.
Scoring lower on the self-assessment doesn't mean that we are performing more poorly than the previous time. We are getting a deeper understanding of our limitations and need to improve on them. E.g.: School Library under Dimension 4, if the school has a cupboard in the Accounting Office with a few books I would rate myself higher but once I got to the following item of pupils reading books in the school library, they I realized that a library encompasses more than a shelf of books. This has a positive impact on my next self-assessment, though expressed by a lower score.
With the trainings, use of the tool and interaction with other head teachers, my understanding has improved. I can now relate my daily work and a new situation at my school with one of the dimension to reflect.
The tools has helped me to detect which areas to focus on when developing my personal action plans.
Some head teachers focus on all dimensions at the same time, while others prefer to work on a few dimensions but discover that actually working on one dimension has a positive influence on others.
Some head teachers use the tool as a guideline to develop their action plans, others have used it to appoint teachers and pupils into departments and clubs as well as delegating some of their responsibilities to the school community and parents associations.
I realize that it is not enough to have only one meeting/year with parents and now I hold at least a meeting with them every term. This is also the case for motivating teachers by rewarding them. The school used to reward only teachers in final years on basis of the number of pupils who have succeeded. Now I can link the final success to the path taken by pupils since their start at the school.
The self – assessment tool is extended to the life of the school. The tool allows my teachers to positively criticize my work among themselves and advice on how one can improve their

performance.
Some head teachers still relate their low performance to infrastructure or teachers and thus assess themselves more positively as they should. For example on Dimension 2 (ICT): since they don't have enough computers there is no way they can provide ICT facilities to teachers or pupils. They relate the lack of computers to the fact that they are not connected to the electricity grid and don't reflect on what can be done at their level.
Group A: Trained and coached by SEOs
The scoring scale is not yet well understood by some head teachers: "What makes two head teachers different if one rates himself 3 and the other one 2? What activities have led to 3 not to 2?"
They first thought there was another untold purpose of conducting the self-assessment hence biasing answers given. Once I was back in my school, I realized that there were no repercussions, and I became more comfortable in the following assessments.
The most common mistake was to rate themselves against the school' "ideal situation" rather than assessing the reality. This was later corrected as they got trained and understood the use of the self-assessment tool. I believe that later self-assessments reflect more our skills than the first ones.
With time, some scores went up while others decreased, but this does not mean that we are performing poorly. As they get familiar with the tool, they are more willing to give realistic answers.
Now the self-assessment has a dual purpose for head teachers: first, it allows me to assess the reality (shortcomings, gaps) and on the other hand, it helps me to set objectives of improvement.
The dimensions are tackled by available means by the head teachers: they balance and make priorities depending on the financial means but do also work on other dimensions that do not require immediate "financial" means.
They now know which challenges they can address at their level and ones beyond their abilities. E.g.: if a school is not connected to the national grid, it is understandable to not get the "One laptop per child" and the school administration to not either invest in computers as long as they are not connected. Their reports will be handwritten and they will rate themselves very poorly at ICT question (rate at as low as 1) and indicate in their action plans what they intend to do: working with the Sector for the administration to lobby for electricity.
Head teachers now use the tool as "aide-mémoire" as before the trainings, their objectives were too vague but today the tool helps them to set and stick to their objective , others have displayed the tool in their office while others have designed a self-assessment tool for their teachers using the 360 self-assessment model. I always carry the the tool with me, including when I attend other trainings (not given by VVOB).
The eight dimensions cover all aspects of the school life and are interrelated among themselves. E.g.: You cannot establish external relationship if you have failed to build internal relationship. If I encounter an issue that does not appear in the tool I can know exactly under which dimension it falls.
Head teachers' performance contracts (Imihigo) is developed on the basis of the tool. Used during weekly meetings with teachers as they guide the agenda for pedagogical meetings and resolutions taken reflect the assessment. E.g.: it is now the teachers who develop their timetables of classes.

Question 2: How is your relation with your SEO?

Group C: Trained, not coached
We do not know which channels to use to get our issues to the next levels (lobbying and advocacy). E.g.: I feel that I am not part of the recruiting process. We are obliged to take any teacher sent to us. I would like the maternity leave to be extended to three months but do not know to whom I can present such requests.
SEOs are seen as they don't know their roles in supporting schools as they act as police. Head teachers have developed a cohesion spirit among themselves to be able to cope with SEOs.
On the other hand, the tool has given the opportunity for better relationships within the school community and also with local authorities as their role is clearer to head teachers and to SEOs.
Group A: Trained and coached by SEOs
The agenda of the meetings is known, SEOs send phone messages with it and during intervision sessions, head teachers can add new items in accordance to the needed support.
Now SEOs have an advisory role rather than "police" and their role is more reinforced as understand the role of head teacher as a leader and can do lobbying for schools at higher administrative levels.
Use of head teachers networks, outside intervision sessions and during intervision sessions SEOs intervene on problems that could not be solved at head teachers' level.
Head teachers would like to have more discussions during intervision sessions and develop stronger peer – to – peer networks.

Though we, at the beginning of the training, explained to all participants the intended purpose of the focus group Interviews and how they were selected randomly, and we explained that we guarantee the anonymity of participants, members of Group C (group that is not coached) who participated into the focus group Interviews showed a few signs of mistrust towards the use of the information to be collected. Some participants were reluctant to answer some questions and worried that their names and information given might be disclosed to SEOs and their authorities. Even after further clarifications, they asked for a copy of the Consent Form and they interrupted the interview to raise issues related to anonymity of the answers before responding to some questions such as "How do you score yourself? Is there an evolution in how you score yourself?" For Group A, the interviewer went through the same process of discussing about how the interview will be conducted and the purpose of the Consent Form. The focus group members of group A spontaneously declined the offer to sign the consent form. This can be interpreted as trust in the project and reflects how the relationship between the head teachers and the project has evolved over time.

Conclusions and discussion

We gave a description of the state of mind of HTs at the start and beginnings of the IfE project on coaching school leadership in primary education in Rwanda. In the framework of the project, three experimental groups of HTs have received input on effective school leadership through training and/or coaching by their SEOs. The rationale or theory of change behind the approach is that trained and coached HTs will influence the practice of their teachers and that in turn, teachers will influence learning outcomes of their pupils in primary education.

The findings of this study are limited to an assessment of the self-reported state of mind of HTs. To measure the perception of HTs and teachers involved in the project, a self-assessment tool on school leadership dimensions has been developed. Self-report data is collected from the different experimental groups.

Related to head teacher's perception on their own performance on different dimensions of school leadership, mean scores on these dimensions – as presented in the 360 degrees feedback and self-assessment tool – show that before the start of the project, most HTs perceive room for professional improvement, especially when it comes to "reshaping conditions for teaching and learning (dimension 2)", "enriching the curriculum (dimension 4)" and "enriching teacher quality (dimension 5)". Remarkable however, is the significant difference in how the different experimental groups completed the self-assessment tool. It seems that HTs from group A, who have been trained together with their SEOs and who knew they would be coached by these SEOs, are somehow more self-critical or might feel controlled when completing the self-assessment tool. It might be that there was a lack of trust, making HTs from this group more careful in the assessment of their personal school leadership capacities. In comparison, HTs from group C, who have been trained but do not receive coaching from their SEO, score significantly higher throughout the first year of the project, or in other words, might be less critical towards themselves and show no evolution or reflective process. For HTs of group B, who have not been trained, but who are being coached by their SEOs, the mean scores on the self-assessment tool are somehow in between the mean scores for group A and C.

These findings made clear that the 360 degrees feedback, only used as a self-assessment tool, should not be used as a tool to measure the capacities and training needs of HTs on aspects of school leadership. On the other hand, the tool seems to be useful as a capacity development tool, especially when the conditions are there for continuous self-reflection, as is the case for experimental group A and B. Through focus group interviews with head teachers of group C and A, the different uses of the 360 tool, other than a six month self–assessment tool, shows different degrees of ownership developed by the users. At the same time, the tool can be seen as a barometer of the atmosphere and relationship of trust between a trainee (the head teacher) and her/his coach (the SEO).

Even though a clear cut answer to the different self-assessment scores at the start of the professional development trajectory cannot be given, it is clear that the quantitative data needs to be complemented with qualitative data on how self-assessment took place throughout the trajectory.