



QUEEN RANIA TEACHER ACADEMY
أكاديمية الملكة رانيا لتدريب المعلمين



School Networks for Instructional Improvement in Jordan

IMPACT AND WAYS FORWARD

Table of Contents

List of Acronyms	4
Dedication	5
Preface	6
Executive Summary	8
Key Findings	10
Conclusion	16
Recommendations and Policy Implications	17
Introduction	19
About Queen Rania Teacher Academy	20
Subject Specific Schools Network Program	21
Training Workshops and Reflective Meetings:	21
Onsite Support and Coaching	22
Schools Network Research Study	23
Evaluation Design and Methodology	23
Purpose	23
Design	23
Sample	23
Conceptual Framework	24
Sources of Evidence	25
Teacher Questionnaire	26
School Leader Questionnaire	26
Instructional Training “Coaching” Observation Tool	26
Field Researcher Observation Tool	26
Students’ Achievements Tests	26
Characteristics of Participants	28
Audience and Stakeholders	29
Study Limitations	30
Ethical Considerations of the Study	30
Key Findings	31
SNP Training Quality Satisfaction Rates	31
Change in the Beliefs, Attitudes, and Instructional Practices of Teachers	32
Change in the Beliefs, Attitudes, and Instructional Leadership Practices of School Leaders	37
Change in the Achievement Outcomes of Students	41
Grade Six Students Overall Results	42
Grade Ten Students Overall Results	43
Grade Six Arabic Reading	44
Grade Six Arabic Writing	45
Grade Ten Arabic Reading	46
Grade Ten Arabic Writing	47
Grade Six English Reading	48
Grade Six English Writing	49
Grade Ten English Reading	50

Table of Contents

Grade Ten English Writing	51
Grade Six Mathematics	52
Grade Ten Mathematics	53
Grade Six Science	54
Grade Ten Science	55

List of Acronyms

CPD	Continuous Professional Development
CPRE	Consortium for Policy Research in Education
EGRA	Early Grade Reading Assessment
ERfKE	Education Reform for the Knowledge Economy
FA	Formative assessment
GAC	Global Affairs Canada
HE	High expectations
HRD Strategy	Human Resources Development Strategy
IRT	Item Response Theory
LR	Likelihood Ratio
M	Mean
MOE	Ministry of Education
NAfKE	National Assessment for Knowledge Economy
NCHRD	National Center for Human Resources Development
OISE	Ontario Institute for Studies in Education
PCK	Pedagogical content knowledge
PLCs	Professional learning communities
PISA	Program for International Student Assessment
QRTA	Queen Rania Teacher Academy
RCT	Randomized Controlled Trials
SC	School culture
SNP	Schools Network Program
SNRS	Schools Network Research Study
TBA	Teachers' beliefs and attitudes
TBI	Team-based instruction
TIMSS	Trends in International Mathematics and Science Study
VSA	Various student activities

Dedication

This impact evaluation report was a significant undertaking, accomplished over three years, thanks to the unwavering commitment, perseverance, and diligence of a team of outstanding professionals to whom we dedicate this work.



We express our sincerest gratitude to the following Queen Rania Teacher Academy (QRTA) study program team members:

- Evaluation Study Team, which managed daily activities and follow-ups: Abeer Hakouz, Kamal Lahlouh, Farah Mkhiehir, and Hadeel Al Abed.
- Monitoring and Evaluation Team, which supported data collection and management: Rana Al Khayat and Omar Jatkar.
- Subject Programs Instructional Trainers, who provided training workshops and conducted onsite support and coaching: Dr. Talaat Salem, Mahmoud Abu Jassar, Sohaib Al Omari, Jumanah Jabr, Haneen Arikat, Randa Dawood, Mohammad Bani Issa, Mohammad Najjar, Samia Shawabkeh, Mohammed Alzoubi, Dr. Ibrahim Al-'Emosh, Lubna Aldereiny, Firas Jdeatawi, Salem Mobaideen, and Fayez Abu-Muala.
- Director of Continuous Education, who guided technical operations and supported daily activities: Hala Taher.
- Subject Program Managers, who guided and managed the program's technical operations: Feras Al Omari, Taraf Ghanem, Maysoon Massoud, Dr. Sumaya Al Mohtaseb, Dr. Eman Al Zghoul, and Dr. Sana'a Al Riyahi.
- Field researchers, who provided field support: Abdalrahman Al Agha, Suzan Srou, Bassam Qanadilo, Nada Ababneh, Enas Lahlouh, Nosaiba Eid, Walaa Taif, Sinareet Dedan, Muath Dodakh, Hala Mobaideen, Sandra Nashqwai, Ahmad Jbouri, Aya Madi, Qais Ahmaro, Mohammed Alsoub, and Ahmad Tabazza.



Ontario Institute for Studies in Education:

Research Team, which supported the study design, data analysis, and reporting: Dr. Malcolm Cunningham, Dr. Ruth Childs, and Dr. Kathy Broad.



National Center for Human Resources Development:

Management and Research Team, which managed student field examinations: Dr. Abdullah Ababneh, Dr. Khattab Abu Lebdeh, Dr. Imad Ababneh, and Dr. Ahmad Tweissi.



Ministry of Education (MOE) Team:

- Training Directorate, which supported and facilitated fieldwork: Hafs Abu Mallouh and Thaera Abu Daieh.
- The teachers, principals, assistant principals, education supervisors, and students of 38 schools across three MOE directorates in north/ Irbid, central/ Amman and south/ Karak.

To QRTA management:

Dr. Osama Obeidat, Hala Taher, and Abed Shamlawi: Their belief in the power of monitoring and evaluation enabled the team to provide this evidence-based report to best support the success of teachers' professional development programs - in addition to former QRTA CEO, Haif Bannayan, for his support and efforts in developing and launching this study.

To QRTA support functions:

Project Management Unit, Logistics and Procurement, Transportation, Finance, Human Resources, and Communications Teams, which supported official communications, administration, and logistical follow-ups.

Canada To Global Affairs Canada

whose generous financial support as part of the 'Scaling-Up Teacher Professional Development' project was vital in making this study a reality and enabling us to provide evidence on the impact of subject-specific teaching on student learning.

Preface



In a sixth-grade classroom, students sat at their desks while the Arabic teacher explained the importance of writing. A student raised her hand and asked, “Ms. Sara, why are we learning this?” Surprised by the question, Sara hesitantly replied, “Because writing is important, and you’ll need it in the future.” The student nodded her head and went back to chatting with her classmates, drowning out the teacher’s voice.

Reflecting on her student’s question, Sara headed to the staffroom to find someone with whom she can discuss more effective ways to engage students in writing.

Days later, the interaction was still on Sara’s mind. She recalled Zeina, a quiet teacher who had recently moved to her school and whom Sara had admired for her outlook on teaching and learning. Sara and Zeina spoke for over an hour about different approaches to teaching writing. They agreed to read more about the applications of such approaches, implement them in the classroom, and reflect on those experiences together. This collaboration continued throughout the year, and the more confident Sara and Zeina grew in their research-based teaching, the more tangible the students’ progress became.

Sara and Zeina’s discussions marked the beginning of what is referred to as professional learning communities (PLCs); groups of educators who frequently meet to collaborate, learn, and develop professionally in order to best serve student learning¹. PLCs have proven to be the most effective form of professional development, especially when coupled with a cycle of workshops and reflective meetings that target the needs of teachers’ pedagogical content knowledge (PCK), and skills development.

Sara and Zeina’s case is not unique. In fact, it is common among countless teachers across Jordan who are equally keen to elevate their teaching capabilities. They understand that attending training workshops is not enough to address their needs, and by extension, the needs of their students. For years, teachers around the world have voiced similar concerns, while research has proven that it is not where professional development is provided but its design features that matter most².

Within the local context, there have been multiple efforts to address and redesign the professional development of teachers on a system-wide level. In 2008, the Ministry of Education (MOE) launched a component within its Education Reform for the Knowledge Economy (ERfKE) program that targeted “the quality of all elements of the teaching and learning continuum in order to ensure the achievement of quality learning outcomes for all children.” In recognition of the value that skilled teachers bring to education, Jordan set qualifying and certifying teachers as a key priority within the Human Resources Development (HRD) Strategy launched in September 2016. The strategy addresses Jordan’s need to qualify teachers before they join the profession (i.e., ‘pre-service’) and the necessity to improve the skills of existing teachers through additional training and effective continuous professional development (CPD).

In March 2008, Queen Rania Teacher Academy (QRTA) and the MOE partnered with the Teachers College, Columbia University and the Columbia University Middle East Research Center to propose a scalable project for teacher professional development.

This partnership resulted in the creation and implementation of a high-quality professional development model that supports the use of effective instruction in Jordanian schools. The model is based on establishing and supporting networks of schools as vehicles of change, providing professional development for teams of teachers in core subjects and leadership training for education supervisors, principals, and assistant principals, in order to reinforce the desired changes in classroom practices. The model was also set out as a larger strategy for establishing overlapping school networks. QRTA continued to provide professional development offerings focused on specific subject areas and work together with PLCs that are part of larger networks. At this juncture, QRTA launched its flagship subject-specific professional development program for in-service teachers titled Schools Network Program (SNP).

¹ DuFour and Eaker, 1998.

² Corcoran, 2007; Desimone, 2009; Darling-Hammond et al., 2009; Odden, Archibald, Fermanich, & Gallagher, 2002; Timperly, Wilson, Barrar, and Fung, 2008.

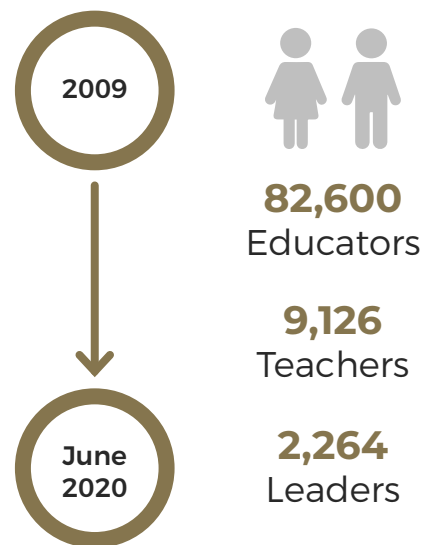
The QRTA SNP connects teachers of the same subjects, creates opportunities for knowledge exchange, enhances peer support, and builds nested professional communities in each school. This model's philosophy requires four main components that, combined, lead to better learning, collaborative and supportive school leadership, teachers committed to learning new skills and the completion of full professional development program activities, ongoing application of newly-acquired practices, and the creation of effective PLCs.



Executive Summary

The Schools Network Research Study (SNRS) impact evaluation report marks the end of the three-year SNRS program. The SNRS addresses the Schools Network Program (SNP) - initiated by Queen Rania Teacher Academy (QRTA) - and investigates its contribution to improving the quality of education in Jordanian schools. The study examines the impact of the SNP program on teaching practices in the four subjects of Arabic, English, Mathematics, and Science, instructional leadership practices, and student learning outcomes in the four subjects in the study schools over three years.

Between 2009 and the end of June 2020, the overall outreach of QRTA exceeded 82,600 educators who benefited from its different programs. In the SNP, 15 cohorts of teachers and school leaders (education supervisors, principals, and assistant principals) participated in professional development workshops. The number of beneficiaries to date reached 9,126 teachers (5,434 females and 3,692 males) and 2,264 leaders (1,287 females and 977 males). Classified by subject, the beneficiaries included 1,625 teachers of Arabic, 2,192 of English, 2,696 of Mathematics, and 2,613 of Science. Given the increasing demand for subject-specific SNP training, QRTA introduced Social Studies as a new network in 2020. The number of beneficiaries until the end of June 2020 amounted to 1,073 teachers (726 females and 347 males) and 100 educational supervisors (30 females and 70 males).



Since the SNP launch in 2009 until 2015, the Consortium for Policy Research in Education (CPRE) at the Teachers College, Columbia University - in collaboration with QRTA - conducted two evaluation studies to measure how the SNP changes the beliefs and practices of teachers and leaders. The first evaluation study, performed in 2011 under the title "Growing School Networks for Instructional Improvement in Jordan, 2009-2010"³, examined the progress and impact of the school network strategy. The second evaluation study, performed in 2015 under the title "Changing Classroom Practices: The Evaluation of the School Network Learning Project in Jordan"⁴, aimed to measure changes in the values and practices of teachers and leaders over time.



Both evaluation studies found positive changes in the instructional practices of teachers and the beliefs and attitudes of teachers and leaders. QRTA leveraged the results to continue enhancing the program to address teacher and leader needs. However, yet to be studied was the extent to which the SNP impacts student learning.



Consequently, in the fall 2016 - under its partnership with Global Affairs Canada (GAC) for a project titled "Scaling-Up Teacher Professional Development", QRTA decided to further study the SNP by investigating the transfer of its impact from teachers to students and whether or not it enhances students' achievements. In response, QRTA launched SNRS in 2016 - in cooperation with the MOE, University of Toronto - Ontario Institute for Studies in Education (OISE), and the National Center for Human Resources Development (NCHRD).

³ Link to the study: https://repository.upenn.edu/cpre_researchreports/61/

⁴ Link to the study: https://repository.upenn.edu/cpre_researchreports/83/

01
 How does the SNP change the beliefs, attitudes, and instructional practices of teachers?

02
 How does the SNP change the beliefs, attitudes, and instructional leadership practices of school leaders?

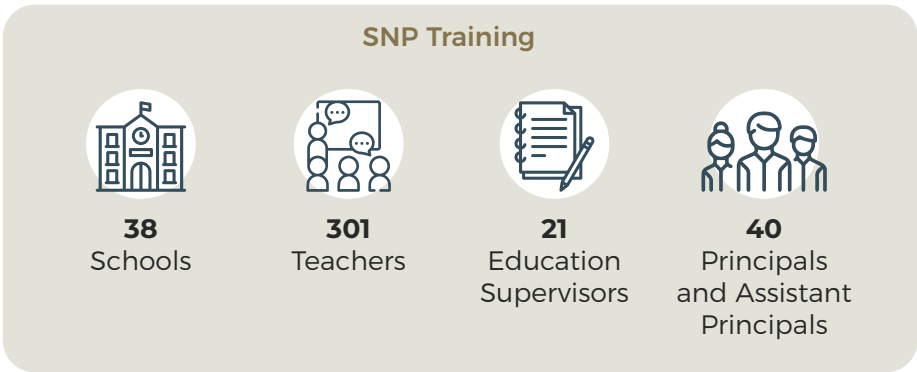
03
 Does the SNP have an impact on the achievement outcomes of students?

The study's underlying assumption suggested that the SNP training model would promote a change not only in the performance of school teachers and leaders, but in the achievements of students as well – provided all SNP components were applied in a structured and efficient manner. Hence, the study focused on three evaluation questions:

The study adopted an experimental design, utilizing elements of the Randomized Controlled Trials (RCT) approach (i.e., treatment and control groups) and the collection and analysis of both quantitative and qualitative data to draw conclusions and findings and propose recommendations.

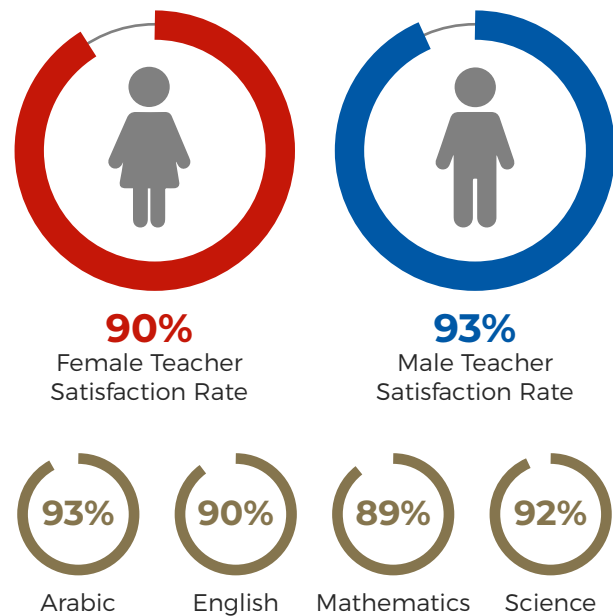
A total of 38 schools in three MOE directorates across the three Jordanian regions participated in the study. These schools were divided into 19 pairs based on similar characteristics, then one school of each pair was randomly selected to undergo the SNP training model. Teachers of the four main subjects (Arabic, English, Mathematics, and Science) and school leaders (educational supervisors, principals, and assistant principals) from each directorate participated in the study. In total, 301 teachers (186 females and 115 males), 21 education supervisors (8 females and 13 males), and 40 principals and assistant principals (26 females and 14 males) benefited from

the SNP training under the scope of this study. The study schools were randomly divided into two groups - namely, the treatment group, which underwent the SNP training, and the control group, which did not undergo the SNP training to serve as a comparison group. Data was gathered over three waves (baseline in fall 2016, mid-term in fall 2018, and post-term in spring 2019), during which a variety of perspectives and sources were collected to enable triangulation and ensure the validity and reliability of the findings. Teachers and school leaders acted as data sources through questionnaires disseminated to both the treatment and control groups. In addition, classroom observations were carried out by QRTA instructional trainers and field researchers in the treatment group schools, and the NCHRD administered students' achievement tests to a sample of sixth and tenth graders in both the treatment and control group schools.



Key Findings

Overall, the evaluation results indicated a high satisfaction rate with the training experience. Classified by gender, the participants' satisfaction rate with the SNP stood at an average of 91% (90% by females and 93% by males). In contrast, satisfaction rates classified by subject were recorded as 93%, 90%, 89% and 92% for Arabic, English, Mathematics, and Science, respectively. The treatment group results under the four aforementioned subjects showed that the subject-specific SNP training had a positive impact on the beliefs, attitudes, and practices of teachers while indicating a positive impact on the students' achievements. Overall, the average scores of students in the different subjects clearly indicated progress when comparing pre-assessment and post-assessment results. A positive change ratio⁵ ranging between 21% - 100% for the sixth graders and 26% - 45% for the tenth graders was found in the treatment group schools. Results also illustrated that the SNP's Instructional Leadership network had a notable positive impact on the beliefs, attitudes, and instructional leadership practices of school leaders.



Change in the Beliefs, Attitudes, and Instructional Practices of Teachers

Overall, the study proved the significant and positive impact of the SNP. The positive changes in the beliefs, attitudes, and instructional practices of teachers among the treatment group schools were spotted in wave three compared to wave one. Meanwhile, the control group schools showed little or no evidence of any changes during the same period. By the end of the study, the treatment group teachers reported a statistically significant positive perception ($p = 0$)⁶ of effective instructional practices, increased collaboration with colleagues, more frequent use of a broader range of instructional strategies, higher confidence in using different instructional strategies ($p = 0.054$), and more positive perceptions of the effectiveness of instructional strategies ($p = 0.045$) compared to the beginning of the study. In contrast, the control group teachers did not report any statistically significant changes in these areas. The analysis showed a statistical significance ($p = 0$) in favor of the treatment group during waves two and three in all areas, indicating the positive impact of the SNP⁷.

The treatment group teachers reported enhanced perception and belief in the importance of effective practices ($M = 79\%$). For example, the teachers demonstrated a growing belief in the importance of improving teaching to help students better acquire the needed knowledge and skills of the subject area. Moreover, their perceptions of their abilities to identify students' misconceptions in the subject area were rated higher.

The treatment group teachers showed a stable or

slight increase in their responses across the three data collection waves, indicating a steady rise in their perceptions of good instruction. In comparison, control group results in this domain remained stable or decreased - except for Science. According to anecdotal feedback received from treatment group teachers across the four subjects, "That their lack of sufficient understanding of the best instructional practices as conveyed in the SNP model prompted them to answer non-objectively the questions of the questionnaire in the first wave; they showed advanced estimates of their skills and themselves. This subsequently led to the convergence of their answers to the questionnaire questions in the second and third waves with those they provided in the first wave." Regardless of the data collection wave, results showed that teachers' perceptions of teaching strategies were positively altered in the treatment group, whereas little or no change was detected in the control group. Overall, the treatment group teachers expressed confidence that they could identify students' misconceptions ($M = 86\%$) and confidence in their own teaching methods ($M = 85\%$). They expressed their strong belief that students could acquire enough knowledge if they tried hard enough ($M = 84\%$) and learn best through practical activities ($M = 88\%$). One item to which teachers in both the treatment and control groups responded similarly was whether teachers at their schools take responsibility for students' achievements ($M = 53\%$) - whereby both groups showed a low percentage of ascribing to this belief.

⁵ The change ratio is calculated as follows: $\frac{[(\text{Post} - \text{Pre})/\text{Pre}] * 100}{1}$. The ratio shows the change percentage compared to the baseline.

⁶ This is the statistical significance "p-value" of paired-sample; t-test of the treatment group responses only in wave one and wave three.

⁷ The full description of findings in this domain is included under the "Key Findings" section

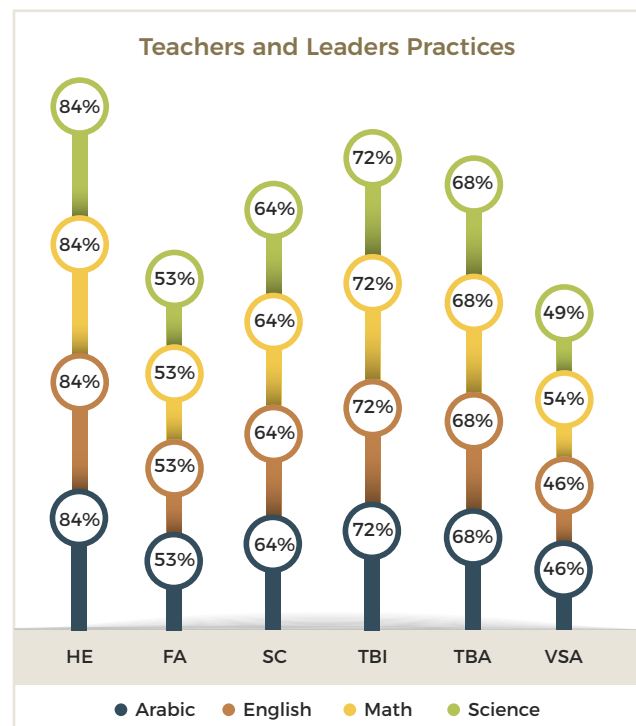
Teachers also reported a positive change in their attitudes towards the targeted instructional practices, including more collaboration with colleagues to plan and discuss lessons and increased use of teaching methods and strategies either on a “daily” or “weekly” basis. Teachers began attending each other’s classes more often and exchanged useful feedback. Moreover, school leaders reported increased confidence in teachers’ attitudes towards instructional practices, including their confidence in teachers’ abilities to design and deliver lesson plans and learning activities for students of all academic levels; leverage new teaching methods and strategies; and use modeling to explain concepts in their respective subject areas.

Besides beliefs and attitudes, teachers demonstrated enhanced instructional practices. In general, teachers who were committed to attending all training activities and were keen on transferring the knowledge they gained to their classroom demonstrated a more positive change in their pedagogical practices. More positive change in the learning environment was observed in schools where the majority of teachers who were engaged in the training program remained in their schools throughout the study duration. This was more evident in the central and northern regions compared to the southern region. In general, observational data and instructional trainers’ and field researchers’ feedback from onsite support visits to schools during the study indicated that central and northern region schools showed slightly more frequent applications of new practices. This is also due to fewer transfers among teachers and school leaders in these regions compared to the southern region. Additionally, this positive change was more evident in girls’ schools than boys’ - except for two boys’ schools⁸ in the central region that showed a very high positive change⁹. Generally, teachers and leaders at girls’ schools exhibited a higher commitment to attending all training activities and encouraged PLCs. Most female teachers transferred the newly acquired practices more frequently to classrooms compared to their male counterparts. These reasons are believed to be behind girls’ schools outperforming boys’ schools.

Furthermore, teachers reported sufficient levels of implementation of the newly acquired practices. Qualitative¹⁰ data analysis of the teachers’ questionnaires showed that most treatment group teachers used different teaching practices, while almost half of the control group teachers said they did not use different teaching practices in lecturing their subjects. Classroom observations¹¹ by instructional trainers and field researchers served as substantial evidence of the behavior of teachers and leaders in waves two and three. Results of observation analysis showed a statistical significance as follows: Arabic ($p = 1e-3$), English ($p = 3e-2$), and Science ($p = 5e-5$), while the statistical significance of Mathematics was not calculated due to the change of trainers for reasons

outside of the program team’s control.

On average, the most frequently reported levels of implementation of the various instructional practices were high across the four subjects. For example, teachers’ average use of various student activities (VSA) during lessons was 46% for Arabic and English, 49% for Science, and 54% for Mathematics. Teachers’ beliefs, attitudes (TBA), and ability to set expectations was 68%. They were able to identify areas and steps for self-development; describe changes made in their practices over time; and connect the changes to both student and professional self-learning. Regarding the indicator on whether or not teachers promote team-based instruction (TBI) and student-centered discussions, the result was 72%. This investigated, for example, if students know how to work in pairs or groups; can easily assign responsibility in order to accomplish tasks; and work with and listen to each other. Evidence on positive school culture (SC) and PLCs reached 64%, as principals and teachers talked about teaching and learning practices at their schools, and teachers discussed working together to plan lessons and visited each other’s classes. Teachers’ use of formative assessment (FA)/adaptive teaching/check for understanding was evident at 53%, as teachers conversed with students, asking them to describe their thinking and adapted lessons in response. Moreover, teachers showed high expectations (HE) of their students and gave them rigorous tasks and clear directions during lessons 84% of the time.



⁸ More description on these two boys’ schools is provided in the leaders’ analysis results section.

⁹ This presents a prospect for future study, probably qualitative, to better understand the factors leading to this high positive change.

¹⁰ This data is from the open-ended questions in the questionnaires.

¹¹ Observational data are nominal since trainers and trainees recorded a given teacher’s behavior as either present or absent. Kruskal-Wallis tests were applied to the observational data to test the null hypothesis that grouping does not result in differences in coaches’ and trainees’ classroom observations. The Kruskal-Wallis test is appropriate because it determines whether or not the population distributions are the same without assuming normal distributions.



Change in the Beliefs, Attitudes, and Instructional Leadership Practices of School Leaders

Overall, the study proved a significant and positive SNP impact. Although slightly less than teachers, leaders among the treatment group schools demonstrated a positive change in their beliefs, attitudes, and instructional leadership practices in waves one and three. The training delivery model of the Instructional Leadership network focuses on raising awareness of the best practices in instructional leadership and encouraging leaders to adopt them. Unlike the subject-specific networks - except for learning walks and attending classroom observations with QRTA instructional trainers - the Instructional Leadership network does not include other onsite support activities. As leaders are often busy with the administrative workload at their schools, their ability to fully implement the instructional leadership practices as expected is adversely affected. This factor explains why the impact on leaders fell behind that of teachers. Still, the control group leaders showed little or no evidence of change over the same period, while the treatment group leaders reported statistically insignificant ($p = 0.945$) positive perceptions pertaining to effective instructional practices, increased collaboration among teachers, and more frequent use of the different instructional strategies among teachers. The positive change in leaders' perceptions regarding their confidence in teachers' use of targeted instructional strategies ($p = 0.62$) or the effectiveness of these instructional strategies ($p = 0.23$) was not statistically significant at the end of the study compared to the beginning. However, the control group leaders did not report any significant changes occurring in these areas.

Treatment group teachers reported significantly more positive perceptions ($p = 0$) on the instructional leadership practices of principals and assistant principals by the end of the study compared to the beginning, whereas control group teachers reported little or no change in their beliefs.

Treatment group leaders reported positive perceptions and beliefs in the importance of effective practices among their teachers ($M = 75\%$), and results were aligned with those reported by the teachers. For example, leaders demonstrated a growing belief in the importance of enhancing teaching to help students better acquire a subject area's knowledge and skills. Moreover, their perception of the teachers' abilities to identify students' misconceptions in the subject area was improved.

Leaders reported positive feedback on their teachers' attitudes towards the targeted instructional practices. For example, they stated that teachers started to collaborate better with colleagues to plan and discuss lessons, teaching methods, and strategies. They also reported that teachers began to attend each other's classes more often and exchange useful feedback. Meanwhile, teachers conveyed improved attitudes towards their leaders' instructional leadership

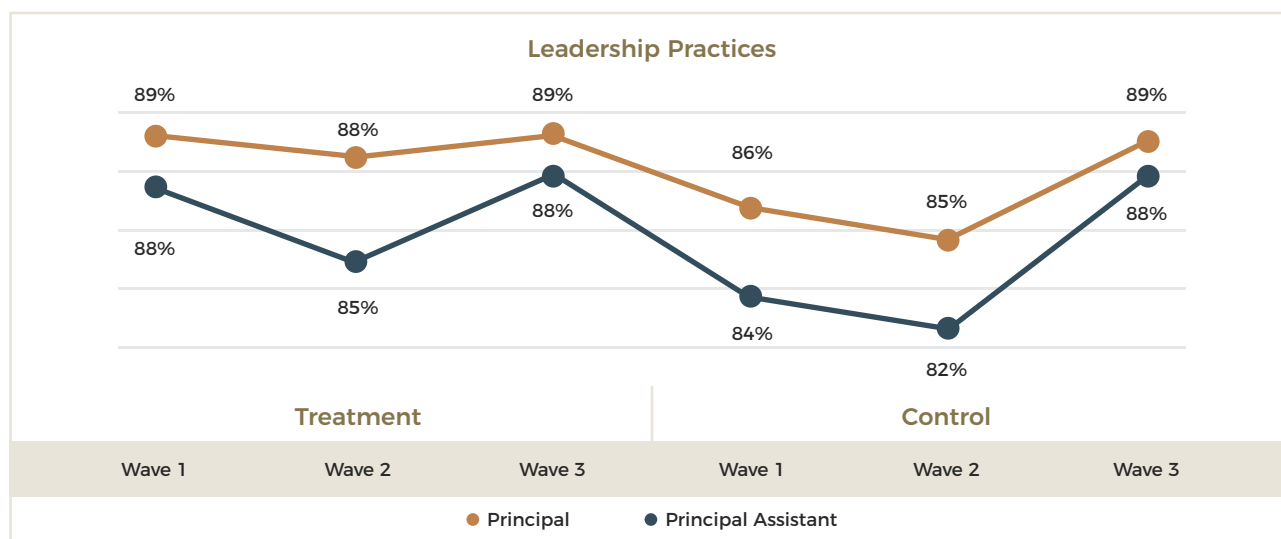
practices. For example, they noted that leaders started encouraging teachers to improve their teaching practices; showing more respect to the teachers' experiences; further appreciating their good work and meticulous follow up on the academic progress of students; accepting constructive feedback from teachers and taking actions accordingly; and calling on teachers to play a meaningful role in setting school development objectives.

In addition to beliefs and attitudes, leaders showed improvement in their instructional leadership practices. Generally, leaders who were committed to attending the training activities or recognized the expected benefits of the SNP were characterized as being more supportive, collaborative, and motivational by their teachers. This encouraged teachers to apply the new skills and knowledge they acquired, eventually showing a more positive change. This trend was generally more evident in girls' schools than boys' schools, except at two boys' schools in the central region, which demonstrated a significantly high positive change due to their leaders' support and encouragement. The principals and assistant principals in these two schools were highly cooperative; they embraced positive views and the potential impact of the SNP. They also facilitated and followed up on their teachers' attendance at the training activities and encouraged them to apply the new teaching practices in the classroom.

Leaders indicated a stable trend in their perceptions of their teachers' instructional capacities. The responses collected throughout the three data collection waves of the treatment group remained steady, whereas the control group showed a slight increase in wave three compared to wave one. Nonetheless, the highest value that the control group scored was similar to that of the treatment group. According to anecdotal feedback received from the treatment group leaders, *"That their lack of sufficient understanding of the best practices in educational leadership prompted them to answer non-objectively the questions of the questionnaire in the first wave; they showed advanced estimates of their skills and themselves. This subsequently led to the convergence of their answers to the questionnaire questions in the second and third waves with those they provided in the first wave."*

The treatment group leaders showed more stable trends in their instructional leadership practices throughout the three data collection waves than those of the control group. In wave one, the treatment group started at a slightly higher baseline, yet this trend was maintained throughout the second and third waves. Alternatively, the control group started at a lower baseline, later showing a fluctuating trend in wave two (decrease) and wave three (increase), which can be attributed to a time-related change and other factors unrelated to the SNP program.

Overall, the analysis showed statistical significance among groups in favor of the treatment group in wave three ($p = 0.002$), while no statistical significance was spotted at the level of the treatment group itself between wave one and wave three ($p = 0.877$) - a fact that supports the partial positive impact of the SNP.



Change in the Achievement Outcomes of Students:

The results attained by students on the standardized tests¹² conducted for both the sixth and tenth graders indicate a higher positive difference in the treatment group compared to the control group across all four subjects (i.e., Arabic, English, Mathematics, and Science).



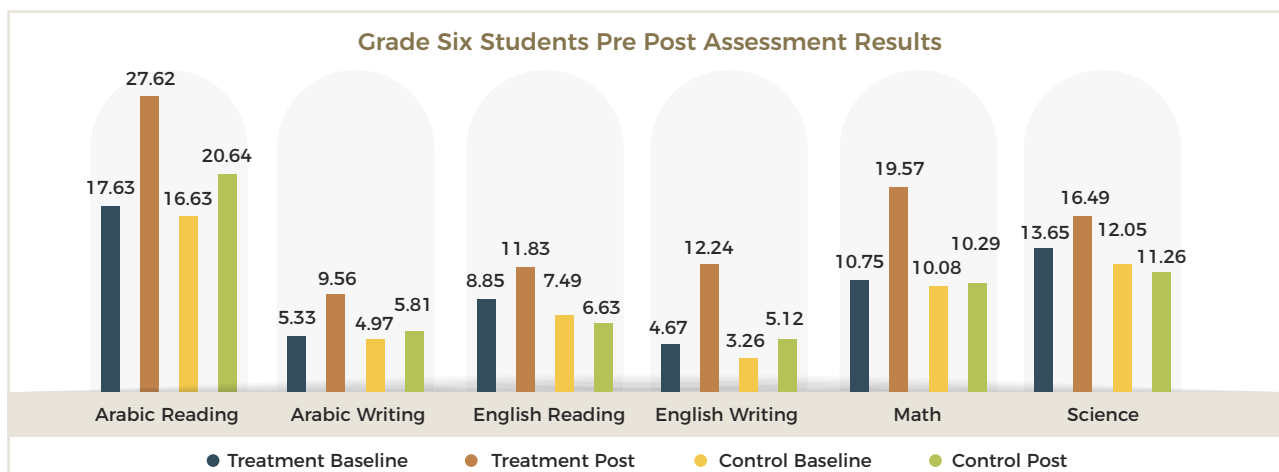
Grade Six Students' Overall Results:

On average, grade six students' scores across all schools increased in wave three (post-assessment) compared to wave one (pre-assessment) for all subjects - except for English Reading in one of the treatment group schools in the southern region. The average scores attained on the various subjects indicated a significant difference in progression from wave one to wave three between the treatment and control group schools. The treatment group schools' overall averages increased, unlike those of the control group.

The highest increase in the treatment group schools was in English Writing, followed by Mathematics, Arabic Writing, Arabic Reading, English Reading, and Science. In the control group schools, English Writing also showed the highest increase, followed by Arabic Reading and Writing, which exhibited a moderate increase. In contrast, the Mathematics increase was minimal, whereas the English Reading and Science results decreased.

Grade Six	Arabic Reading	Arabic Writing	English Reading	English Writing	Math	Science
Change ratio ¹³ in Treatment	57%	79%	34%	100% ¹⁴	82%	21%
Change ratio in Control	24%	17%	-11%	57%	2%	-7%

¹² Students of wave three are different from those examined in wave one. The rationale behind this change is the SNRS's aim to measure the impact of improving the teachers' and leaders' practices on the students' achievements in schools regardless of identities. Teachers are expected to apply the new teaching methods with all students from different grades and should work gradually to upscale learning as they teach higher grades.



This high positive change in sixth-grade student achievements in wave three revealed a high statistically significant difference in Arabic Writing, English Reading, English Writing, Mathematics, and Science (all standing at $p = 0$) in favor of the treatment group. Conversely, no statistically significant difference was spotted in Arabic Reading ($p = 3.06e-1$). Moving forward from wave one to three, this positive change was evident across all four subjects in the treatment group, whose sixth-graders scored a statistically significant higher average in wave three compared to

students in the same group during wave one. These differences were registered across all four subjects - namely Arabic Reading, Arabic Writing, English Reading, English Writing, Mathematics, and Science (all standing at $p = 0$). Some of the sixth graders of the control group schools showed increase in the achievement scores from wave one to three, but such increases were substantially and consistently lower in the control group schools and not statistically significant.

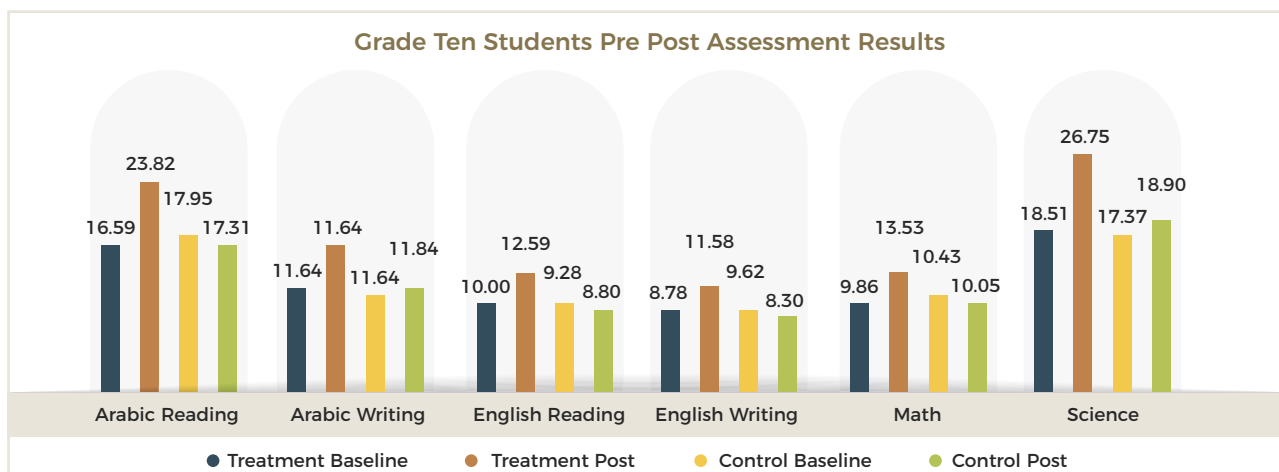


Grade Ten Students' Overall Results:

On average, grade ten students' scores across all schools increased in wave three (post-assessment) compared to wave one (pre-assessment) for all subjects. The average scores attained on the various subjects indicated a significant difference in progression from wave one to wave three between the treatment and control groups. The treatment group schools' overall averages increased in all subjects, unlike those of the control group. The highest increase in the treatment

group schools was in Science, followed by Arabic Reading, Arabic Writing, Mathematics, English Writing, and English Reading. In the control group schools, Arabic Writing showed the highest increase, which was moderate, followed by a small increase in Science. At the same time, Arabic Reading, English Reading, English Writing and Mathematics results decreased.

Grade Ten	Arabic Reading	Arabic Writing	English Reading	English Writing	Math	Science
Change ratio in Treatment	44%	40%	26%	32%	37%	45%
Change ratio in Control	-4%	12%	-5%	-14%	-4%	9%



¹³ Change ratio is calculated as: $\frac{((\text{Post} - \text{Pre})/\text{Pre}) * 100}$. This ratio is used to show change percentage compared to baseline.

¹⁴ Actual ratio is 162%, however, to simplify the message it is presented here 100%.

This high positive change in tenth-grade student achievements in wave three revealed a statistically significant difference in Arabic Reading, Arabic Writing, English Writing, and Science (all standing at $p = 0$) in favor of the treatment group. In contrast, no statistically significant difference was spotted in English Reading ($p = 8.3e-2$) or Mathematics ($p = 4.83e-1$). Moving forward from wave one to three, this positive change was also evident across all four subjects in the treatment group itself, whereby the tenth graders of the treatment group schools scored a statistically significant higher average in wave three compared to students of the same group in wave one for all subjects, except for one - i.e., Arabic Reading, Arabic Writing, English Writing, and Science (all standing at $p = 0$), and Mathematics ($p = 7e-3$). In contrast, the treatment group students who took the English Reading tests did not show a statistically significant difference in their achievement scores between wave one and three ($p = 8.3e-2$). Some of the tenth graders of the control group schools showed increases in the achievement scores from wave one to three, but such increases were substantially and consistently lower in the control group schools and not statistically significant.

For both grades six and ten, the directorates constituted a significant source of variation, as the

northern and central regions scored higher in various subjects. Gender was also a substantial source of variation, as female students achieved higher scores than their male counterparts across all four subjects for grade six. Meanwhile, female students in grade ten achieved higher scores than their male counterparts in all subjects - except for Mathematics. English Writing results showed no statistically significant differences ($p = 1.22e-1$) even with female students scoring higher than males. Reasons behind this variation are expected to be due to the higher level of commitment that participants in northern and central regions showed compared to the south and the higher level of commitment that female participants displayed compared to males in attending all the training program activities. The greater and more frequent use of the newly acquired skills inside the classrooms - which the instructional trainers observed more in northern and central regions, particularly in girls' schools - was another main factor. Although there was no strong evidence on the effectiveness of PLCs, instructional trainers' ongoing visits to schools throughout the course of the study showed that PLCs were, in general, more robust in northern and central regions compared to the south and more substantial in girls' schools compared to boys' schools.



Conclusion

Reference to the main purpose of the SNRS, which aimed at investigating the contribution of the QRTA SNP to the improvement of education in Jordanian schools, the following conclusions can be drawn from the detailed findings of this evaluation:



Change in the Beliefs, Attitudes, and Instructional Practices of Teachers: The SNRS provided solid evidence that the SNP had a positive effect on the beliefs, attitudes, and targeted instructional practices of subject teachers throughout the three-year study.



Change in the Beliefs, Attitudes, and Instructional Leadership Practices of School Leaders: The SNRS provided some evidence, although not as strong, that the SNP had a positive effect on the beliefs, attitudes, and the targeted instructional leadership practices of school leaders throughout the three-year study.



Change in the Achievement Outcomes of Students: The SNRS provided strong and compelling evidence that the SNP had a positive effect on the achievement outcomes of students across all four subjects due to the improved beliefs, attitudes, and instructional practices of teachers and school leaders throughout the three-year study.

The evaluation results of this meticulously designed study revealed that the SNP had a very positive impact on teachers' beliefs, attitudes, and instructional practices and the improved achievement outcomes of the sixth and tenth-grade students. Moreover, the study showed that the SNP had some positive effects on school leaders' instructional leadership practices. As indicated by the study, the quality of the QRTA SNP training delivery was high.

Based on these results, the educational community and decision-makers in Jordan can validate that students whose teachers and school leaders participated in the SNP and transferred their learning into practice in their schools outperformed their counterparts, whose teachers and school leaders did not participate in the program. Moreover, the SNP participants demonstrated consistent and continued positive change in instructional practices and leadership while creating a more interactive and engaging learning environment for students than teachers and school leaders who did not participate in the SNP.



Recommendations and Policy Implications

The SNRS presented compelling evidence that the SNP had a positive effect on education quality in schools. This was manifested in improving the beliefs, attitudes, and practices of teachers, leaders to some extent, and the achievement outcomes of students - in addition to the vast accumulated knowledge earned by the evaluation and program teams during the study that directly fed into revising the SNP for better outcomes. Accordingly, this report proposes the following set of recommendations:



Subject-Specific Professional Development of Teachers:

Teachers who are competent in their pedagogical content knowledge (PCK) are more likely to positively impact their students' learning process and progress. This was evident across the study's four designated subjects, whereby results revealed the positive impact of the subject-specific pedagogy training on educational capacity-building in the specific subject area. As such, the study findings suggest endorsing the SNP for teachers' professional development, and:

- Expanding the subject-specific SNP to accommodate more teachers in the four respective subjects, i.e., Arabic, English, Math, and Science - a step that is expected to better support student learning and enhance the learning outcomes of Jordan's educational system.
- Expanding the SNP model to include more subjects, such as Social Studies¹⁵.
- Proposing the SNP as a core program for teachers' professional development and integrating it into teachers' career paths.



Instructional Leadership Professional Development of Leaders:

Supportive and instructionally capable school leaders will more likely positively impact teaching practices and the learning environment than less cooperative, instructionally incapable leaders. Based on the study findings, schools with cooperative and supportive leaders, who completed the entire Instructional Leadership program, as well as those who remained at their school over the study timeframe demonstrated better results than schools with changing and/or uncooperative leadership. Hence, the study recommends:

- Expand the Instructional Leadership program to accommodate more school leaders, a step that is expected to promote better learning outcomes at schools.
- Boosting leaders' participation in Instructional Leadership programs in order to better equip them with the skills, knowledge, and practical experiences of instructional leadership that ultimately promote effective teaching and learning in schools.
- Facilitating more opportunities for school leaders to learn about the teachers' subject-specific SNP in the leaders' Instructional Leadership program, thus helping them better understand the new practices that teachers need to perform to enhance the student learning process. Throughout this study, there was evident variation in the levels of knowledge in this aspect among educational supervisors who were introduced to the subject-specific SNP as opposed to the principals and assistant principals who were not. This prompted the study team to include reflective meetings under the Instructional Leadership program to introduce principals and assistant principals to the teachers' training. This step proved to be of added value to the program.
- Including more onsite activities to support leaders in schools, strengthen school leader networking, and build a practical implementation model of the instructional leadership practices in real-life situations. Currently, onsite support for school leaders is limited to a learning walk, joining subject-teacher coaches during their classroom observations, and facilitating PLCs. This form of support proved insufficient to engage school leaders in effective and lasting instructional leadership practices. In contrast, the positive effect of the onsite support on the SNP teachers was evident.
- Including instructional leadership as an essential element of leaders' professional development and integrating it into their career paths.

¹⁵ QRTA in collaboration with the MOE launched a Social Studies network in 2020.



Professional Learning Communities:

The existence of a collaborative culture and ongoing practices in the form of PLCs on both the school and directorate levels would improve collaboration and exchange of expertise. PLCs remain untapped and require further support to ensure their effective implementation through a well-established structure and guiding principles. Therefore, the study recommends:

- Institutionalizing the PLC culture in schools by incorporating them into the school schedule; building capacity within schools to lead effective PLCs; and monitoring and supporting PLC activities through education supervisors and school principals.
- Considering more flexible regulations for teachers and leaders to meet frequently with their colleagues at the school and directorate levels.
- Encouraging and supporting school leaders to utilize PLCs as a vehicle to building teachers' capacities and including them in school development plans, alongside the resources required to enable their implementation.
- Creating an online platform that serves as a virtual PLC for teachers and school leaders to exchange knowledge, expertise, and innovative ideas.



SNP's Arabic and English Writing:

Despite the significant positive changes that the Arabic and English programs had on enhancing students' writing skills, teachers faced several challenges with the writing component. Integrating it into the curriculum required additional time and effort - over and above teaching the existing curriculum - to implement the structured methodology presented in the network programs that seek to build students' writing skills possible. Furthermore, inadequate writing skills among students in the middle and higher grades added to the time required to enhance this capability. Therefore, the study recommends:

- Integrating the writing methodology presented in the network programs, which has proven successful, into the existing English and Arabic curricula to gradually and subtly hone students' writing skills, starting from the early grades.



Onsite Support and Coaching:

The ongoing onsite support provided to schools appeared to be among the key factors that facilitated the program's success. Coaching visits contributed to increased commitment among teachers and leaders, enabling teachers to grasp and discuss effective applications of the new teaching strategies modeled by trainers in real-life classroom situations while inspiring them to apply these new teaching practices. Hence, the study recommends:

- Planning more coaching activities in the professional development programs for both subject-specific and Instructional Leadership programs, eventually contributing to better implementation of new practices.

With these recommendations in mind, we summarize by calling for mobilization to scale up and expand the SNP to include more schools across in Jordan and additional subjects to impact students' achievement positively across different grade levels.



Introduction

In recognition of the importance and value of skilled teachers to education, Jordan described the challenge of qualifying and certifying teachers as a key priority within its ongoing education reform efforts, one of which is the Human Resources Development (HRD) Strategy launched in September 2016. The strategy addresses Jordan's need to qualify teachers before they join the profession (i.e., "pre-service") and the necessity to improve the skills of existing teachers through additional training and new effective continuous professional development (CPD).

Over the last few decades, there has been consistent improvement in educational access and provision in Jordan; however, student learning outcomes at all levels remain a major concern. At the primary level, according to the Early Grade Reading Assessment (EGRA), only around 20% of second- and third-grade students achieve the basic level of reading comprehension.^{16 17 18} Regarding Jordan's results for the secondary level (eighth grade), the Trends in International Mathematics and Science Study (TIMSS) results in 2011 showed a significant decrease since 2007 and were even lower than in 1999.^{19 20}

Jordan's efforts towards education reform show signs of positive progress in students' learning. For example, in the 2018^{21 22} Program for International Student Assessment (PISA), Jordan displayed improvements in students' scores in all three subjects (Reading, Mathematics, and Science) compared to previous PISA rounds since 2006. Consequently, sustaining education reform efforts and investing in building teachers' and leaders' pedagogical knowledge and practices remain important factors to maintain this progress in students' achievements and build further upon it. Moreover, because schools represent vehicles of change within communities, it is through them that prospects for future generations of Jordanian children must improve. For schools to fulfill their duty and meet the Kingdom's expectations, they need qualified teachers and leaders who can uplift education in Jordan to meet international performance standards. Educators²³ with improved pedagogical knowledge and practices and enhanced instructional and leadership skills are called upon to serve as inspiring leaders who stimulate and motivate positive change in their schools.



¹⁶ Alghad (2013) <http://www.alghad.com/articles/565944>

¹⁷ RTI International (2012) Executive Summary: Student Performance in Reading and Mathematics, Pedagogic Practice, and School Management in Jordan, prepared for USAID/Jordan.

¹⁸ QRTA & ALECSO (2014) Teacher Education and Recruitment Policies: The Case of Jordan.

¹⁹ Beatty, A., & Pritchett, L. (2012) From Schooling Goals to Learning Goals: How Fast Can Student Learning Improve? CDG Policy Paper O12, Center for Global Development, Washington, DC.

²⁰ TIMSS & PIRLS International Study Center. (2012) TIMSS 2011 International Results in Mathematics.

²¹ <https://www.oecd.org/pisa/PISA%202018%20Insights%20and%20Interpretations%20FINAL%20PDF.pdf>.

²² <https://gpseducation.oecd.org/CountryProfile?primaryCountry=JOR&treshold=10&topic=PI>.

²³ Educators include: Teachers, supervisors, principals and assistant principals.

About Queen Rania Teacher Academy

Queen Rania Teacher Academy (QRTA) is an independent non-profit organization committed to the vision of Her Majesty Queen Rania Al Abdullah of supporting and empowering educators with the skills, recognition, and support necessary to excel in their classrooms. QRTA was launched in 2009 under the patronage of Her Majesty Queen Rania Al Abdullah and in partnership with the Teachers College, Columbia University and the Jordanian Ministry of Education (MOE). QRTA aspires to contribute to the empowerment of educators and equip them with the knowledge and skills to positively influence future generations of Jordanian and Arab children

by spearheading efforts in teacher professional development.

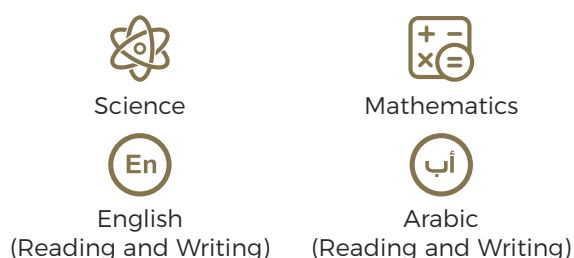
Since its establishment, QRTA has worked diligently to improve teaching quality in the Jordanian education system and stimulate educational advancements across the Arab world. It has strived to carry out its mission by expanding its reach across Jordan and the region, tackling development issues and introducing quality CPD opportunities for teachers. QRTA has gained recognition as a change agent working to enhance education in the region, with an overall outreach exceeding 82,600 educators who benefited from its various programs.



Subject Specific Schools Network Program

The partnership between QRTA, the Teachers College, Columbia University, and the MOE resulted in devising and implementing a high-quality professional development model that supports the use of effective instruction in Jordan's schools. The model is based on establishing and supporting networks of schools as vehicles, providing professional development for teams of teachers in core subjects, as well as leadership training for principals, assistant principals, and education supervisors, in order to support the desired changes in classroom practices.

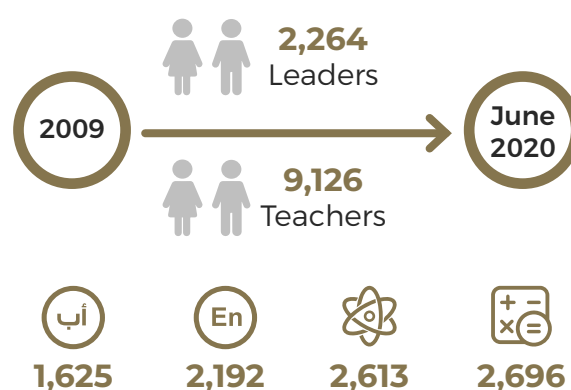
The QRTA SNP is subject-specific in:



In addition, it incorporates instructional leadership training for school principals, assistant principals, and directorate supervisors, given that effective instructional leaders ensure teachers' quality performances and the delivery of the highest quality instruction every day. The SNP connects teachers in the same disciplines, creates opportunities for knowledge exchange, enhances peer support, and builds nested PLCs. To achieve successful results, the philosophy of the SNP model requires the availability of collaborative and supportive school leadership, teachers committed to learning new skills and completing the full professional development program, ongoing application of the newly acquired practices, and the creation of effective PLCs to exchange knowledge and expertise.

Since 2009 and until the end of June 2020, 15 cohorts of teachers and school leaders (education supervisors, principals, and assistant principals) have participated in professional development workshops through the SNP. The number of beneficiaries to date

reached 9,126 teachers (5,434 females and 3,692 males) and 2,264 leaders (1,287 females and 977 males). Classified by subject, the beneficiary teachers included 1,625 teachers of Arabic, 2,192 of English, 2,696 of Mathematics, and 2,613 of Science. Given the increasing demand for subject-specific SNP training, QRTA introduced Social Studies as a new network in 2020. The numbers of beneficiary teachers and educational supervisors until the end of June 2020 amounted to 1,073 (726 females and 347 males) and 100 (30 females and 70 males), respectively.



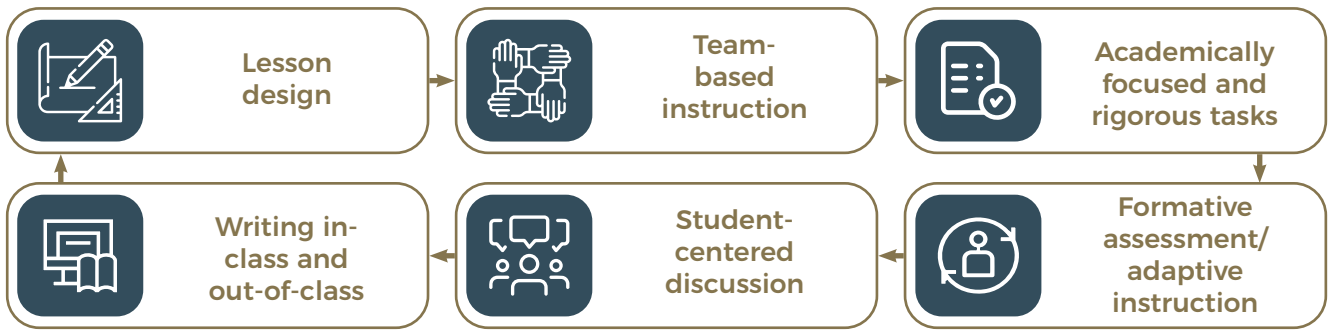
The SNP is unique in its structure for not being a one-off educational experience but rather a continuum of practical professional development activities to scaffold teachers' knowledge and skills over two years. Each network usually encompasses 15 to 20 schools within a given geographic area and is designed to address a specific content area. The SNP is intended to promote changes in teachers' and school leaders' beliefs, attitudes, and practices, which will lead to positive classroom changes, as demonstrated through students' achievement gains. The program offers regular professional development workshops, onsite support, and opportunities for collaboration across schools. The structure of the SNP consists of:



Training Workshops and Reflective Meetings:

Foundational to altering teachers' and school leaders' beliefs, attitudes, and instructional practices, the SNP comprises a series of embedded workshops and reflective sessions that focus on building teachers' capacities in the six "Big Effective" instructional practices. These are lesson design, team-based instruction (TBI), academically focused and rigorous tasks, formative assessment/adaptive instruction, student-centered discussion, and writing in-class and out-of-class. Teachers attend six workshops and six reflective meetings during their training, while school leaders attend six workshops on instructional leadership. In this study, in particular, four reflective sessions²⁴ were added to the Instructional Leadership program to build a better understanding among leaders about the SNP and SNRS.

²⁴ Details of workshop topics are presented in annex 1.



Onsite Support and Coaching:

During the program, QRTA instructional trainers followed up with trainees through monthly visits per subject per school, allowing them to gauge what sense teachers and school leaders made of their workshop experiences, assess how their instructional practices and leadership evolved, and support their professional development. Instructional trainers observed lessons in progress, monitored professional growth, and provided support and advice to teachers and leaders. Thus, information derived from onsite school interactions with teachers and school leaders contextualized essential ideas and skills introduced in the QRTA workshops. Onsite visits provided important opportunities for QRTA instructional trainers to encourage and support incremental changes in practice by engaging educators to consider pedagogical questions relevant to their daily practices.



Schools Network Research Study



Since the SNP launch in 2009 and throughout its implementation until 2015, the Consortium for Policy Research (CPRE) in Education at the Teachers College, Columbia University - in collaboration with QRTA - conducted two evaluation studies to measure how the SNP changes the beliefs and practices of teachers and leaders. The first evaluation study, performed in 2011 under the title "Growing School Networks for Instructional Improvement in Jordan, 2009-2010,"²⁵ examined the progress and impact of the school network strategy. The second evaluation study, performed in 2015 under the title "Changing Classroom Practices: The Evaluation of the School Network Learning Project in Jordan"²⁶, aimed to measure changes in the values and practices of teachers and leaders over time.



Both evaluation studies found positive changes in the instructional practices of teachers and the beliefs and attitudes of teachers and school leaders. QRTA widely used the results to continue enhancing the program to address teacher and leader needs. However, the extent to which the SNP impacted student learning was not yet studied.



Consequently - in fall 2016 - under its partnership with Global Affairs Canada (GAC) for a project titled "Scaling-Up Teacher Professional Development", QRTA decided to further study the SNP by investigating the transfer of its impact from teachers to students and whether or not it enhances students' achievement. In response, QRTA launched SNRS in 2016 - in cooperation with the MOE, University of Toronto, Ontario Institute for Studies in Education (OISE), and the National Center for Human Resources Development (NCHRD).

Evaluation Design and Methodology



Purpose

To measure the impact of QRTA's SNP training model on changing teachers' and leaders' performances and how this affects students' achievement. In addition, to investigate a holistic and integrative school-wide improvement program in the core subjects of Science, Mathematics, and English and Arabic literacy, as well as instructional leadership.



Design

The SNRS used an experimental design approach for the evaluation to measure the impact of leaders' and teachers' professional development on students' achievements by utilizing Randomized Controlled Trials (RCT) using a two-group pre-post design²⁷ approach (i.e., treatment and control groups). The SNP was delivered to schools in the treatment group, whereas schools in the control group did not receive the SNP and were used for comparison purposes only.



Sample

To ensure the comparability of the treatment and control groups, 38 schools²⁸ that had not previously been involved in SNP were selected. These schools were grouped into 19 pairs matched on school level (elementary or secondary), region (central/ Amman, north/ Irbid, south/ Karak), and the gender of the schools (boys', girls', or mixed). Within each pair, one of the schools was randomly selected to receive the SNP, thus entering the treatment group, while the other was assigned to the control group²⁹. This was done to ensure that any differences in students' achievements could be attributed to being randomly assigned to the treatment group (relative to the control group) and not to other variables within the sampling variation limits. The 19 pairs of schools were spread across the three regions, with seven pairs of schools in the central region (Amman), seven pairs in the northern region (Irbid), and five pairs in the southern region (Karak).

²⁵ Link to the study: https://repository.upenn.edu/cpre_researchreports/61/

²⁶ Link to the study: https://repository.upenn.edu/cpre_researchreports/83/

²⁷ Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Belmont, CA: Wadsworth.

²⁸ At the start of the evaluation study, 40 schools (20 pairs) were selected. However, after the first school semester, one of the treatment group schools in the south region requested to withdraw from the study, which led to the removal of its control pair.

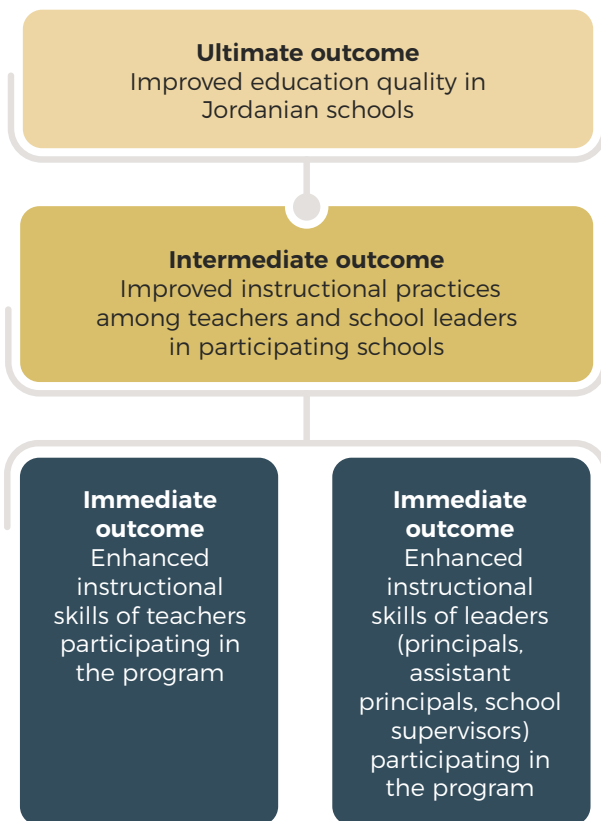
²⁹ For the purpose of this evaluation study, the control group will not receive the SNP and its results will be used for comparison.

Region	TREATMENT		CONTROL	
	Elementary	Secondary	Elementary	Secondary
Amman				
Irbid				
Karak				



Conceptual Framework

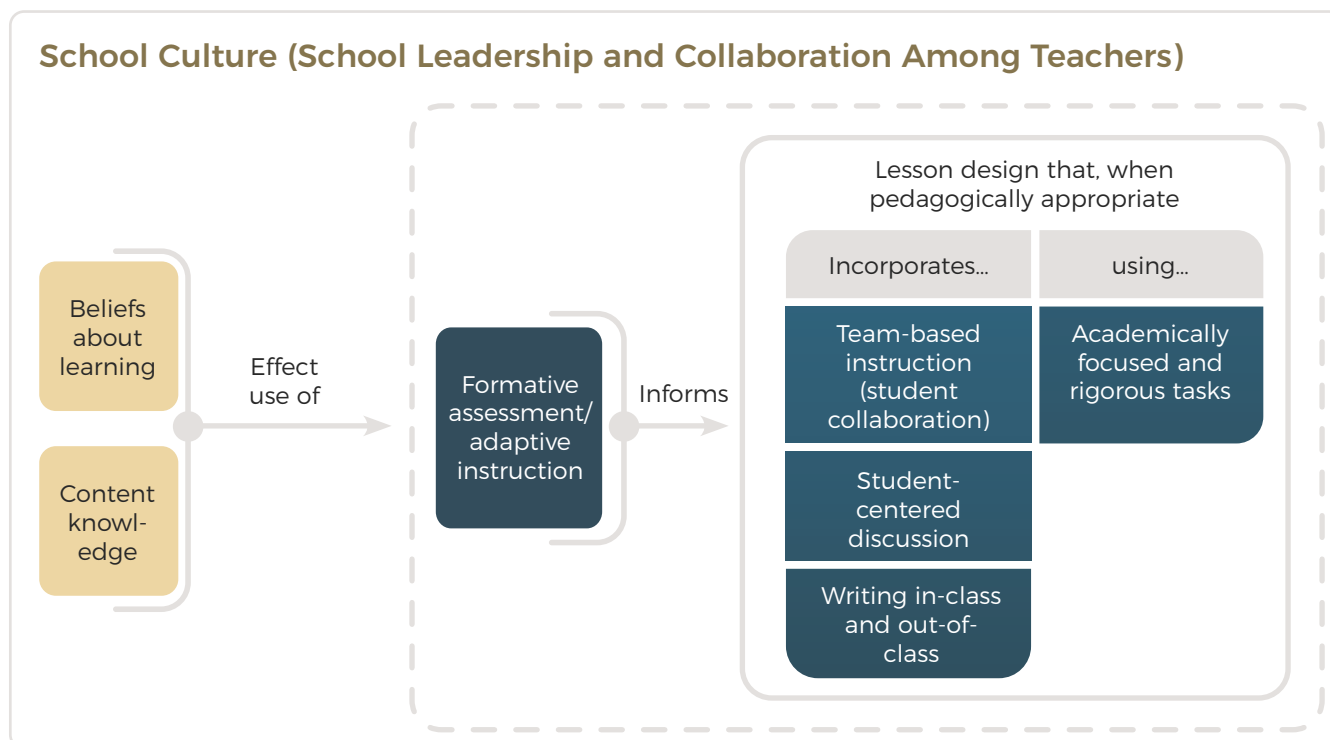
The underlying assumption was that QRTA’s SNP training model should promote change in the teachers’ and leaders’ performances and students’ achievements - if all SNP components were applied in a structured and efficient manner. This assumption aligned with the expected outcomes of the SNRS:



In addition to this study assumption and based on an extensive review of the QRTA SNP materials conducted by the OISE research team at the beginning of this assignment, the evaluation was designed depending on the relationships among the six “Big Effective” instructional practices of the SNP (lesson design, team-based instruction, academically focused and rigorous tasks, formative assessment/adaptive instruction, student-centered discussion, and writing in-class and out-of-class). These SNP instructional practices were connected with the relationship of teachers’ beliefs about learning and their content knowledge and the school culture’s influence - including school leadership and teacher collaborations.

The design³⁰ assumed that teachers’ use of formative assessment/adaptive instruction informs their lesson design, which, when pedagogically appropriate, incorporates team-based instruction (student collaboration), student-centered discussion, and writing in-class and out-of-class, using academically focused and rigorous tasks. The use of these instructional practices is affected by a teacher’s beliefs on learning and their content knowledge. Chances to use these instructional practices are affected by school culture, especially school leadership and collaboration among teachers (community of practice activities).

³⁰ QRTA Evaluation Plan, Schools Network Research Study, prepared by Ruth Childs & Kathryn Broad, OISE, October 2016.



Sources of Evidence:

Throughout the study, both quantitative and qualitative data were collected from schools, teachers, and leaders to create a full profile of all participating schools and investigate changes over time. Data was collected on training quality through post-workshop feedback surveys from teachers and leaders. Data on teachers’ and school leaders’ beliefs, attitudes, and practices were collected through questionnaires, classroom observations, assessments of teachers’ portfolios, and the success rates of the MOE accreditation exam. Students’ achievements were

measured through a test administered to a sample of students under the supervision of the NCHRD. Data was collected in three waves (2016-2017, 2017-2018, 2018-2019) to address the purposes of the study. Teacher and school leader questionnaires were collected in the first, second, and third years of the study. Classroom observations by QRTA instructional trainers and field researchers were undertaken in the second and third years. Students’ examination data was collected in the first and third years of the study.

Source of Evidence	Intended Outcome			
	Enhanced instructional skills of teachers participating in the schools’ networks	Increased instructional leadership skills of school leaders participating in the schools’ networks	Improved instructional practices of teachers and school leaders	Improved education quality
Teacher Questionnaire	x	x	x	
School Leader Questionnaire	x	x	x	
School Observation Forms	x	x	x	
Teacher MOE Accreditation Exam	x		x	
Teacher Portfolio	x		x	
Student Tests				x

Note: The Teacher MOE Accreditation Exam and the Teacher Portfolio are requirements for participating teachers to receive MOE accreditation for their participation in the Schools Network Program.



Teacher Questionnaire

Administered to participating teachers in both the treatment and control group schools in each data collection wave. Questions required respondents to select from among five choices (Likert-scale). Seven items were on teachers' perceptions of effective instructional practices, six items on whether/how teachers collaborated with colleagues, 20 on teachers' interactions with school principals, the same 20 were used to gauge teachers' interactions with supervisors,

and 22 items on teachers' beliefs on effective instructional practices. In the last section, each of the 22 items represented an instructional strategy. For each, teachers were asked to indicate the extent to which they: (1) had used the strategy in the last six months, (2) were confident about using the strategy in their subject area, and (3) thought the strategy was effective.



School Leader Questionnaire

Administered to principals, assistant principals, and education supervisors in both the treatment and control group schools. Questions required respondents to select from among five choices (Likert-scale). Seven items addressed school leaders' perceptions of effective instructional practices; five addressed their collaboration with teachers, 20 assessed school leaders' interactions with principals and assistant principals, and 20 assessed school leaders' interactions with supervisors. Finally, 22 items

gauged school leaders' beliefs on the instructional strategies they were confident about, and 22 items were concerned with their beliefs on effective instructional strategies. To ensure data triangulation and reliability, teachers from the treatment and control group schools completed questionnaires concerned with how effectively principals, assistant principals, and education supervisors encouraged teachers to incorporate improved instructional strategies in their professional practice.



Instructional Training "Coaching" Observation Tool

Completed during onsite support visits in treatment group schools by QRTA instructional trainers "coaches", who indicated whether or not they observed the listed instructional practices. The total number of items identified was used to indicate the degree to which teachers and school leaders successfully incorporated what they learned in the workshops into their practices. The observations contained main sections, each with sub-items. These sections were: (1)

the variety of student activities, (2) teachers' beliefs, attitudes, and instructional practices, (3) team-based instruction and student-centered discussion, (4) school culture and professional learning communities, (5) formative assessment/adaptive teaching, and (6) high expectations and rigorous tasks. The observations collected data applicable to instructional leadership practices and PLCs at schools.



Field Researcher Observation Tool

A second observation instrument completed by the field researchers accompanying QRTA instructional trainers during onsite support visits to provide additional information intended to corroborate instructional trainers' observations. During the instructional trainers' school visits, field researchers were asked to upload photographs of the physical walls of classrooms, lesson plans, seating plans, and student activities and interactions. The observations contained main sections, each with sub-items. These

sections were: (1) the diversity of student activities, (2) the availability of educational resources, and (3) the instructional leadership practices and professional learning communities, which field researchers were requested to collect as applicable. This observation tool required field researchers to estimate the time teachers and students spent engaging in different activities during lessons (e.g., teacher talking to the class, students working in groups, etc.).



Student Achievement Tests

QRTA contracted the NCHRD to administer Arabic, English, Mathematics, and Science tests (the study's four subjects) to a sample of students in grades six and ten. The tests were the same between waves, but the students were chosen from different cohorts

in December 2016 (pre-assessment) and December 2018 (post-assessment). The rationale behind this was linked to the study's objective, which was to measure schools' overall achievements and the overall improvement in teachers' progress on the

school-level, not the achievements and progress of individual students. Class sections were randomly selected, then samples of students from those sections were randomly chosen and required to complete two tests administered on the same day. The total time to complete the two tests was set at 75 minutes for grade six and 90 minutes for grade ten. Eight test types were created for each grade: Arabic Reading, Arabic Writing one (Personal Narrative and Article), Arabic Writing two (Personal Narrative and Informational Writing), English Reading, English Writing one (Personal Narrative and Article), English Writing two (Personal Narrative and Informational Writing), Mathematics, and Science.

The Science, Mathematics, and Arabic Reading assessments are part of the National Assessment for Knowledge Economy (NAfKE), which was developed as part of ERfKE. Meanwhile, the Arabic Writing

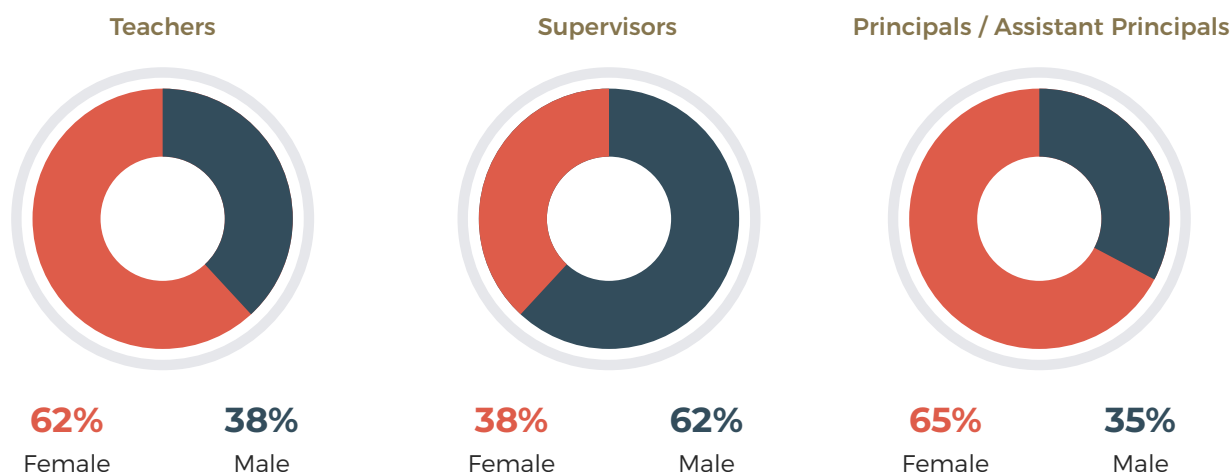
and English Reading and Writing assessments were developed by QRTA - in coordination with the MOE and NCHRD³¹ and administered to grade six and ten students. The Mathematics and Science tests aimed to measure skills such as knowing, applying, and problem-solving, in addition to content knowledge of the subject. The language tests were designed to assess students' knowledge, application, reasoning, retrieval, and understanding of Arabic and English language tasks. Writing one was considered an easier test because it included questions requiring students to write personal narratives and essays. Writing two tests were considered more difficult because students were required to write informational pieces, in addition to personal narratives. Since the purpose of the analyses was not to compare the sub-competencies within writing, these tests were combined.



³¹ The development of the tests for Arabic writing and English reading and writing went through a pilot phase to a sample of students at one of MOE schools in central region. This was to ensure validity, relevance and accuracy of the tests' questions and measures. Based on pilot results the actual tests that were used in the SNRS were prepared.

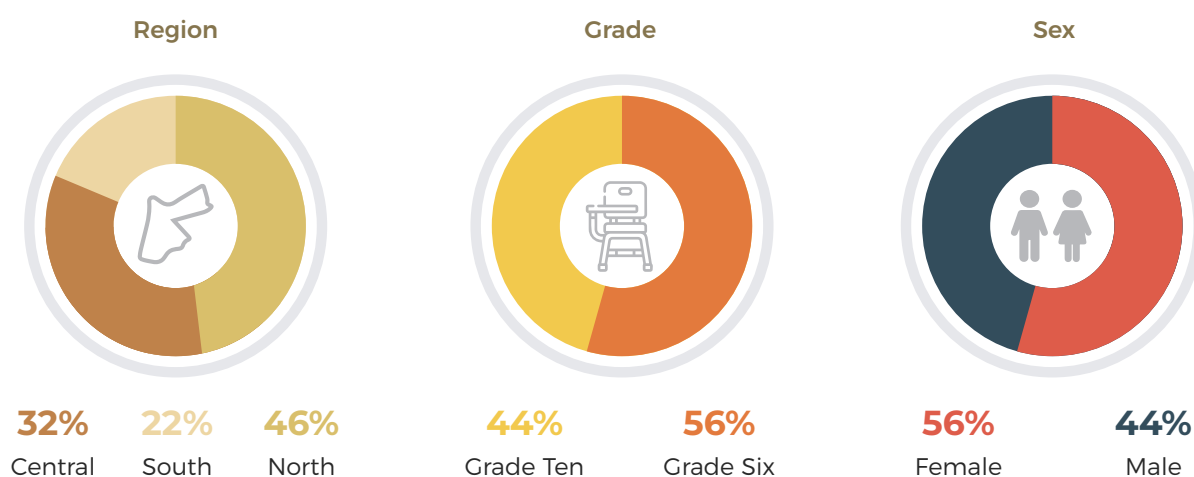
Characteristics of Participants

A total of 362 participants benefited from the training program in 19 treatment group schools across the three regions. The beneficiaries consisted of 301 teachers (186 females and 115 males), 21 education supervisors (8 females and 13 males), and 40 principals and assistant principals (26 females and 14 males).



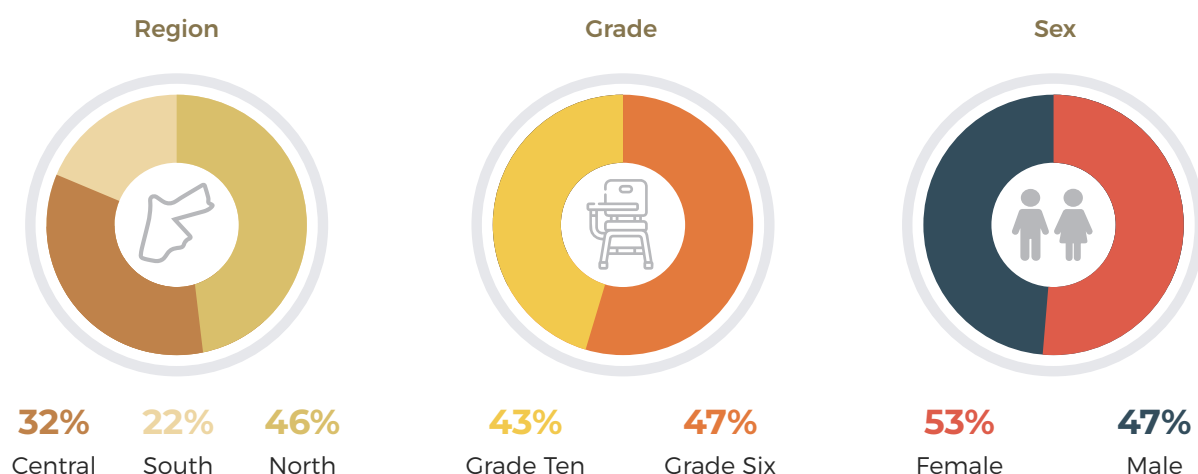
Out of the 301 teachers who received the training program, 212 teachers applied for the MOE certification exam, out of which 203 passed, representing a success rate of 96% (98% females and 91% males) across all subjects. The success rate was highest for Mathematics at 100%, followed by Arabic at 97%, English at 94%, and Science at 93%. The highest success rate was in north/ Irbid at 97%, followed by central/ Amman and south/ Karak at 95%. The overall evaluation results showed that northern and central regions were more committed to completing the training program and transferring the new practices to their schools, possibly explaining why the success rates were higher in these regions.

During wave one of the data collection (baseline), a total of 1,732 students took the test. Students from both the treatment and control groups, combined, were distributed as per the variables below:



³² List of participating schools is presented in annex 6.

During wave three of data collection (post-assessment), a total of 1,827 students took the test. As explained above, students from wave three were not the same as those who took the wave one test. The rationale behind this is that the SNRS aims to measure the effect of improving teachers' and leaders' practices on students' achievements in the school regardless of who they are. Teachers are expected to apply their new teaching methods with students from different grades and work gradually with them to scaffold and upscale learning as they move to higher grades. Students from both the treatment and control groups, combined, were distributed as per the variables below:



The distribution of students from the treatment and control groups across the 38 participating schools and the four subjects who took the subject tests are displayed in the tables below:

	Arabic Reading	Arabic Writing	English Reading	English Writing	Math	Science
Grade Six Students						
Treatment baseline	107	216	106	215	106	106
Treatment post	112	233	116	229	116	116
Control baseline	108	223	108	218	111	108
Control post	128	256	126	256	127	126
Grade Ten Students						
Treatment baseline	74	151	73	147	78	73
Treatment post	77	145	73	149	75	73
Control baseline	93	192	94	187	97	94
Control post	107	211	105	218	101	105

Audience and Stakeholders

This study's targeted stakeholders included strategic decision-makers within the education sector, particularly those concerned with teachers' performances and professional development; individuals engaged in implementation; and individuals who participated in and benefited from the SNP. These included the QRTA Board of Trustees, team members, donors, partners, educators, researchers, and the MOE policy and decision-makers, who informed on the process of education reform.

Study Limitations

As in any evaluation and/or research study, some limitations were encountered and considered:

- Small sample size, given the available financial resources.
- The dropout of one of the schools, which did not maintain attendance to the workshops.
- The transfers of teachers and leaders to other schools. To accommodate this, teachers' and leaders' data was tracked and updated regularly throughout the study.
- Some data was self-reported, questionnaire data. To account for this, data was collected from a variety of sources, enabling its triangulation.
- Transfers among the program team members (QRTA instructional trainers and field researchers) due to administrative issues beyond the scope and control of the study and the need to ensure consistency of collected data between different instructional trainers and field researchers. To accommodate this, frequent moderation and standardization sessions were conducted among the team members during the study to ensure unified data collection to the greatest extent possible.

Ethical Considerations of the Study

To ensure that the study was conducted ethically and complied with high standards, several issues were considered and addressed:

- The best interests of the participating teachers, leaders, and students were the primary consideration. Before the project began, QRTA sought permission from the MOE to include the participating schools.
- The study team did not put any participants (students or educators) at risk and ensured they did not come to any emotional or other harm.
- The project team and school staff informed all participants about the study, and their consent to take part was obtained.
- Participating teachers were fully informed about the evaluation aims and methods and agreed to participate in the project.
- Consents were obtained from all related participants (MOE, students, and educators) in case of any recordings, videotaping, or official visits during the study.
- Only data essential to the evaluation was collected. All data was treated confidentially, stored securely, and made available only to persons involved in the evaluation study.
- The participating teachers and leaders will be debriefed about the conclusions of the evaluation.



Key Findings

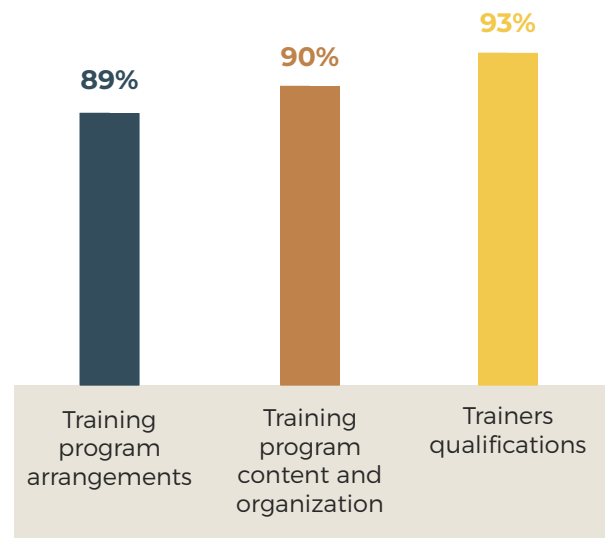
Overall, the SNRS results provided solid and compelling evidence that the SNP positively affected teachers' practices and students' achievements across all four subjects (Arabic, English, Mathematics, and Science). It also delivered some evidence, although

not as strong, that the SNP of Instructional Leadership had a positive effect on school leaders' instructional leadership practices. This section of the report offers details of data analysis³² results that support the study's key findings.

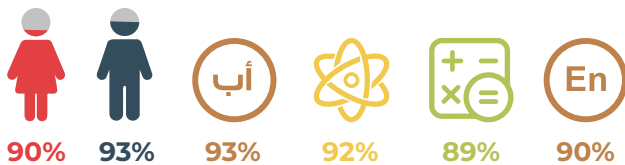
SNP Training Quality Satisfaction Rates:

Overall, participants assessed their training experiences highly with an average satisfaction rate of 91% (90% female and 93% males), and presented a wealth of information on how the program advanced their learning, which is explained further in the following report sections. Participants also acknowledged the high qualifications of the SNP instructional trainers and how this supported their learning experiences. The data collected from the post-workshop questionnaire showed that the overall average level of satisfaction of the training quality was positive and high for all subjects (Arabic 93%, Science 92%, English 90%, and Mathematics 89%).

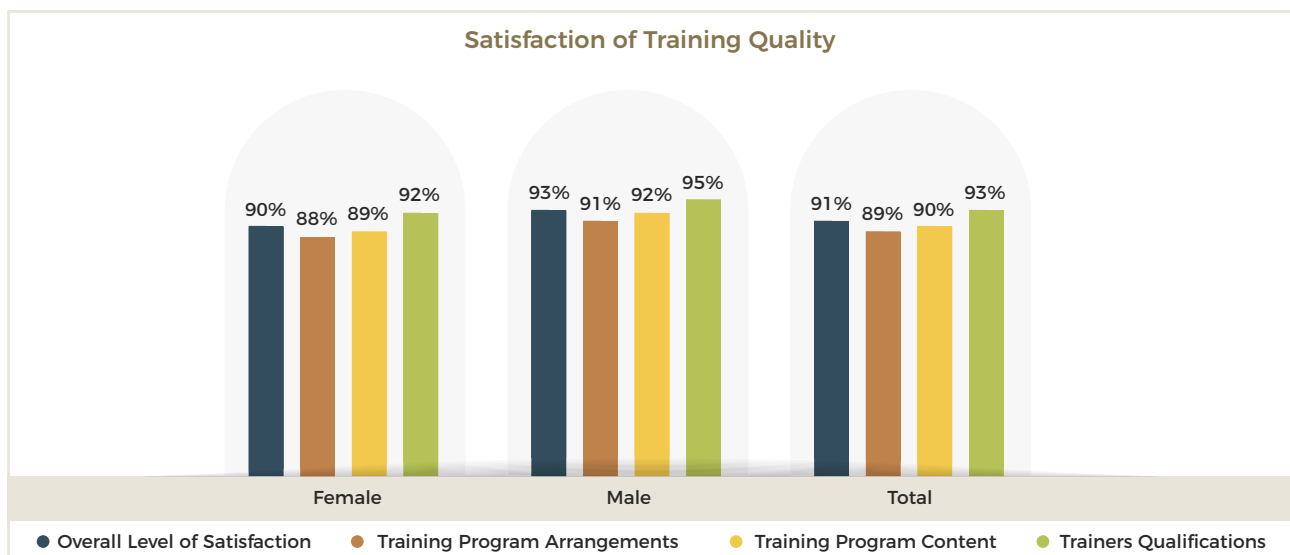
Level of satisfaction was measured in three main domains:



Satisfaction Rate



The satisfaction rates were measured across three main domains:



“ I learnt how to be a writer, not a person who only knows a few words and doesn't know how to use them ”
-English Teacher

³² One major observation found in analyzing teachers' and leaders' questionnaires is that respondents in both the treatment and control groups tended to give high scores in wave one, not fully understanding the concept of the SNP. However, in waves two and three, respondents in the treatment group became more reserved in their responses as they built a better understanding of the concepts, while the control group maintained their high scores to the questions.

Furthermore, participants were asked to rate their satisfaction with their knowledge and learning from the training program. Data analysis showed an average of 88% (87% females and 90% males). The average value for the different subjects was 89% for Arabic, 88% for Science, 87% for English, and 87% for Mathematics. Analysis results of the program quality showed that both females and males were highly satisfied with the program, and no significant variations were spotted in this regard. This would be because the QRTA SNP program is delivered to participants, whether females or males, with the

same approach and delivery model.

Moreover, whenever the QRTA team found that more support was required for certain groups of teachers, leaders, schools, or directorates, it was extended, regardless of gender. Having said this, and based on QRTA's experience in the education field, boys' schools usually require more support to implement and sustain new practices. Thus, QRTA took into consideration the specific needs of boys' schools and the challenges that faced male educators throughout the continuous development program activities.



Change in the Beliefs, Attitudes, and Instructional Practices of Teachers:

Overall, the study proved the significant and positive impact of the SNP. The positive changes in the beliefs, attitudes, and instructional practices of teachers among the treatment group schools were spotted in wave three compared to wave one. Meanwhile, the control group schools showed little or no evidence of any changes over the same period. By the end of the study, the treatment group teachers reported significantly more positive perceptions ($p = 0$)³³ of effective instructional practices, increased collaboration with colleagues, more frequent use of a broader range of instructional strategies, higher confidence in using different instructional strategies ($p = 0.054$), and more positive perceptions of the effectiveness of instructional strategies ($p = 0.045$) compared to the beginning of the study. In contrast, the control group teachers did not report any significant changes in these areas. The analysis

showed a statistical significance ($p = 0$) in favor of the treatment group during waves two and three in all areas, indicating the positive impact of the SNP.

The treatment group teachers reported enhanced perception and belief in the importance of effective practices ($M = 79\%$)³⁴. For example, the teachers demonstrated a growing belief in the importance of improving teaching to help students better acquire the needed knowledge and skills of the subject area. Moreover, their perceptions of their abilities to identify students' misconceptions in the subject area were improved. The distribution of the overall average scores (M) regarding teachers' perceptions per subject in both the treatment and control groups is displayed in the table below, where improvement was evident across all four subjects, and no major variations were observed:

	TREATMENT			CONTROL		
	Wave 1	Wave 2	Wave 3	Wave 1	Wave 2	Wave 3
Arabic	76%	79%	79%	79%	79%	79%
English	78%	77%	78%	78%	79%	75%
Math	76%	76%	76%	78%	79%	76%
Science	77%	81%	82%	75%	80%	76%

Teachers also reported a positive change in their attitudes towards the targeted instructional practices, including more collaboration with colleagues to plan and discuss lessons and increased use of teaching methods and strategies either on a daily or weekly basis. Teachers began attending each other's classes more often and exchanged useful feedback. Moreover, school leaders reported increased confidence in the teachers' attitudes towards instructional practices, including their confidence in the teachers' abilities to design and deliver lesson plans and learning activities for students of all different academic levels; leverage new teaching methods and strategies; and use modeling to explain concepts in the respective subject areas.

The treatment group teachers showed a stable or slight increase in their responses across the three data collection waves, indicating a steady rise in their perceptions of good instruction. In comparison, control group results in this domain remained stable or decreased - except for Science. According to anecdotal feedback received from treatment group teachers across the four subjects, *"That their lack of sufficient understanding of the best instructional practices as conveyed in the SNP model prompted them to answer non-objectively the questions of the questionnaire in the first wave; they showed advanced estimates of their skills and themselves. This subsequently led to the convergence of their answers to the questionnaire questions in the*

³³ This is the statistical significance "p-value" of paired-sample; t-test of the treatment group responses only in wave one and wave three.

³⁴ This is the total mean for the whole domain for perceptions of all teachers from the four subjects and three regions.

second and third waves with those they provided in the first wave.” Regardless of the data collection wave, results showed that teachers’ perceptions of teaching strategies were positively altered in the treatment group, whereas little or no change was detected in the control group. Overall, the treatment group teachers expressed confidence that they could identify students’ misconceptions (M = 86%) and confidence in their own teaching methods (M = 85%). They expressed a strong belief that students could acquire enough knowledge if they tried hard enough (M = 84%) and learn best through practical activities (M = 88%). One item to which teachers in both the treatment and control groups expressed a lower belief was whether teachers at their schools take responsibility for students’ achievements (M = 53%).

Overall, analysis³⁵ of teachers’ perceptions showed a statistically significant difference in favor of the treatment group in both wave two and wave three (p

= 0)³⁶ and a statistically significant difference within the treatment group itself between wave one and wave three (p = 0)³⁷, indicating the positive impact of the SNP.

Positive change in teachers’ attitudes towards collaborating with one another was also observed. Treatment group teachers attributed this improvement to PLCs, during which they would meet with colleagues from the same subject area to discuss lesson planning, teaching methods, and strategies. Teachers started attending each other’s lessons, welcomed teachers of different subject areas to attend their lessons, and exchanged useful feedback on their teaching methods and strategies. Overall, across the four subjects, the percentage of treatment group teachers who reported increased forms of collaboration rose compared to the control group teachers.

	Arabic				English			
	TREATMENT		CONTROL		TREATMENT		CONTROL	
	Wave1	Wave3	Wave1	Wave3	Wave1	Wave3	Wave1	Wave3
Never	13%	3%	10%	4%	14%	6%	18%	9%
Less than once in month	39%	20%	26%	21%	31%	18%	29%	33%
Once or twice monthly	31%	42%	34%	43%	41%	44%	28%	28%
Once or twice a week	14%	30%	24%	19%	11%	30%	20%	24%
Daily	3%	5%	6%	12%	3%	2%	6%	5%

	Math				Science			
	TREATMENT		CONTROL		TREATMENT		CONTROL	
	Wave1	Wave3	Wave1	Wave3	Wave1	Wave3	Wave1	Wave3
Never	14%	10%	11%	4%	14%	8%	18%	6%
Less than once in month	36%	16%	37%	24%	32%	36%	36%	43%
Once or twice monthly	35%	42%	32%	37%	29%	41%	29%	36%
Once or twice a week	14%	30%	16%	24%	21%	13%	13%	12%
Daily	2%	2%	4%	10%	4%	2%	3%	4%

Across the four subjects, data showed a consistent and stable positive change among the treatment group teachers, mostly on the meeting “once or twice a week” basis, which the SNP encourages. The control group teachers also showed a change in this indicator, but not consistently. This latter change most probably did not reflect the same understanding of teachers’ collaborations as presented within the SNP model.

Moreover, evidence from the qualitative data gathered from teachers’ questionnaires showed that more than half of the treatment group teachers worked with their colleagues on building PLCs. Teachers expressed that communication amongst them had increased, whereby they began organizing reflective meetings;

“ **The interaction and the meetings between teachers increased in order to exchange experience, opportunities for cooperation, and increase collaboration among us**

- English Teacher

”

³⁵ Whenever the control or treatment groups are not mentioned, it means there was no statistical significance. This applies to all sections below.

³⁶ This is the statistical significance “p” value of ANCOVA to test whether or not there is a statistical significance in responses between wave one and wave two, wave two and wave three, wave one and wave three, and between the treatment and control groups.

³⁷ This is the statistical significance “p” value of paired-sample t-test for the treatment group responses between wave one and wave three only.

exchanging teaching experiences, strategies, and visits; and providing each other with constructive feedback. Some teachers in the control group mentioned that they exchanged experiences and discussed teaching practices among themselves; however, this is not believed to be structured as per the PLCs promoted by the SNP.

In addition to beliefs and attitudes, teachers demonstrated enhanced instructional practices. In general, teachers who were committed to attending all the training activities and were keen on transferring their knowledge to their classroom demonstrated a more positive change in their pedagogical practices. More positive change in the learning environment was also observed in schools where the majority of teachers engaged in the training program remained in their schools throughout the study duration. This was more evident in the central and northern regions compared to the southern region. In general, observational data and instructional trainers' and field researchers' feedback from onsite support visits to schools during the study showed that central and northern region schools had a slightly higher application of new practices. This is also due to fewer transfers among teachers and school leaders in these regions compared to the southern region. Additionally, this positive change was more evident in girls' schools than boys' - except for two boys' schools³⁸ in the central region that showed a very high positive change. Generally, teachers and leaders at girls' schools exhibited a higher commitment to

attending all training activities and encouraged PLCs. Female teachers transferred the newly acquired practices more frequently to classrooms compared to their male counterparts. These reasons are believed to be behind girls' schools outperforming boys' schools.

The treatment group teachers positively expressed how frequently they implemented new instructional strategies - such as asking questions of varying difficulty levels, using observations to assess students' understanding and plan teaching accordingly, and delivering lesson plans and activities designed to accommodate students' different academic skills. The treatment group teachers' responses showed a significantly increased frequency in using instructional strategies on a "daily" and "once or twice a week" basis. This change occurred across the four subjects and was consistent and stable. Arabic teachers reported the highest frequency in using the new instructional strategies on a "daily" basis, followed by English and Mathematics, while Science remained the same. On the "once or twice a week" basis, though, Science reported the highest frequency, followed by Mathematics, Arabic, and English. The control group teachers, in contrast, showed a slight positive change only in Mathematics and Science and was limited to the "once or twice a week" basis. However, changes in the control group most likely do not reflect the same understanding of instructional strategies as conveyed by the SNP training model since they did not undergo the training program.

Daily basis	TREATMENT			CONTROL		
	Pre	Mid	Post	Pre	Mid	Post
Arabic	44%	55%	51%	45%	38%	39%
English	35%	32%	41%	45%	43%	37%
Math	41%	41%	46%	48%	44%	48%
Science	46%	40%	46%	40%	34%	40%
Once or twice a week basis	Pre	Mid	Post	Pre	Mid	Post
Arabic	33%	36%	36%	38%	40%	35%
English	39%	42%	40%	41%	37%	34%
Math	33%	33%	37%	33%	32%	34%
Science	35%	42%	44%	35%	33%	36%

During waves two and three, teachers were asked if they had used different teaching practices and student activities than in previous years. Qualitative data analysis of the teachers' questionnaires showed that most treatment group teachers used different teaching practices while almost half of the control group teachers said they did not use different

teaching practices. In general, most of the treatment group teachers mentioned that they utilized group work, modeling, traffic light cards, learning through play techniques, and role-playing. They also began integrating more technology in their teaching, using the internet and interactive boards.

³⁸ More information on these two boys' schools is provided in the section below headed "Change in the Beliefs, Attitudes, and Instructional Leadership Practices of School Leaders".

“ I used active learning teaching strategies to engage students in the lesson. I also used teaching strategies based on inquiry

- Science Teacher

”

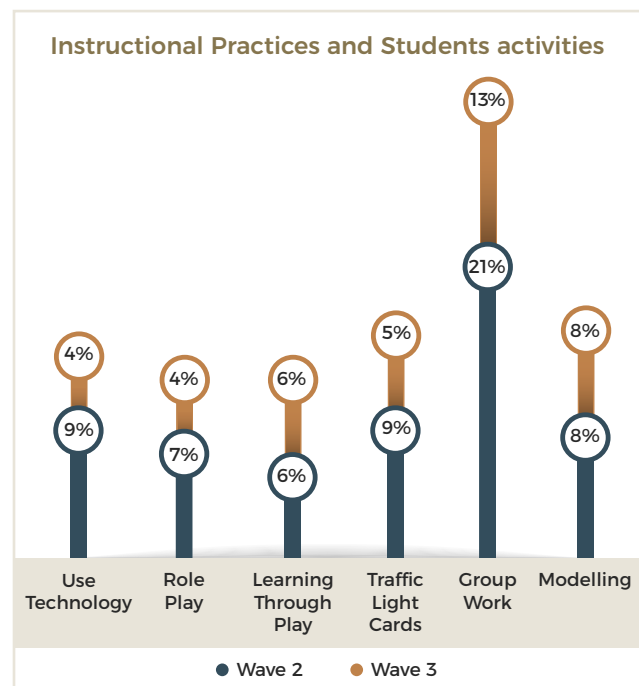
On the subject level, Arabic teachers reported the highest frequency in using the reading and writing instructional strategies during wave two (61%, 15%) and wave three (49%, 41%). English teachers reported an increase in applying reading and writing instructional strategies during wave three (17%, 27%). More than half of the Mathematics teachers reported using manipulatives in teaching the subject during wave two and wave three, while other Mathematics teachers said that they used mini boards, interactive boards and the internet to deliver certain lessons. More than one-third of Science teachers mentioned that they applied the “5Es” inquiry model and used school laboratories to perform scientific experiments. Furthermore, most treatment group teachers mentioned that they used new, previously untapped teaching resources in the classroom. For example, Mathematics teachers said they used the manipulatives provided by QRTA in teaching the subject. Arabic teachers stated that they used library books and stories. Teachers across the four subjects mentioned using more of the technological resources available at schools, such as interactive boards, computers, and the internet.

Teachers’ confidence in their abilities to apply the new targeted instructional practices was also measured throughout the SNRS. Analyzed results revealed a stable or slight increase in teachers’ responses across the three data collection waves for the treatment group, indicating a steady rise in teachers’ confidence

“ Using the computer has allowed students to learn more and go beyond the curriculum

- Arabic Teacher

”



to use the different instructional strategies in their subject teaching. Control group results showed stable or decreased average scores in this domain. Overall, and regardless of the data collection wave, the treatment group teachers expressed high confidence levels in encouraging class discussions that engaged students ($M = 89\%$); planning the beginning of lessons to engage and activate students’ previously-acquired knowledge ($M = 88\%$); and using observations to assess students’ understanding ($M = 88\%$). They also expressed high levels of confidence in asking questions of varying difficulty levels ($M = 87\%$) and questions to assess students’ understanding and facilitate lesson planning ($M = 87\%$); allowing students to choose topics for their assignments and projects ($M = 80\%$); and using lecturing in teaching ($M = 80\%$) as needed.

Overall, analysis of teachers’ confidence showed a statistically significant difference in favor of the treatment group in both wave two and wave three ($p = 0$) and a statistically significant difference within the treatment group itself between wave one and wave three ($p = 0.054$), indicating the positive impact of the SNP. The distribution of overall average scores (M) for confidence per subject in the treatment and control groups is explained in the table below, where the average was high for all four subjects and no major variations among them were spotted:

	TREATMENT			CONTROL		
	Wave1	Wave2	Wave3	Wave1	Wave2	Wave3
Arabic	82%	86%	86%	83%	82%	83%
English	84%	80%	83%	86%	85%	81%
Math	82%	81%	84%	87%	84%	85%
Science	86%	87%	87%	84%	84%	83%

The treatment group teachers' beliefs in the effectiveness of the different teaching strategies showed a statistically significant change between waves one and three. Analysis results indicated a stable or slight increase in teachers' responses across the three data collection waves for the treatment group, while the control group presented stable or fluctuating average scores in this domain. Regardless of the data collection wave, the treatment group teachers' responses were slightly higher (M = 84%) than in the control group (M = 83%). Female teachers in the treatment group (M = 85%) reported slightly higher averages than males (M = 84%), while, in the control group, female and male teachers reported

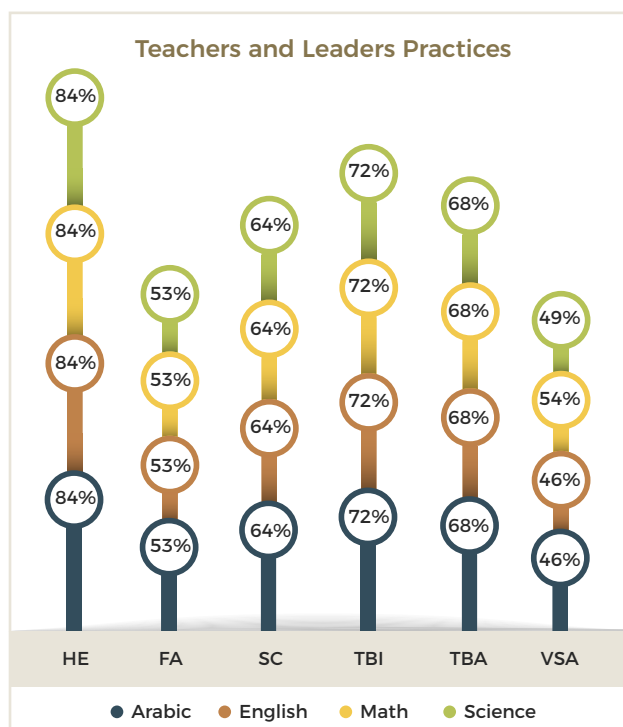
the same average scores (M = 83%).

Overall, analysis of this domain on the effectiveness of new teaching strategies revealed a statistically significant difference in favor of the treatment group in both wave two and wave three (p = 0) and a statistically significant difference within the treatment group itself between wave one and wave three (p = 0.045), indicating the positive impact of the SNP. The distribution of the overall average scores (M) for effectiveness per subject in the treatment and control groups is explained in the table below, where averages were also high for all four subjects and no major variations among them were spotted:

	TREATMENT			CONTROL		
	Wave1	Wave2	Wave3	Wave1	Wave2	Wave3
Arabic	83%	86%	86%	84%	82%	81%
English	84%	83%	83%	82%	83%	83%
Math	81%	83%	86%	86%	85%	86%
Science	85%	85%	86%	83%	84%	82%

Teachers also reported sufficient levels of implementation of the newly acquired practices. Qualitative data analysis of teachers' questionnaires showed that most treatment group teachers applied different teaching practices, while almost half of the control group teachers said they did not use different teaching practices. Also, classroom observations³⁹ by instructional trainers and field researchers served as vital evidence on teachers' and leaders' behaviors in waves two and three. Results of the behavior analysis showed a statistical significance as follows: Arabic (p = 1e-3), English (p = 3e-2), and Science (p = 5e-5), while the statistical significance for Mathematics was not calculated due to the change of trainers for reasons beyond the program team's control.

On average, the most frequently reported levels of implementation of the various instructional practices were high across all four subjects. For example, teachers' average use of various student activities (VSA) during lessons was 46% for Arabic and English, 49% for Science, and 54% for Mathematics. Teachers' beliefs, attitudes (TBA), and ability to set expectations was 68%. They were able to identify areas and steps for self-development; describe changes made in their



³⁹ Observational data are nominal since trainers and trainees recorded a given teacher's behavior as either present or absent. Kruskal-Wallis tests were applied to the observational data to test the null hypothesis that grouping does not result in differences in coaches' and trainees' classroom observations. The Kruskal-Wallis test is appropriate because it determines whether or not the population distributions are the same without assuming normal distributions.

practices over time; and connect the changes to both student and professional self-learning. Teachers promoted team-based instruction and student-centered discussions (72%). This investigated, for example, if students know how to work in pairs or groups, can easily assign responsibility in order to accomplish tasks; and work with and listen to each other. Evidence on positive school culture and PLCs reached 64%, as principals and teachers talked about teaching and learning practices at their schools, and

teachers discussed working together to plan lessons and visited each other's classes. Teachers' use of formative assessment (FA)/adaptive teaching/check for understanding was evident at 53%, as teachers conversed with students, asking them to describe their thinking and adapted lessons in response. Moreover, teachers showed high expectations (HE) of their students and gave them rigorous tasks and clear directions during lessons 84% of the time.



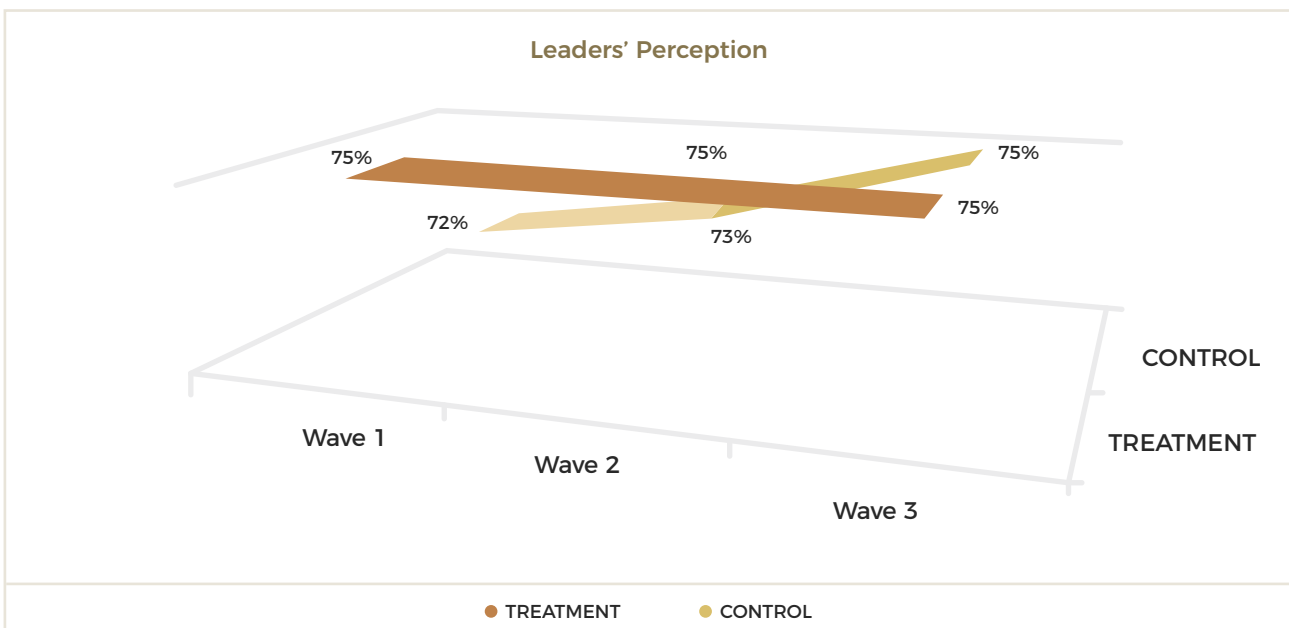
Change in the Beliefs, Attitudes, and Instructional Leadership Practices of School Leaders

Overall, the study proved the significant and positive impact of the SNP. Although slightly less than teachers, leaders among the treatment group schools demonstrated a positive change in their beliefs, attitudes, and instructional leadership practices in waves one and three. The training delivery model of the Instructional Leadership network focuses on raising awareness on the best practices in instructional leadership and encouraging leaders to adopt them. Unlike the subject-specific networks - except for learning walks and attending classroom observations with QRTA instructional trainers - the Instructional Leadership network does not include other onsite support activities. As leaders are often busy with the administrative workload at their schools, their ability to fully implement the instructional leadership practices as expected is adversely affected. This factor explains why the impact on leaders fell behind that of teachers. Still, the control group leaders showed little or no evidence of change over the same period, while the treatment group leaders reported statistically insignificant ($p = 0.945$) positive perceptions pertaining to effective instructional practices, increased collaboration among teachers, and more frequent use of the different instructional

strategies among teachers. The positive change in leaders' perceptions was not statistically significant ($p = 0.23$) regarding their confidence in their teachers' use of the targeted instructional strategies or the effectiveness of these instructional strategies ($p = 0.62$) at the end of the study compared to the beginning. However, the control group leaders did not report any significant changes occurring in these areas.

Treatment group teachers reported significantly more positive perceptions ($p = 0$) pertaining to the instructional leadership practices of principals and assistant principals by the end of the study compared to the beginning, whereas control group teachers reported little or no change in their beliefs.

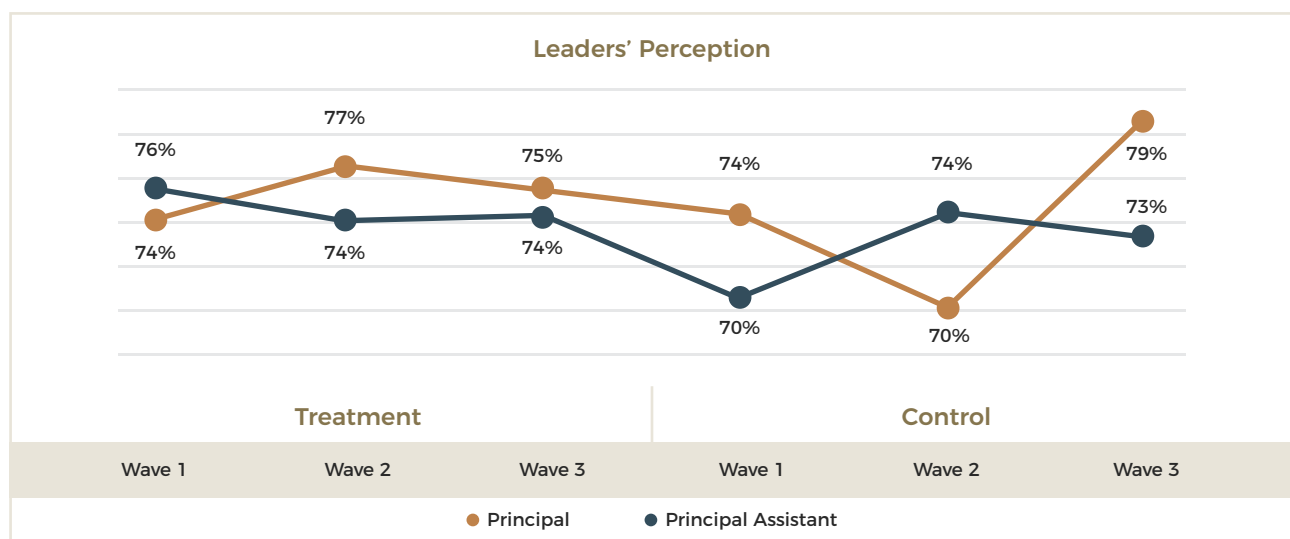
Treatment group leaders reported positive perceptions and beliefs in the importance of effective instructional practices among their teachers ($M = 75\%$), and results were aligned with those reported by the teachers. For example, leaders demonstrated a growing belief in the importance of enhancing teaching to help students better acquire a subject area's knowledge and skills. Moreover, their perception of the teachers' abilities to identify students' misconceptions in the subject area was improved.



Leaders indicated a stable trend in their perception of their teachers' instructional capacities. The responses collected throughout the three data collection waves of the treatment group remained steady, whereas the control group showed a slight increase in wave three compared to wave one. Nonetheless, the highest value that the control group scored was similar to that of the treatment group. According to anecdotal feedback received from the treatment group leaders, *"That their lack of sufficient understanding of the best practices in educational leadership prompted them to answer non-objectively the questions of the questionnaire in the first wave; they showed advanced estimates of their skills and themselves. This subsequently led to the convergence of their answers to the questionnaire questions in the second*

and third waves with those they provided in the first wave."

Regardless of the wave, the treatment group leaders' perceptions did not differ between females and males, as both were at (M = 75%), while in the control group, females were at (M = 76%) and males at (M = 63%). Here, we assume that due to participation in the SNP, the treatment group leaders - both females and males - built better perceptions of their teachers compared to the control group. Based on directorates, the treatment group scored slightly higher results (central (M = 77%), north (M = 75%), and south (M = 72%)) compared to the control group (central (M = 75%), north (M = 75%), and south (M = 69%)).



Overall, analysis of leaders' perceptions showed a statistically significant difference between groups in favor of the treatment group only in wave two ($p = 0.018$), while no statistical significance was found in wave three ($p = 0.088$). No statistically significant difference was found within the treatment group itself between waves one and three ($p = 0.945$), which supports the partial positive impact of the SNP.

Leaders reported positive feedback on their teachers' attitudes towards the targeted instructional practices. For example, they recounted that teachers started to collaborate with colleagues better to plan and discuss lessons, teaching methods, and strategies. They added that teachers started to attend each other's classes more often and exchange useful feedback. Meanwhile, teachers conveyed improved attitudes towards their leaders' instructional leadership practices. For example, they noted that leaders started encouraging teachers to improve their teaching practices; showing more respect to the teachers' experiences; further appreciating their good work and meticulous follow up on the academic progress of students; accepting constructive feedback from teachers and taking actions accordingly; and calling on teachers to play a meaningful role in setting school development objectives.

Moreover, evidence from the qualitative data analysis of the leaders' questionnaires showed that, during waves two and three, most treatment group leaders mentioned that the way teachers were collaborating at school had changed compared to previous years. They also indicated that communication between themselves and the teachers and among the teachers had improved. Leaders also mentioned that teachers exchanged classroom visits and discussed teaching the subject together. One school leader said, "There is a difference in the way teachers interact with each other compared to previous years. Teachers exchanged views and benefited from each other's experiences, which was reflected inside the classroom". Moreover, leaders said that teachers used different teaching practices and student activities compared to previous years, as they applied the new instructional strategies learned in the SNP. According to leaders, teachers applied active learning strategies and collaborative work and used supportive teaching strategies. Most treatment group leaders stated that the way they worked with teachers at their schools had changed. Leaders also said that they applied new instructional leadership practices as they attended and observed classes, gave feedback, and enhanced communication with their teachers. Meanwhile, the

“ In previous years, classroom activities were few, whereas now there are continuous activities inside classrooms, and they are interesting. Teaching is also different as a result of teachers’ training in the SNP ”

- School Leader

qualitative data analysis of the control group leaders’ questionnaires showed that there was no change in teachers’ collaborative work, and they did not use different teaching practices or student activities than in previous years. Moreover, the way leaders worked with their teachers did not change compared to previous years.

In addition to leaders’ ratings of themselves, the treatment group teachers also reported significantly more positive perceptions of their principals’ and assistant principals’ instructional leadership skills at the end of the study (M = 82%) than at the beginning (M = 76%).

Overall, analysis of how teachers perceived instructional leadership by their leaders showed a statistically significant difference between groups in favor of the treatment group in wave three ($p = 0$) only, while in wave two, this difference was in favor of the control group ($p = 0$). A statistically significant difference within the treatment group itself between wave one and wave three ($p = 0$) was also found, which supports the partial positive impact of the SNP.

In addition to beliefs and attitudes, leaders showed improvement in their instructional leadership practices. Generally, leaders who were committed to attending the training activities or recognized the expected benefits of the SNP were characterized as being more supportive, collaborative, and

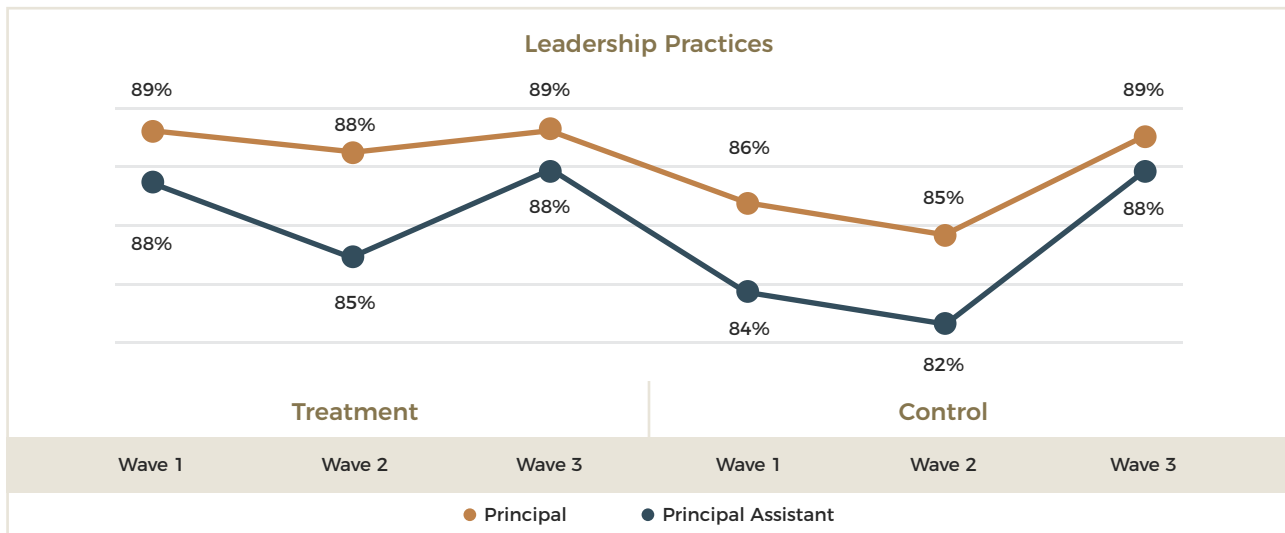
motivational by their teachers. This encouraged teachers to apply the new skills and knowledge they acquired, eventually showing a more positive change. This trend was generally more evident in girls’ schools than boys’ schools, except at two boys’ schools in the central region, which demonstrated a significantly high positive change due to their leaders’ support and encouragement. The principals and assistant principals in these two schools were highly cooperative; they embraced positive views and the potential impact of the SNP. They also facilitated and followed up on their teachers’ attendance at the training activities and encouraged them to apply the new teaching practices in the classroom.

Similar to the results observed among teachers, the treatment group leaders indicated improved collaboration among their teachers in the form of PLCs. Leaders stated that teachers started meeting with colleagues from the same subject area to discuss lesson planning, teaching methods, and strategies. Teachers also started attending each other’s lessons, welcomed teachers from different subject areas to attend their lessons, and exchanged useful feedback on their teaching methods and strategies. Overall, the percentage of treatment group leaders who reported increased forms of collaboration changed positively over time compared to the control group leaders.

	TREATMENT		CONTROL	
	Wave 1	Wave 3	Wave 1	Wave 3
Never	5%	2%	16%	3%
Less than once in month	33%	26%	28%	29%
Once or twice monthly	43%	58%	34%	34%
Once or twice a week	18%	14%	16%	28%
Daily	1%	0%	7%	5%

The treatment group leaders showed more stable trends in their instructional leadership practices throughout the three data collection waves than those of the control group. In wave one, the treatment group started at a slightly higher baseline, yet this trend was maintained throughout the second and

third waves. Alternatively, the control group started at a lower baseline, later showing a fluctuating trend in wave two (decrease) and wave three (increase), which can be attributed to a time-related change and other factors unrelated to the SNP program.



Overall, analysis of this domain showed a statistically significant difference between groups in favor of the treatment group only in wave three ($p = 0.002$). Yet, no statistically significant difference was found in the treatment group itself between wave one and wave three ($p = 0.877$), which supports the partial positive impact of the SNP.

Moreover, regardless of the data collection wave, the treatment group leaders reported a slightly higher average score ($M = 77\%$) of instructional leadership skills between education supervisors compared to the control group leaders ($M = 72\%$). Education supervisors' participation in the subject-specific SNP could be a strong reason behind the enhanced support they extended to teachers and their better understanding of the new instructional practices that the SNP encourages. Male leaders in the treatment group scored higher ($M = 80\%$) compared to females ($M = 74\%$). Male leaders in the control group also scored higher ($M = 73\%$) compared to females ($M = 71\%$). These findings align with the public sector supervision trends in Jordan, which is more male-dominant due to its nature, fieldwork focus and longer hours, consequently making it less appealing to females.

Overall, analysis of this domain showed a statistically significant difference between groups in favor of the treatment group only in wave two ($p = 0$). In wave three, this difference was in favor of the control group ($p = 0$). No statistically significant difference was found within the treatment group itself between wave one and wave three ($p = 0.213$), while there was a significant difference between wave one and wave three for the control group ($p = 0$), which supports the partial positive impact of the SNP.

Regardless of the data collection wave, leaders reported high average scores on their beliefs in the effectiveness of teaching strategies and their teachers' use of them - such as asking questions of varying difficulty levels, drawing on students' prior knowledge and experiences to develop lesson plans, requesting students to explain their thinking and justifications to each other, and encouraging engaging class

discussions and question-asking by the students, etc. To this end, the treatment group leaders scored slightly higher ($M = 84\%$) than those in the control group ($M = 82\%$).

Overall, analysis of leaders' beliefs in the effectiveness of teaching strategies and their teachers' use of them showed a statistically significant difference between groups in favor of the treatment group in both wave two ($p = 0$) and wave three ($p = 0.01$), while no statistically significant difference was found within the treatment group itself between wave one and wave three ($p = 0.62$), which supports the partial positive impact of the SNP.

Male leaders, regardless of the data collection wave, in both the treatment group ($M = 85\%$) and control group ($M = 91\%$) reported higher averages than females in the treatment group ($M = 83\%$) and control group ($M = 79\%$). This finding is interesting because, in general, female schools outperform male schools. This could be explained by the gender differences in self-reporting (males tend to appraise their skills more highly than females). The findings from leaders' level of confidence in their teachers' abilities to use teaching strategies confirmed this by showing that male leaders were rated lower than females. Regardless of the data collection wave, both the treatment and control groups scored the same average ($M = 73\%$), with females in the treatment group scoring ($M = 75\%$) and in the control groups ($M = 74\%$) and males in the treatment group scoring ($M = 71\%$) and in the control group ($M = 72\%$). Under the "effectiveness" and "confidence" domains, the northern region scored highest, followed by the treatment group's southern and central regions.

Overall, the analysis showed a statistically significant difference between groups in favor of the treatment group in wave three ($p = 0.023$) only. In contrast, for wave two, this difference among groups was statistically significant in favor of the control group ($p = 0.002$). No statistically significant difference within the treatment group itself was found between wave one and wave three ($p = 0.23$), which supports the partial positive impact of the SNP.



Change in the Achievement Outcomes of Students:

Overall, student test results for both grades six and ten across all subjects provided compelling evidence of higher achievement scores among the treatment group schools compared to the control group schools, and higher achievement scores by the treatment group at the end of the program compared to the beginning.

Andersen's Likelihood Ratio test was used to compare sub-populations. An aggregate students' achievement score was combined (treatment, control) and modeled as a single dataset using a partial credit model. Students' aggregate scores were then plotted by data collection waves. Plots were examined visually to determine if there was a good fit between the difficulty of the items used in the test and the model estimate of students' abilities to answer. In most cases, the difficulty range of the items aligned with the achievement range of the test-takers. Simple Linear Regression and Wilcoxon tests of statistical significance of probability (p-value) were conducted.

Hypothesis testing⁴⁰ was performed in two ways:

1. Vertically between the treatment and control groups in both waves one and three to check for any differences. Testing in wave one aimed at checking if both groups were the same (comparable), and wave three aimed at identifying differences as a result of the SNP.
2. Horizontally between waves one and three for the same group to check for any statistically significant differences as a result of the SNP.

Responses from students in the control group schools are shown as blue text, points, and lines, while responses from students in the treatment group schools are shown as red text, points, and lines. Control and treatment data ellipses (blue and red) illustrate data tendencies and degree of overlap, which gives a visual sense of the changes between waves one and three: no change, trending upward, trending downward. The control and treatment lines of best fit (simple linear regression results) with standard errors (gray shading around the lines) also illustrate data trends and similarities (a narrower grey area indicates fewer errors and more reliable results). The grey area is narrower from the middle and gets larger towards the end in wave three. Because there was no mid-point assessment (wave two) for students, the line represents a line of best fit.

18 out of 19 treatment group schools showed an improvement in students' overall average scores across all four subjects in standardized tests, representing 95%. The average total scores of students in the treatment group schools per subject per grade in each school show that students' test results for grade six improved across all schools - except for one female school in the southern region in English Reading⁴¹ - and improved across all schools for grade ten. The number of girls' and mixed schools in the treatment group that showed improvement in standardized tests was 10 out of 11, representing 91%, and all eight boys' schools showed a positive change in students' overall average scores, representing 100%. The below table shows the details of the treatment group schools that improved students' scores:

	Number of treatment schools	Number of schools that improved students' scores	Percentage of improved schools out of total number of treatment schools	Percentage of improved schools out of number of schools for each grade
Number of treatment schools	19	18	95%	-
Number of schools which have grade six	15	14	74%	93%
Number of schools which have grade six only	9	8	42%	89%
Number of schools which have grade ten	10	10	53%	100%
Number of schools which have grade ten only	4	4	21%	100%

⁴⁰ Hypothesis testing (Probability) is (p-value), which is used in statistics to measure the statistical significance. When the p-value is less than 5%, there is a statistically significant effect. When the p-value is more than 5%, there is no statistically significant value, and change (increase or decrease) is attributed to error. The Wilcoxon test is used to test the probability (p-value for statistical significance) of students' achievements and observations. p-value is presented using the "e" mathematical exponential. Statistical significance values in the report are shortened using "e" following research writing best practices. For example, "3.06e-1" is equal to: 3.6 times 10 to the power -1 (1/(10 to the power 1)), which is equal to 0.36. When (p > 5e-2), there is a statistical significance, when it is more, there is no statistical significance.

⁴¹ Students' test results in the pre-assessment phase were very high for this school compared to all other scores whether in this subject or across all subjects. Even though there were quality assurance measures during students' testing by NCHRD and QRTA, we still expect that students had some sort of support to answer the questions in this particular school, but there is no strong evidence on this.

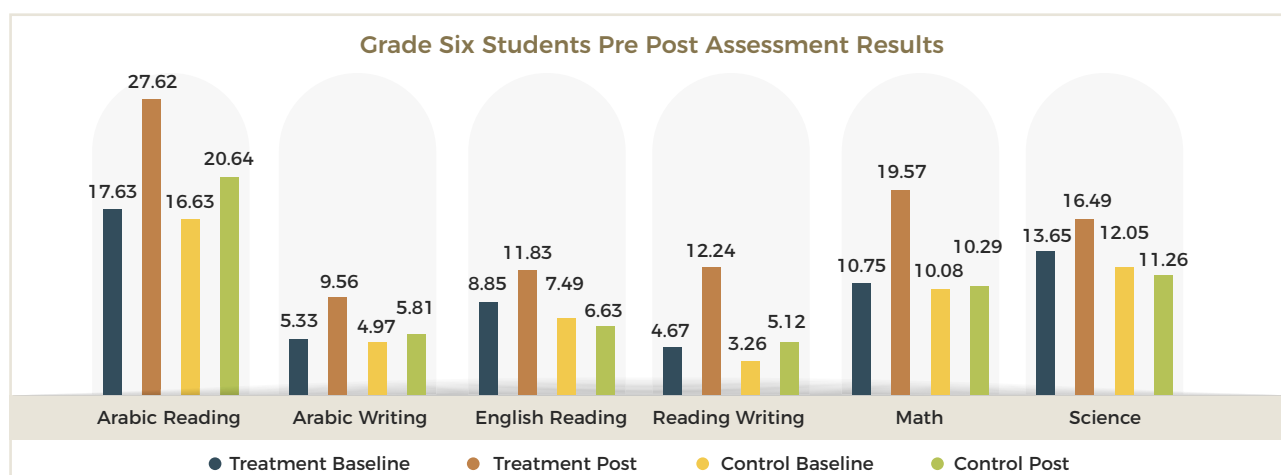


Grade Six Students' Overall Results

On average, students' scores across all schools for grade six increased in wave three (post-assessment) compared to wave one (pre-assessment) for all subjects, except English Reading in one of the treatment group schools in the southern region. Average scores for the different subjects clearly indicated the significant progression from wave one to wave three results between the treatment and control groups. The treatment group schools' overall averages increased across all subjects, while the

control group's scores did not. The highest increase in the treatment group schools was in English Writing, followed by Mathematics, Arabic Writing, Arabic Reading, English Reading, and Science. In the control group schools, English Writing showed the highest increase, followed by Arabic Reading and Writing, which showed a moderate increase. Also, the Mathematics increase was very small, whereas the English Reading and Science results decreased.

Grade Six	Arabic Reading	Arabic Writing	English Reading	English Writing	Math	Science
Change ratio⁴² in Treatment	57%	79%	34%	100%	82%	21%
Change ratio in Control	24%	17%	-11%	57%	2%	-7%



This high positive change in the sixth graders' achievements in wave three revealed a highly statistically significant difference in Arabic Writing, English Reading, English Writing, Mathematics, and Science (all standing at $p = 0$) in favor of the treatment group, while no statistically significant difference was spotted in Arabic Reading ($p = 3.06e-1$). Hence, moving forward from wave one to three, this positive change was also evident across all four subjects in the treatment group itself. The sixth graders of the treatment group scored a statistically significant

higher average in wave three compared to students in the same group during wave one across all four subjects, namely Arabic Reading, Arabic Writing, English Reading, English Writing, Mathematics, and Science (all standing at $p = 0$). Some of the sixth graders of the control group schools showed increase in achievement scores from wave one to three, but such increases were substantially and consistently lower than the increase which happened in the treatment group schools.

⁴² Change ratio is calculated as: $\frac{((\text{Post} - \text{Pre})/\text{Pre}) * 100}$. This ratio is used to show how the change percentage compared to baseline.

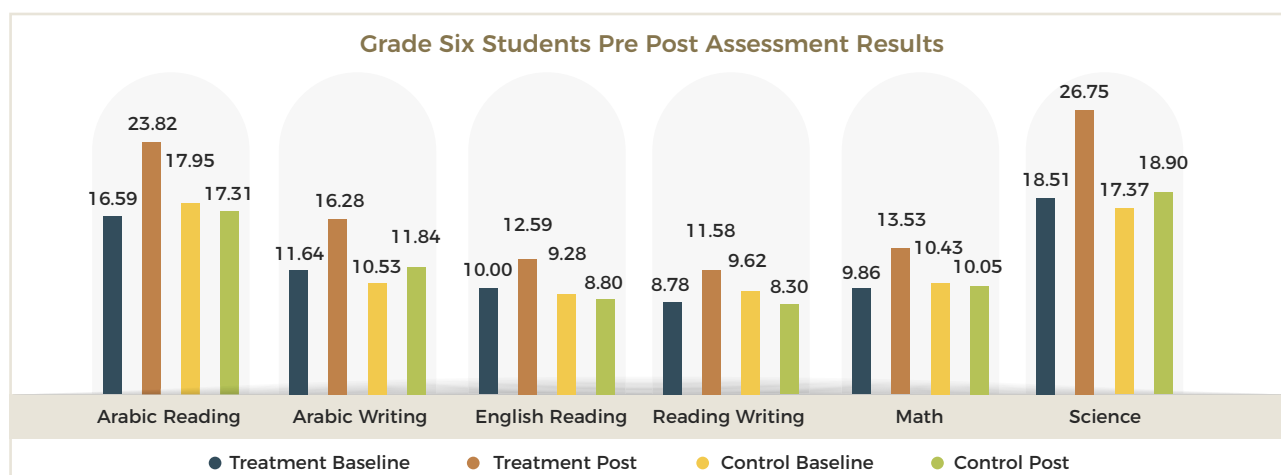


Grade Ten Students' Overall Results

On average, students' scores across all schools for grade ten increased in wave three (post-assessment) compared to wave one (pre-assessment) for all subjects. Average scores for the different subjects clearly indicated the significant progression from wave one to wave three results between the treatment and control groups. The treatment group schools' overall averages increased across all subjects, while the control group's scores did not. The highest increase in

the treatment group schools was in Science, followed by Arabic Reading, Arabic Writing, Mathematics, English Writing, and English Reading. In the control group schools, Arabic Writing showed the highest increase, which was moderate, followed by a small increase in Science, and decreases in Arabic Reading, English Reading, English Writing and Mathematics results.

Grade Ten	Arabic Reading	Arabic Writing	English Reading	English Writing	Math	Science
Change ratio in Treatment	44%	40%	26%	32%	37%	45%
Change ratio in Control	-4%	12%	-5%	-14%	-4%	9%



This high positive change in tenth graders' achievements in wave three also revealed a statistically significant difference in Arabic Reading, Arabic Writing, English Writing, and Science (all standing at $p = 0$) in favor of the treatment group, while no statistically significant difference was spotted in English Reading ($p = 8.3e-2$) or Mathematics ($p = 4.83e-1$). Hence, moving forward from wave one to three, this positive change was also evident across all four subjects in the treatment group itself. The tenth graders of the treatment group schools scored a statistically significant higher average in wave three compared to students of the same group in wave one for all subjects, except for one; Arabic Reading, Arabic Writing, English Writing, and Science (all standing at $p = 0$), and Mathematics at ($p = 7e-3$), whereas treatment group students who took the English Reading tests did not show a statistically significant difference in their achievement scores between waves one and three ($p = 8.3e-2$). Some of the tenth graders of the control group schools showed increases in achievement scores from wave one to three, but such increases were substantially and consistently lower than the increase which happened in the treatment group schools.

For both grades six and ten, the directorates constituted a significant source of variation, as the

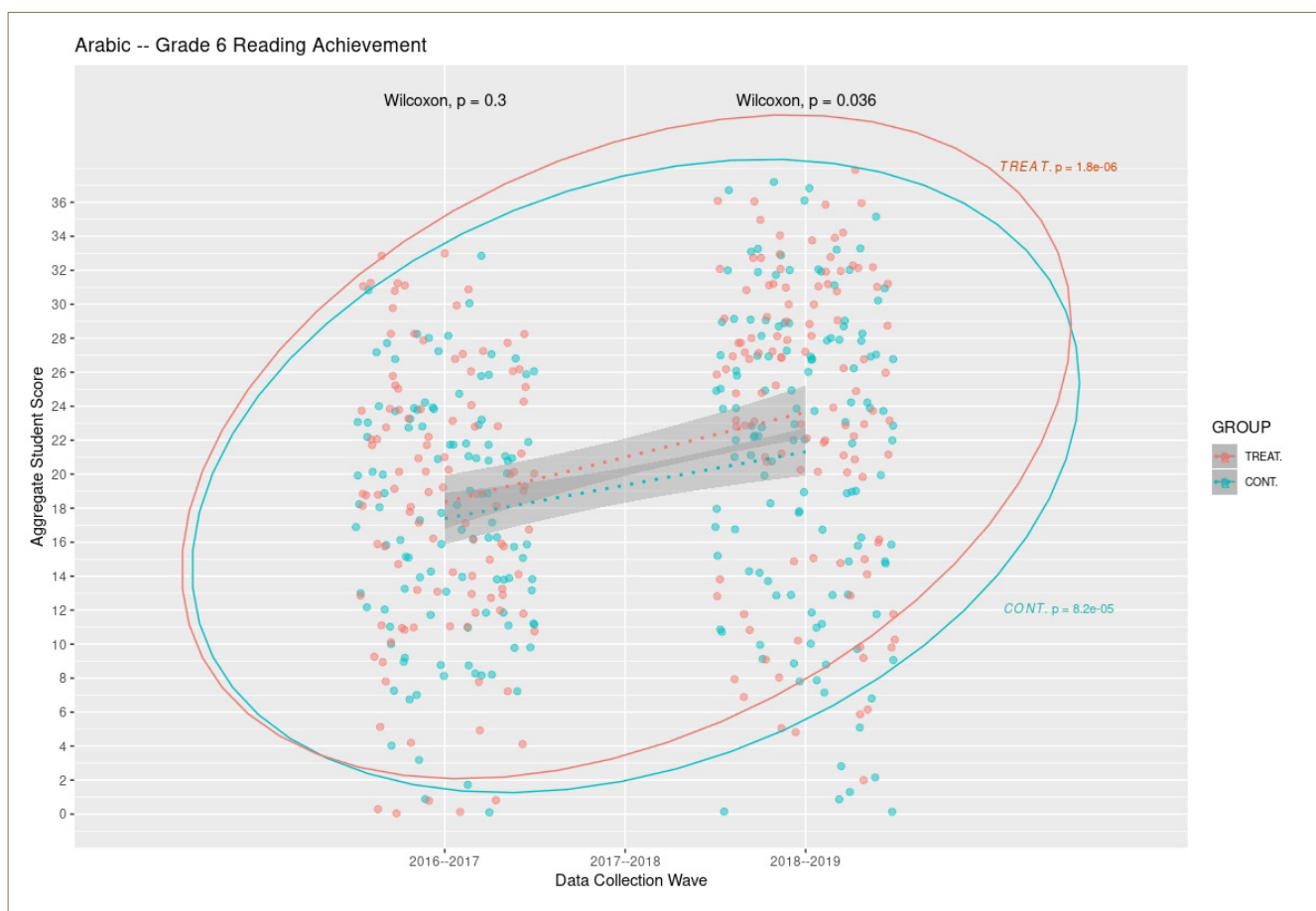
northern and central regions achieved higher scores in various subjects. Gender was also a substantial source of variation, as female students achieved higher scores than their male counterparts across all four subjects for grade six. Meanwhile, female students in grade ten achieved higher scores than their male counterparts in all subjects - except for Mathematics, where male students took the lead. English Writing results also showed no statistically significant differences ($p = 1.22e-1$) even with female students scoring higher than males. Reasons behind this variation are expected to be due to the higher level of commitment that participants in northern and central regions showed compared to the south and the higher level of commitment female participants displayed compared to males in attending all the training program activities. Also, instructional trainers observed greater and more frequent use of the newly acquired skills inside the classrooms in the northern and central regions and in girls' schools. In addition, although there was no strong evidence on the effectiveness of PLCs, instructional trainers' ongoing visits to schools throughout the course of the study showed that PLCs were, in general, more robust in the northern and central regions compared to the south and more substantial in girls' schools compared to boys' schools.



Grade Six Arabic Reading

A total of 219 students from both the treatment and control group schools completed the Arabic Reading test in wave one, and 236 students completed the same test in wave three. Grouping (treatment, control) was not a significant source of variance ($p = 3.06e-01$). That said, from the combined data of waves one and three, the treatment group schools scored higher ($M = 22.62$) than the control group schools ($M = 18.64$). Wave one (pre-assessment baseline) treatment group result was ($M = 17.63$) and increased in wave three (post-assessment) to ($M = 27.62$), while wave one control

group result was ($M = 16.63$) and increased in wave three to ($M = 20.64$). Comparison of data collected in different waves was very significant ($p = 0$). Students in wave one ($M = 17.13$), regardless of the group, scored lower average scores than students in wave three ($M = 23.90$). Students' gender was a significant source of variance ($p = 0$), with females scoring higher ($M = 21.96$) than males ($M = 19.45$). The difference between directorates was also significant ($p = 1e-03$), with north/ Irbid scoring ($M = 20.52$), central/ Amman ($M = 21.52$), and south/ Karak ($M = 19.79$).



Positive slope tendencies are clearly seen in the data ellipses and regression lines, as illustrated in the figure. Analysis results supported our assumption that the treatment and control group samples were comparable in wave one⁴³ but not wave three⁴⁴. Statistical tests of responses by data collection wave indicated no significant difference between groups in wave one ($p = 3e-01$), but a change and significant differences⁴⁵ in wave three ($p = 3.6e-02$) between

the control and treatment group results were noted, indicating a greater progression within the treatment group and a difference in favor of the treatment group. Statistical tests of students' results across data collection waves showed significant differences in both the treatment group ($p = 1.8e-06$) and control group ($p = 8.2e-05$) but smaller progression within the control group. This change can be attributed to the SNP for the treatment group.

⁴³ This means that students' achievement scores in both treatment and control group schools showed a statistical similarity in wave one (pre-assessment baseline) and started the study from the same point. This way, the change that happens in wave three in these group using the RCT design of this study will be statistically studied and attributed to the impact of the SNP.

⁴⁴ This means that there is a statistical difference between the treatment and control groups in wave three (post-assessment), which means there is change. This change in the treatment group is attributed to the SNP due to using the RCT as a design for this study.

⁴⁵ Significance difference tests were calculated at confidence level 95% and error margin at 5% for all exams across the four subjects.

Grade Six Arabic Writing

A total of 449 students from both the treatment and control group schools completed the Arabic Writing test in wave one, and 479 students completed the same test in wave three. Writing one and two tests' data were combined and modeled as a single dataset using a partial credit model. Findings showed that Arabic Writing tests one and two differed significantly ($p = 0$). Writing test one results (waves one and three, and the treatment and control groups combined) were associated with a lower average total score ($M = 05.89$) compared to Writing test two ($M = 06.99$). Grouping (treatment, control) also significantly differed ($p = 0$). Students in the treatment group schools scored higher ($M = 07.52$) than students in the

control group schools ($M = 05.42$). Wave one result for the treatment group was ($M = 05.33$) and increased in wave three to ($M = 09.56$), while wave one result for the control group was ($M = 04.97$) and increased in wave three to ($M = 05.81$). Comparison of data collected in different waves was also very significant ($p = 0$). Students in wave one ($M = 05.15$), regardless of the group, scored lower average scores than students in wave three ($M = 07.60$). Students' gender was a very significant source of variance ($p = 0$), with females scoring higher ($M = 08.65$) than males ($M = 04.40$). The difference between directorates was also significant ($p = 0$), with north/ Irbid scoring ($M = 05.63$), central/ Amman ($M = 07.49$), and south/ Karak ($M = 06.26$).



Positive slope tendencies are clearly seen in the data ellipses and regression lines, as illustrated in the figure. A positive slope exists only in the analysis of the treatment group. Analysis results supported our assumption that the treatment and control group samples were comparable in wave one but not wave three. Statistical tests of responses by data collection wave indicated no significant difference between groups in wave one ($p = 1.7e-01$), but a

change and significant differences in wave three ($p = 2.6e-08$) between the control and treatment results were noted, indicating a greater progression within the treatment group and a difference in favor of the treatment group. Statistical tests of students' results across data collection waves indicated significant differences in the treatment group ($p = 7.8e-09$) but not in the control group ($p = 1.3e-01$). This change can be attributed to the SNP for the treatment group.

Grade Ten Arabic Reading

A total of 167 students from both the treatment and control group schools completed the Arabic Reading test in wave one, and 184 students completed the same test in wave three. Grouping (treatment, control) was a significant source of variance ($p = 0$). From data in waves one and three combined, students in the treatment group schools scored higher ($M = 20.28$) than students in the control group schools ($M = 17.61$). Wave one result for the treatment group was ($M = 16.59$) and increased in wave three to ($M = 23.82$), while wave one result for the control group was (M

$= 17.95$) and decreased in wave three to ($M = 17.31$). Comparison of data collected in different waves was very significant ($p = 0$). Students in wave one ($M = 17.35$), regardless of the group, scored lower average scores than students in wave three ($M = 20.03$). Students' gender was a very significant source of variance ($p = 0$), with females scoring higher ($M = 20.46$) than males ($M = 15.96$). The difference between directorates was also significant ($p = 3.6e-02$), with north/ Irbid scoring ($M = 18.76$), central/ Amman ($M = 19.54$), and south/ Karak ($M = 16.90$).



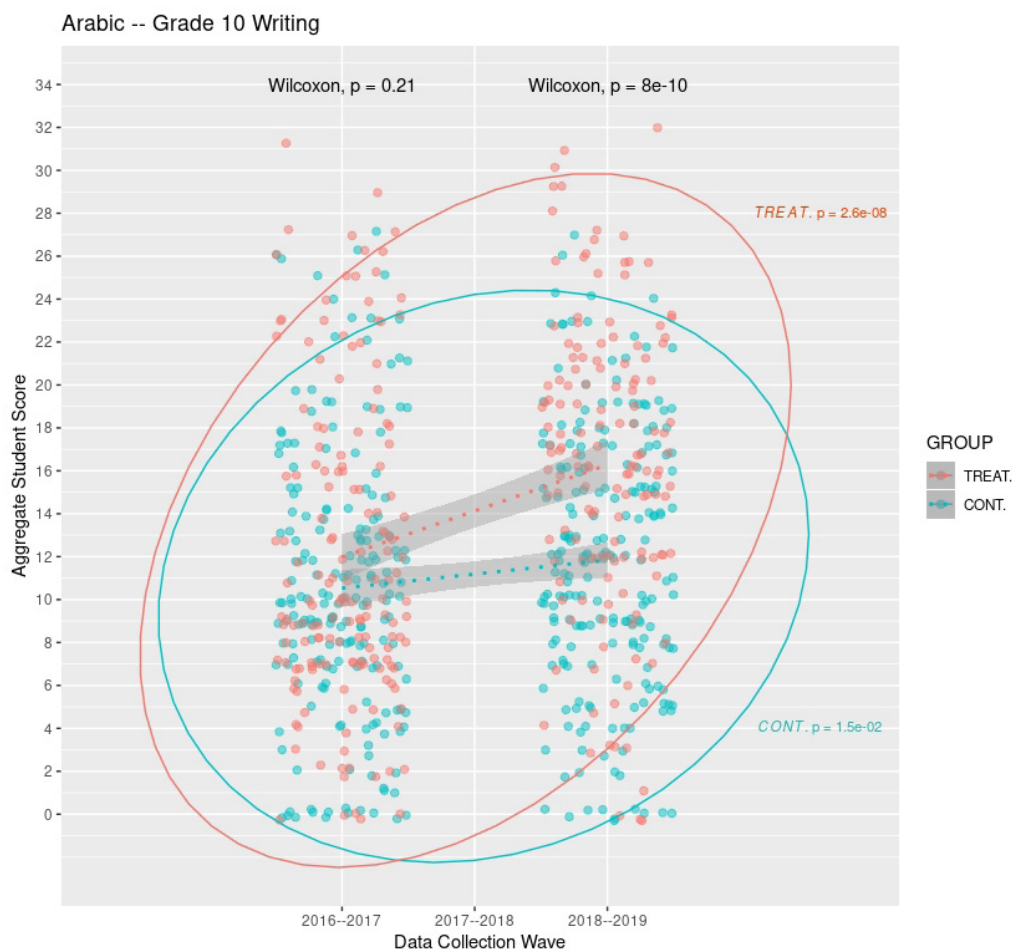
Positive slope tendencies are clearly seen in the data ellipses and regression lines, as illustrated in the figure. A positive slope exists only in the analysis of the treatment group. Analysis results supported our assumption that the treatment and control group samples were comparable in wave one but not wave three. Statistical tests of responses by data collection wave indicated no significant difference between groups in wave one ($p = 1.8e-01$), but a change and very significant differences in wave three ($p = 1.0e-08$)

between the control and treatment results, which indicates a greater progression within the treatment group and a difference in favor of the treatment group. Statistical tests of students' results across data collection waves indicated significant differences in the treatment group ($p = 3.2e-08$) but not in the control group ($p = 3.2e-01$), for which the average score **decreased**. This change can be attributed to the SNP for the treatment group.

Grade Ten Arabic Writing

A total of 343 students from both the treatment and control group schools completed the Arabic Writing test in wave one, and 356 students completed the same test in wave three. Writing one and two tests' data were combined and modeled as a single dataset using a partial credit Item Response Theory (IRT) model. Findings showed that Arabic Writing tests one and two significantly differed ($p = 0$). Writing test one results (in waves one and three, with the treatment and control groups combined) were associated with a higher average total score ($M = 13.33$) compared to Writing test two ($M = 11.38$). Grouping (treatment, control) also significantly differed ($p = 0$). Students in the treatment group schools scored higher ($M = 13.91$) than students in the control group schools (M

$= 11.22$). Wave one result for the treatment group was ($M = 11.64$) and increased in wave three to ($M = 16.28$), while wave one result for the control group was ($M = 10.53$) and increased in wave three to ($M = 11.84$). Comparison of data collected in different waves was also very statistically significant ($p = 0$). Students in wave one ($M = 11.02$), regardless of the group, scored lower average scores than students in wave three ($M = 13.65$). Students' gender was a very significant source of variance ($p = 0$), with females scoring higher ($M = 14.90$) than males ($M = 08.24$). The difference between directorates was also significant ($p = 0$), with north/ Irbid scoring ($M = 12.26$), central/ Amman ($M = 13.21$), and south/ Karak ($M = 11.06$).



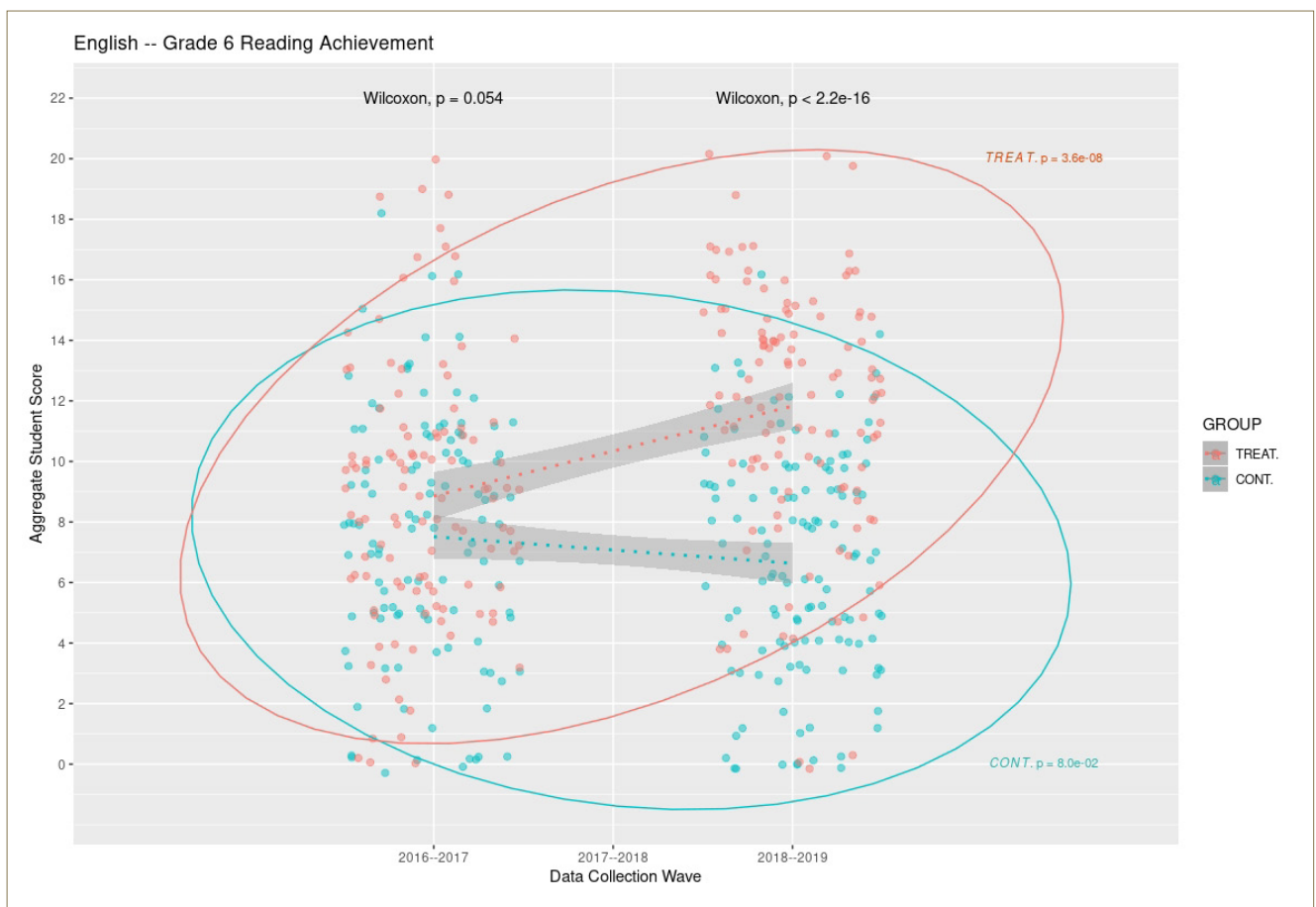
Positive slope tendencies are clearly seen in the data ellipses and regression lines, as illustrated in the figure. A positive slope exists only in the analysis of the treatment group. Analysis results supported our assumption that the treatment and control group samples were comparable in wave one but not wave three. Statistical tests of responses by data collection wave indicated no significant difference between groups in wave one ($p = 2.1e-01$), but a change and significant differences in wave three ($p = 8.0e-10$)

between the control and treatment results, which indicates a greater progression within the treatment group and a difference in favor of the treatment group. Statistical tests of students' results across data collection waves indicated significant differences in both the treatment group ($p = 2.6e-08$) and control group ($p = 1.5e-02$) but smaller progression among the control group. This change can be attributed to the SNP for the treatment group.

En Grade Six English Reading

A total of 214 students from both the treatment and control group schools completed the English Reading test in wave one, and 242 students completed the same test in wave three. Grouping (treatment, control) was a significant source of variance ($p = 0$). From data in waves one and three combined, students in the treatment group schools scored higher ($M = 10.41$) than students in the control group schools ($M = 07.03$). Wave one result for the treatment group was ($M = 08.85$) and increased in wave three to ($M = 11.83$), while wave one result for the control group was (M

$= 07.49$) and **decreased** in wave three to ($M = 06.63$). Comparison of data collected in different waves was very significant ($p = 0$). Students in wave one ($M = 08.16$), regardless of the group, scored lower average scores than students in wave three ($M = 09.12$). Students' gender was a very significant source of variance ($p = 0$), with females scoring higher ($M = 09.88$) than males ($M = 07.55$). The difference between directorates was also significant ($p = 4.0e-02$), with north/ Irbid scoring ($M = 08.18$), central/ Amman ($M = 09.24$), and south/ Karak ($M = 08.61$).



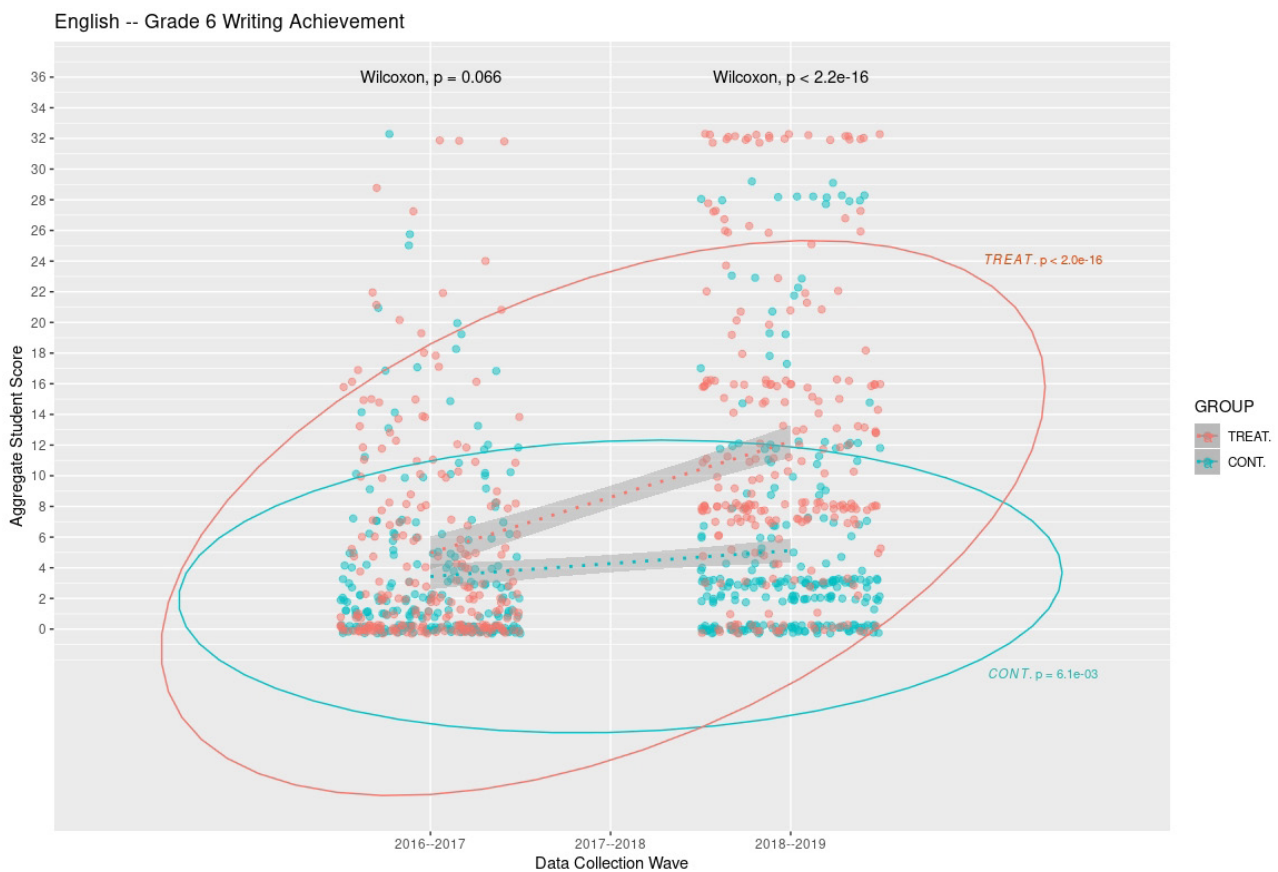
Positive slope tendencies are clearly seen in the data ellipses and regression lines, as illustrated in the figure. A positive slope exists only in the analysis of the treatment group. Analysis results supported our assumption that the treatment and control group samples were comparable in wave one but not wave three. Statistical tests of responses by data collection wave indicated no significant difference between groups in wave one ($p = 5.4e-02$), but a change and significant differences in wave three ($p = 2.2e-16$)

between the control and treatment results, which indicates a greater progression within the treatment group and a difference in favor of the treatment group. Statistical tests of students' results across data collection waves indicated significant differences in both the treatment group ($p = 3.6e-08$) and control group ($p = 8.0e-02$), but a smaller change among the control group, for which the average score **decreased**. This change can be attributed to the SNP for the treatment group.

En Grade Six English Writing

A total of 433 students from both the treatment and control group schools completed the English Writing test in wave one, and 445 students completed the same test in wave three. Writing one and two tests' data were combined and modeled as a single dataset using a partial credit model. Findings showed that English Writing tests one and two significantly differed ($p = 0$). Writing test one results (waves one and three, with the treatment and control groups combined) were associated with a lower average total score ($M = 05.77$) compared to Writing test two ($M = 06.92$). Grouping (treatment, control) also significantly differed ($p = 0$), as students in the treatment group schools scored substantially higher ($M = 08.57$) than students in the control group schools ($M = 04.27$). Wave

one result for the treatment group was ($M = 04.67$) and increased in wave three to ($M = 12.24$), while wave one result for the control group was ($M = 03.26$) and increased in wave three to ($M = 05.12$). Comparison of data collected in different waves was also very significant ($p = 0$). Students in wave one ($M = 03.96$), regardless of the group, scored lower average scores than students in wave three ($M = 08.48$). Students' gender was a very significant source of variance ($p = 0$), with females scoring higher ($M = 08.24$) than males ($M = 04.51$). The difference between directorates was also significant ($p = 0$), with north/ Irbid scoring ($M = 05.60$), central/ Amman ($M = 07.39$), and south/ Karak ($M = 06.05$).



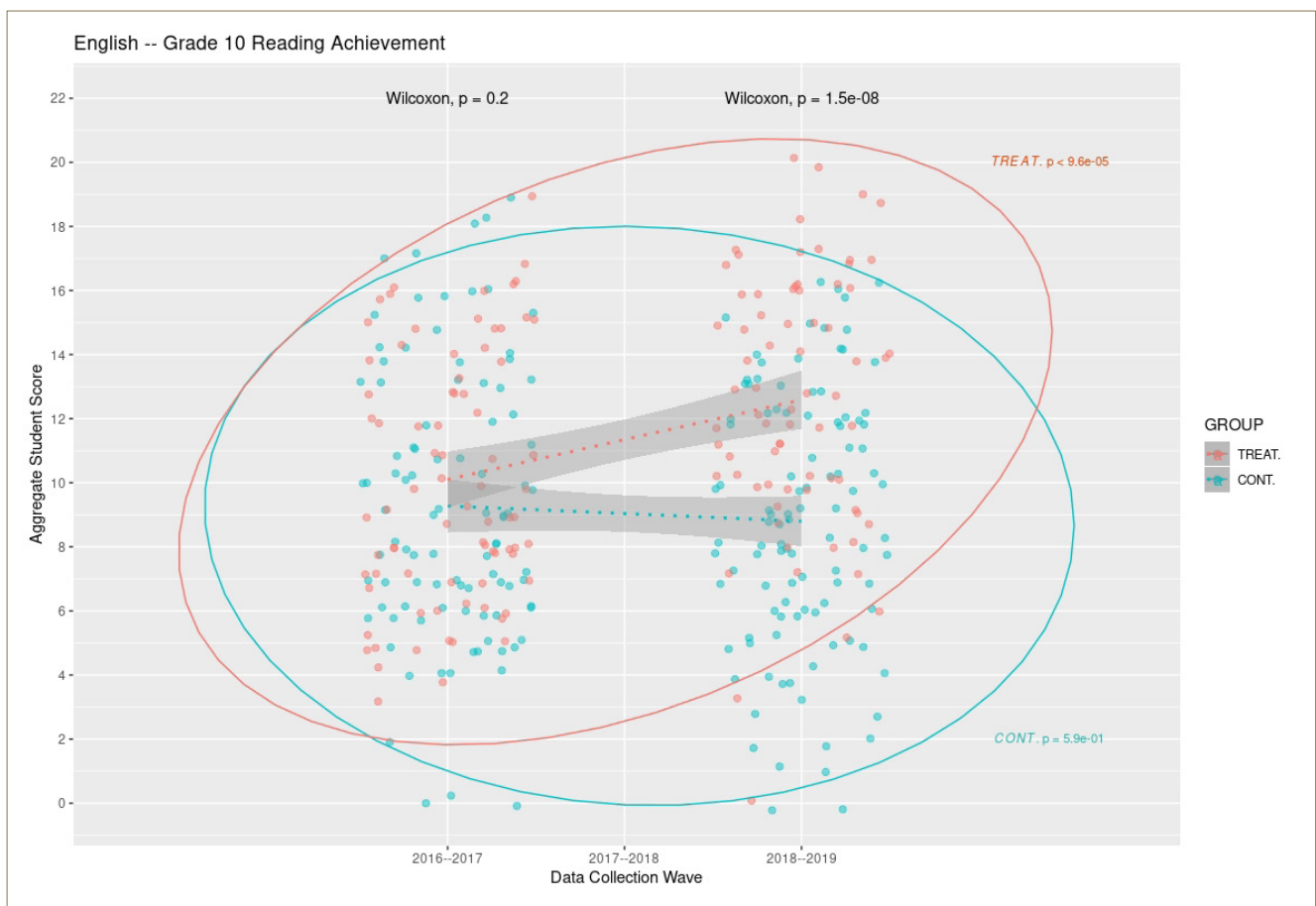
Positive slope tendencies are clearly seen in the data ellipses and regression lines, as illustrated in the figure. A positive slope exists only in the analysis of the treatment group. Analysis results supported our assumption that the treatment and control group samples were comparable in wave one but not wave three. Statistical tests of responses by data collection wave indicated no significant difference between groups in wave one ($p = 6.6e-02$), but a change and highly significant differences in wave three ($p < 2.2e-$

16) between the control and treatment results, which indicates a greater progression within the treatment group and a difference in favor of the treatment group. Statistical tests of students' results across data collection waves indicated significant differences in both the treatment group ($p = 2.0e-16$) and control group ($p = 6.0e-03$), but smaller change among the control group. This change can be attributed to the SNP for the treatment group.

En Grade Ten English Reading

A total of 167 students from both the treatment and control group schools completed the English Reading test in wave one, and 178 students completed the same test in wave three. Grouping (treatment, control) was not a significant ($p = 8.3e-02$) source of variance. From data in waves one and three combined, students in the treatment group schools scored higher ($M = 11.29$) than students in the control group schools ($M = 09.06$). Wave one result for the treatment group was ($M = 10.00$) and increased in wave three to ($M = 12.59$), while wave one for the

control group was ($M = 09.28$) and **decreased** in wave three to ($M = 08.80$). Comparison of data collected in different waves was significant ($p = 3.3e-02$). Students in wave one ($M = 09.60$), regardless of the group, scored lower average scores than students in wave three ($M = 10.35$). Students' gender was also a significant source of variance ($p = 0$), with females scoring higher ($M = 10.48$) than males ($M = 09.18$). The difference between directorates was also significant ($p = 0$), with north/ Irbid scoring ($M = 09.70$), central/ Amman ($M = 11.38$), and south/ Karak ($M = 08.47$).



Positive slope tendencies are clearly seen in the data ellipses and regression lines, as illustrated in the figure. A positive slope exists only in the analysis of the treatment group. Analysis results supported our assumption that the treatment and control group samples were comparable in wave one but not wave three. Statistical tests of responses by data collection wave indicated no significant difference between groups in wave one ($p = 2.0e-01$), but a change and highly significant differences in wave three ($p = .5e-$

08) between the control and treatment results, which indicates a greater progression within the treatment group and a difference in favor of the treatment group. Statistical tests of students' results across data collection waves indicated significant differences in the treatment group ($p = 9.6e-05$) but not in the control group ($p = 5.9e-01$), for which the average score **decreased**. This change can be attributed to the SNP for the treatment group.

En Grade Ten English Writing

A total of 334 students from both the treatment and control group schools completed the English Writing test in wave one, and 367 students completed the same test in wave three. Writing one and two tests' data were combined and modeled as a single dataset using a partial credit model. Findings showed that English Writing tests one and two significantly differed ($p = 0$). Writing test one results (waves one and three, and with the treatment and control groups combined) were associated with a higher average total score ($M = 11.04$) compared to Writing test two ($M = 07.98$). Grouping (treatment, control) also significantly differed ($p = 0$). Students in the treatment group schools scored substantially higher ($M = 10.19$) than students in the control group schools ($M = 08.91$). Wave one result for

the treatment group was ($M = 08.78$) and increased in wave three to ($M = 11.58$), while wave one result for the control group was ($M = 09.62$) and **decreased** in wave three to ($M = 08.30$). Comparison of data collected in different waves was also very significant ($p = 0$). Students in wave one ($M = 09.25$), regardless of the group, scored slightly lower average scores than students in wave three ($M = 09.63$). Students' gender was not a significant source of variance ($p = 1.22e-01$), but females still scored higher ($M = 11.73$) than males ($M = 05.79$). The difference between directorates was significant ($p = 2.9e-02$), with north/ Irbid scoring ($M = 09.62$), central/ Amman ($M = 09.77$), and south/ Karak ($M = 07.99$).



Positive slope tendencies are clearly seen in the data ellipses and regression lines, as illustrated in the figure. A positive slope exists only in the analysis of the treatment group. Analysis results supported our assumption that the treatment and control group samples were comparable in wave one but not wave three. Statistical tests of responses by data collection wave indicated no significant difference between groups in wave one ($p = 8.1e-01$), but a change and highly significant differences in wave three ($p = 1.6e-03$) between the control and treatment results, which indicates a greater progression within the treatment

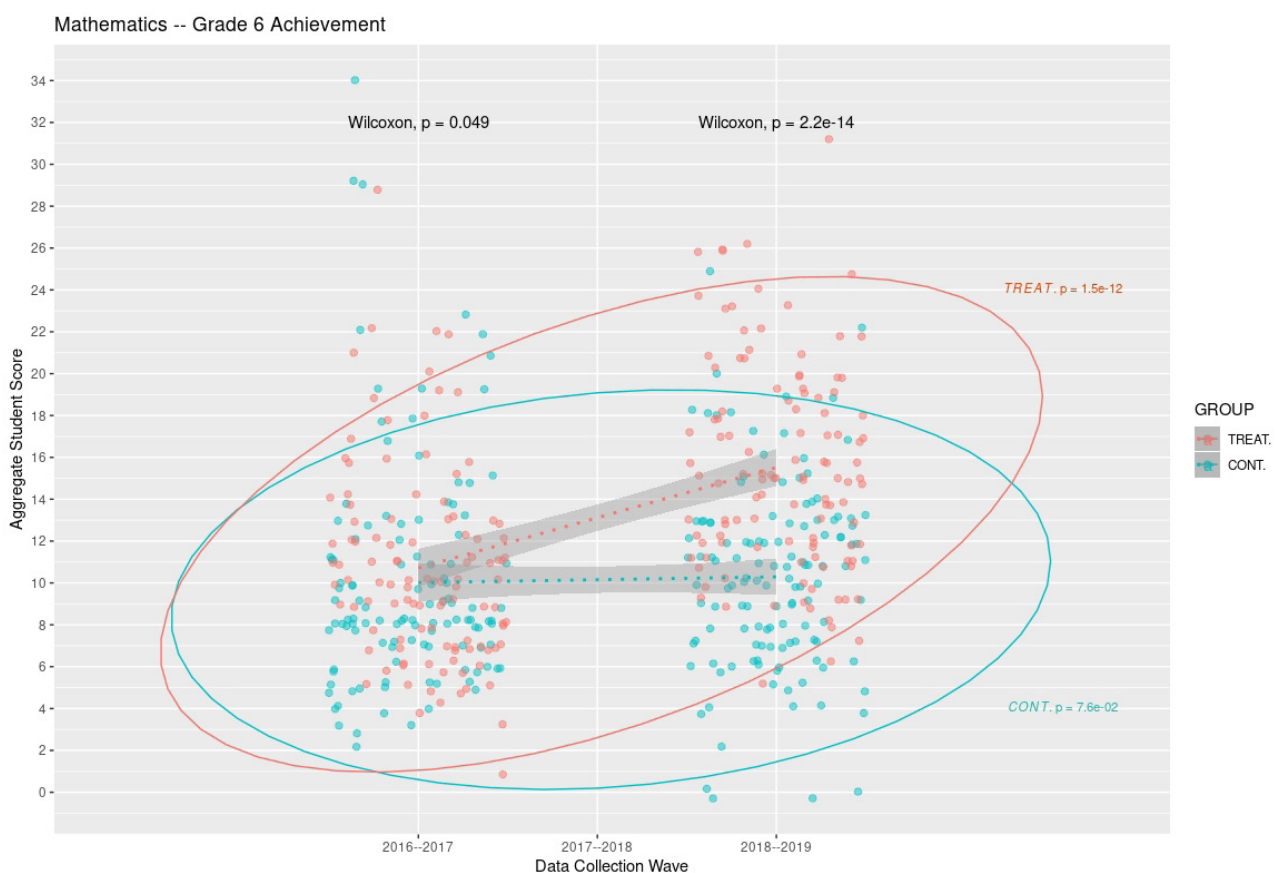
group and a difference in favor of the treatment group. Statistical tests of students' results across data collection waves indicated no significant differences in both the treatment group ($p = 1.2e-01$) and control group ($p = 2.1e-01$). Although average scores of all students changed over time, on average, they increased for the treatment group and **decreased** for the control group. The treatment group change cannot be solely attributed to the SNP - as other factors such as test, group, data collection wave, directorate, or any combination of factors may have contributed to this change.



Grade Six Mathematics

A total of 217 students from both the treatment and control group schools completed the Mathematics test in wave one, and 243 students completed the same test in wave three. Grouping (treatment, control) was a significant source of variance ($p = 0$). From data in waves one and three combined, students in the treatment group schools scored higher ($M = 15.36$) than students in the control group schools ($M = 10.19$). Wave one result for the treatment group was ($M = 10.75$) and increased in wave three to ($M = 19.57$), while wave one result for the control group was ($M = 10.08$) and slightly increased in wave

three to ($M = 10.29$). Comparison of data collected in different waves was very significant ($p = 0$). Students in wave one ($M = 10.41$), regardless of the group, scored lower average scores than students in wave three ($M = 14.72$). Students' gender was a significant source of variance ($p = 0$), with females scoring slightly higher ($M = 12.90$) than males ($M = 12.50$). The difference between directorates was not significant ($p = 7.95e-01$), with average scores being in north/ Irbid ($M = 11.55$), central/ Amman ($M = 13.42$), and south/ Karak ($M = 13.47$).



Positive slope tendencies are clearly seen in the data ellipses and regression lines, as illustrated in the figure. A positive slope exists only in the analysis of the treatment group. In the case of grade six Mathematics, analysis results did not support our assumption that the treatment and control group samples were comparable⁴⁶ in wave one, but it supported our assumptions that the treatment and control group samples were not comparable⁴⁷ in wave three. Statistical tests of responses by data collection wave indicated that there was a significant difference

between groups in wave one ($p = 4.9e-02$), as well as a change and significant differences in wave three ($p = 2.2e-14$) between the control and treatment results, which indicates a greater progression within the treatment group and a difference in favor of the treatment group. Statistical tests of students' results across data collection waves indicated significant differences in the treatment group ($p = 1.5e-12$) but not in the control group ($p = 7.6e-02$), for which the average score increased very slightly. This change can be attributed to the SNP for the treatment group.

⁴⁶ This means that statistical analysis results showed that the treatment and control groups did not have statistical similarities in wave one when the study began.

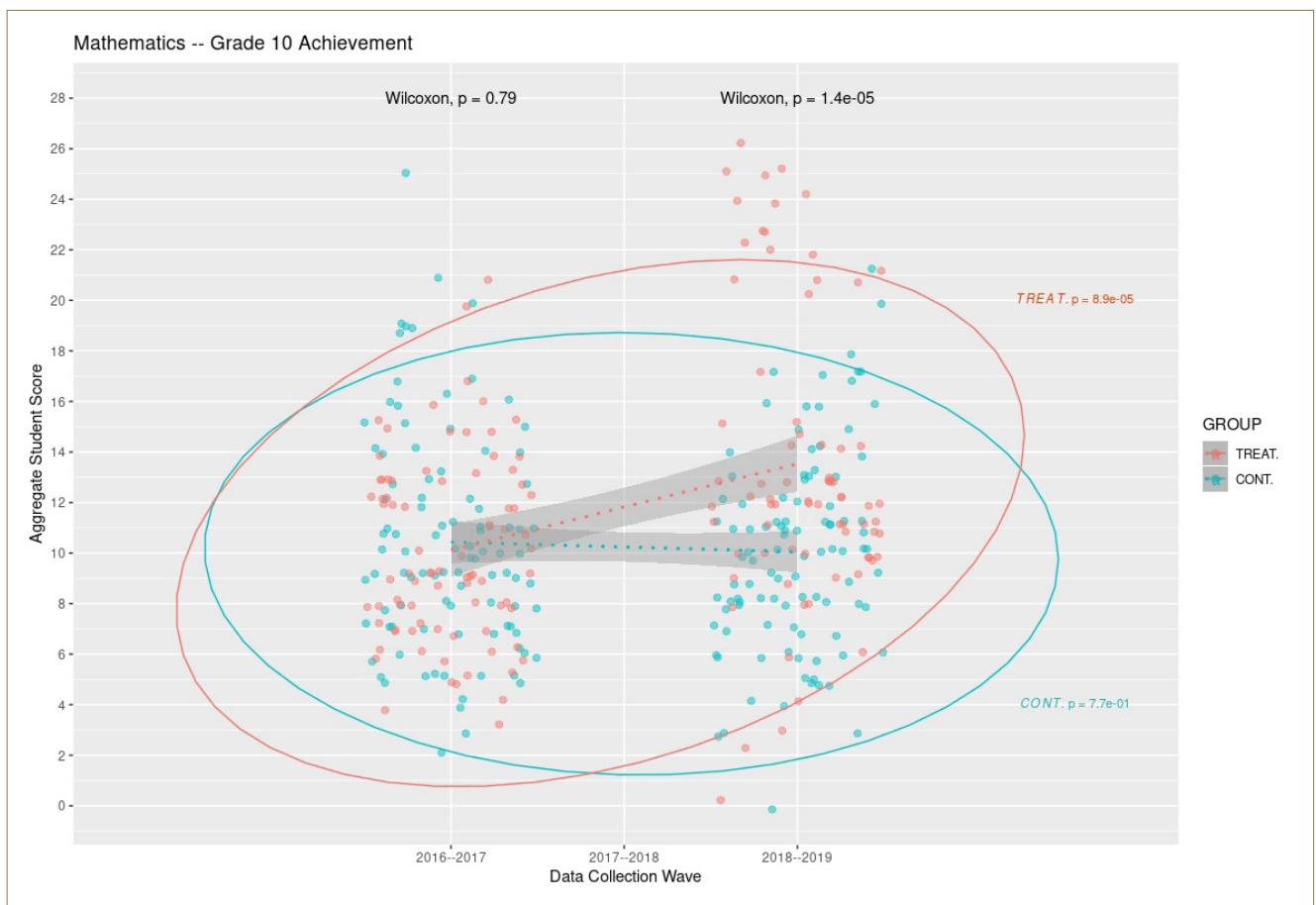
⁴⁷ This means there is a statistical difference between the groups in wave three - which in the case of the treatment group is attributed to the SNP.



Grade Ten Mathematics

A total of 175 students from both the treatment and control group schools completed the Mathematics test in wave one, and 176 students completed the same test in wave three. Grouping (treatment, control) was not a significant source of variance ($p = 4.88e-01$). From data in waves one and three combined, students in the treatment group schools scored higher ($M = 11.66$) than students in the control group schools ($M = 10.24$). Wave one result for the treatment group was ($M = 09.68$) and increased in wave three to ($M = 13.53$), while wave one result for the control

group was ($M = 10.43$) and **decreased** slightly in wave three to ($M = 10.05$). Comparison of data collected in different waves was significant ($p = 0$). Students in wave one ($M = 10.18$), regardless of the group, scored lower average scores than students in wave three ($M = 11.53$). Students' gender was a significant source of variance ($p = 0$), with females scoring higher ($M = 10.67$) than males ($M = 11.15$). The difference between directorates was a significant variable ($p = 0$), with north/ Irbid scoring ($M = 10.82$), central/ Amman ($M = 11.92$), and south/ Karak ($M = 08.85$).



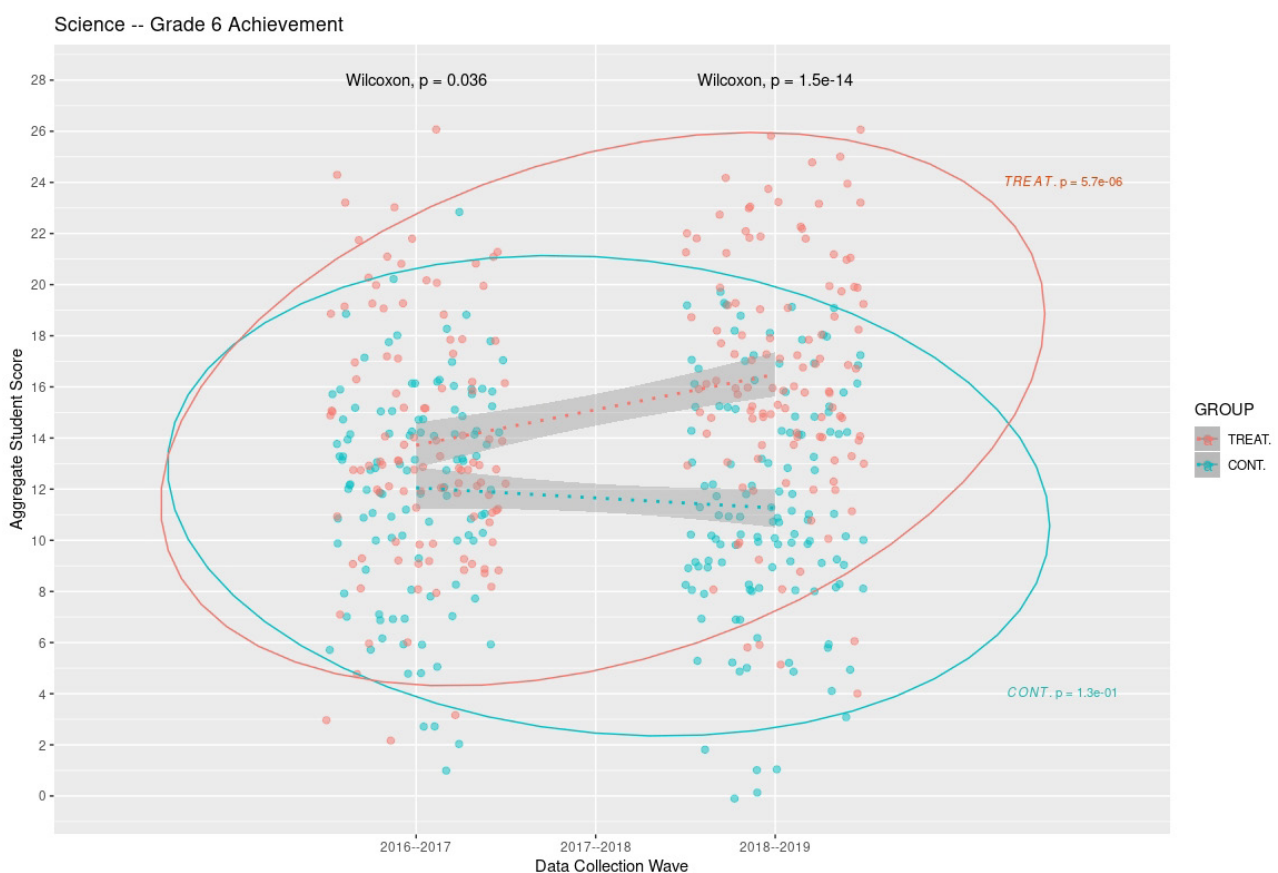
Positive slope tendencies are clearly seen in the data ellipses and regression lines, as illustrated in the figure. A positive slope exists only in the analysis of the treatment group. Analysis results supported our assumption that the treatment and control group samples were comparable in wave one but not wave three. Statistical tests of results by data collection wave indicated no significant difference in wave one ($p = 7.9e-01$), but a change and significant differences

in wave three ($p = 1.4e-05$) between the control and treatment results, which indicates a difference in favor of the treatment group. Statistical tests of responses by data collection wave indicated significant differences in the treatment group ($p = 8.9e-05$) but not in the control group ($p = 7.7e-01$), for which the average score slightly **decreased**. This change can be attributed to the SNP for the treatment group.

Grade Six Science

A total of 214 students from both the treatment and control group schools completed the Science test in wave one, and 242 students completed the same test in wave three. Grouping (treatment, control) was a significant source of variance ($p = 0$). From data in waves one and three combined, students in the treatment group schools scored higher ($M = 15.14$) than students in the control group schools ($M = 11.62$). Wave one result for the treatment group was ($M = 13.65$) and increased in wave three to ($M = 16.46$), while wave one result for the control group was ($M = 12.05$) and **decreased** in wave three to (M

$= 11.26$). Comparison of data collected in different waves was very significant ($p = 0$). Students in wave one ($M = 12.84$), regardless of the group, scored lower average scores than students in wave three ($M = 13.77$). Students' gender was a very significant source of variance ($p = 0$), with females scoring slightly higher ($M = 13.48$) than males ($M = 13.19$). The difference between directorates was not a significant source of variance ($p = 7.95e-01$), with north/ Irbid scoring ($M = 12.85$), central/ Amman ($M = 13.82$), and south/ Karak ($M = 13.38$).



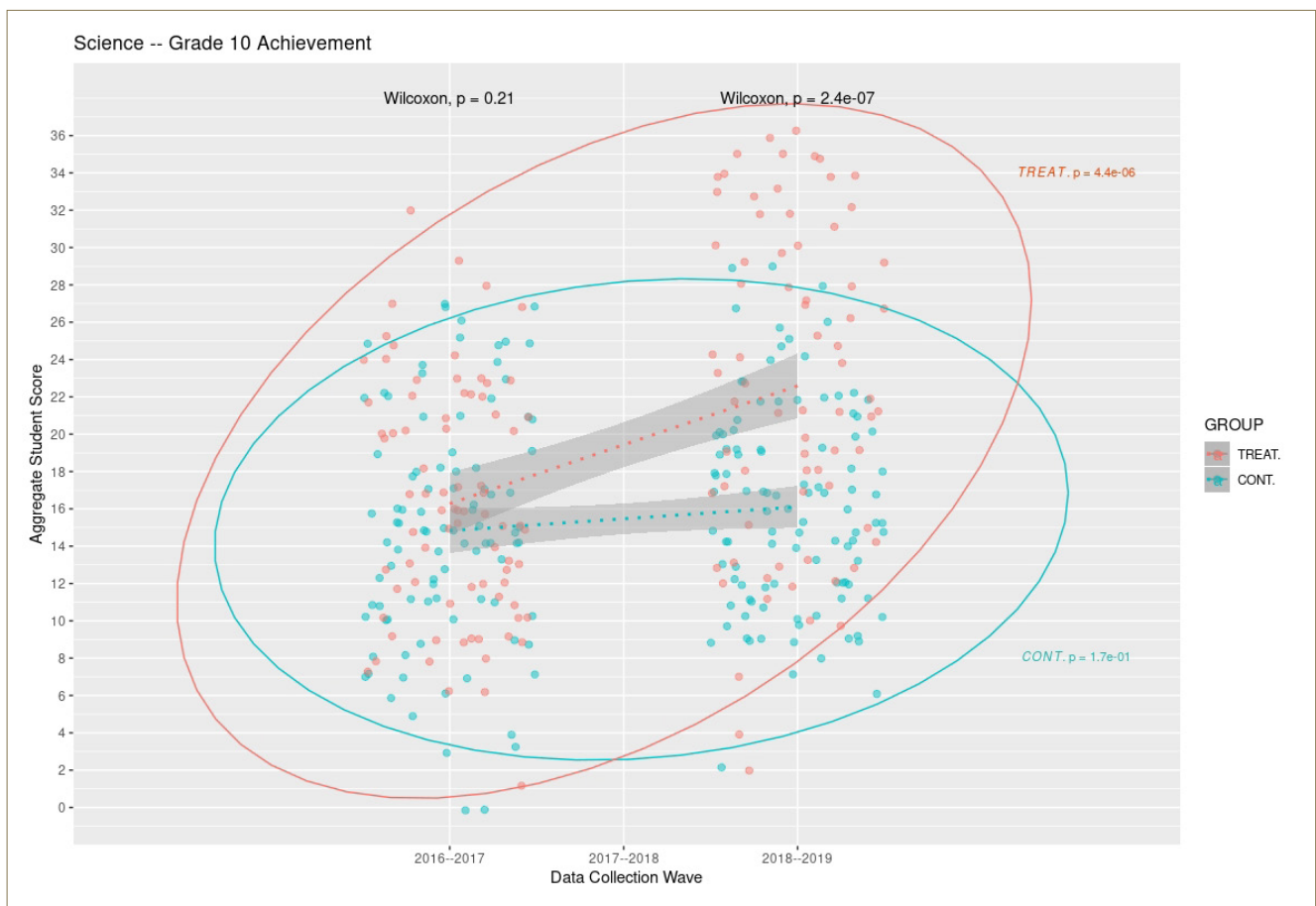
Positive slope tendencies are clearly seen in the data ellipses and regression lines, as illustrated in the figure. A positive slope exists only in the analysis of the treatment group. In the case of grade six Science, analysis results did not support our assumption that the treatment and control group samples were comparable in wave one, but it supported our assumptions that the treatment and control group samples were not comparable in wave three. Statistical tests of responses by data collection wave indicated that there was a significant difference in wave one

($p = 3.6e-01$), as well as a change and significant differences in wave three ($p = 1.5e-14$) between the control and treatment results, which indicates a greater progression within the treatment group and a difference in favor of the treatment group. Statistical tests of students' results across data collection waves indicated significant differences in the treatment group ($p = 5.7e-06$) but not in the control group ($p = 1.3e-01$), for which the average score **decreased**. This change can be attributed to the SNP for the treatment group.

Grade Ten Science

A total of 167 students from both the treatment and control group schools completed the Science test in wave one, and 178 students completed the same test in wave three. Grouping (treatment, control) was a significant source of variance ($p = 0$). From data in waves one and three combined, students in the treatment group schools scored higher ($M = 22.63$) than students in the control group schools ($M = 18.18$). Wave one result for the treatment group was ($M = 18.51$) and increased in wave three to ($M = 26.75$), while wave one result for the control group was (M

$= 17.37$) and increased in wave three to ($M = 18.90$). Comparison of data collected in different waves was very significant ($p = 0$). Students in wave one ($M = 17.87$), regardless of the group, scored lower average scores than students in wave three ($M = 22.12$). Students' gender was a very significant source of variance ($p = 0$), with females scoring higher ($M = 20.79$) than males ($M = 18.85$). The difference between directorates was also a significant source of variance ($p = 4.0e-3$), with north/ Irbid scoring ($M = 19.25$), central/ Amman ($M = 22.86$), and south/ Karak ($M = 17.98$).



Positive slope tendencies are clearly seen in the data ellipses and regression lines, as illustrated in the figure. A positive slope exists only in the analysis of the treatment group. Analysis results supported our assumption that the treatment and control group samples were comparable in wave one but not wave three. Statistical tests of responses by data collection wave indicated no significant difference between groups in wave one ($p = 2.1e-01$), but a change and

highly significant differences in wave three ($p = 2.4e-07$) between the control and treatment results, which indicates a greater progression within the treatment group and a difference in favor of the treatment group. Statistical tests of students' results across data collection waves indicated significant differences in the treatment group ($p = 4.4e-08$) but not in the control group ($p = 1.7e-01$). This change can be attributed to the SNP for the treatment group.



QUEEN RANIA TEACHER ACADEMY
أكاديمية الملكة رانيا لتدريب المعلمين