

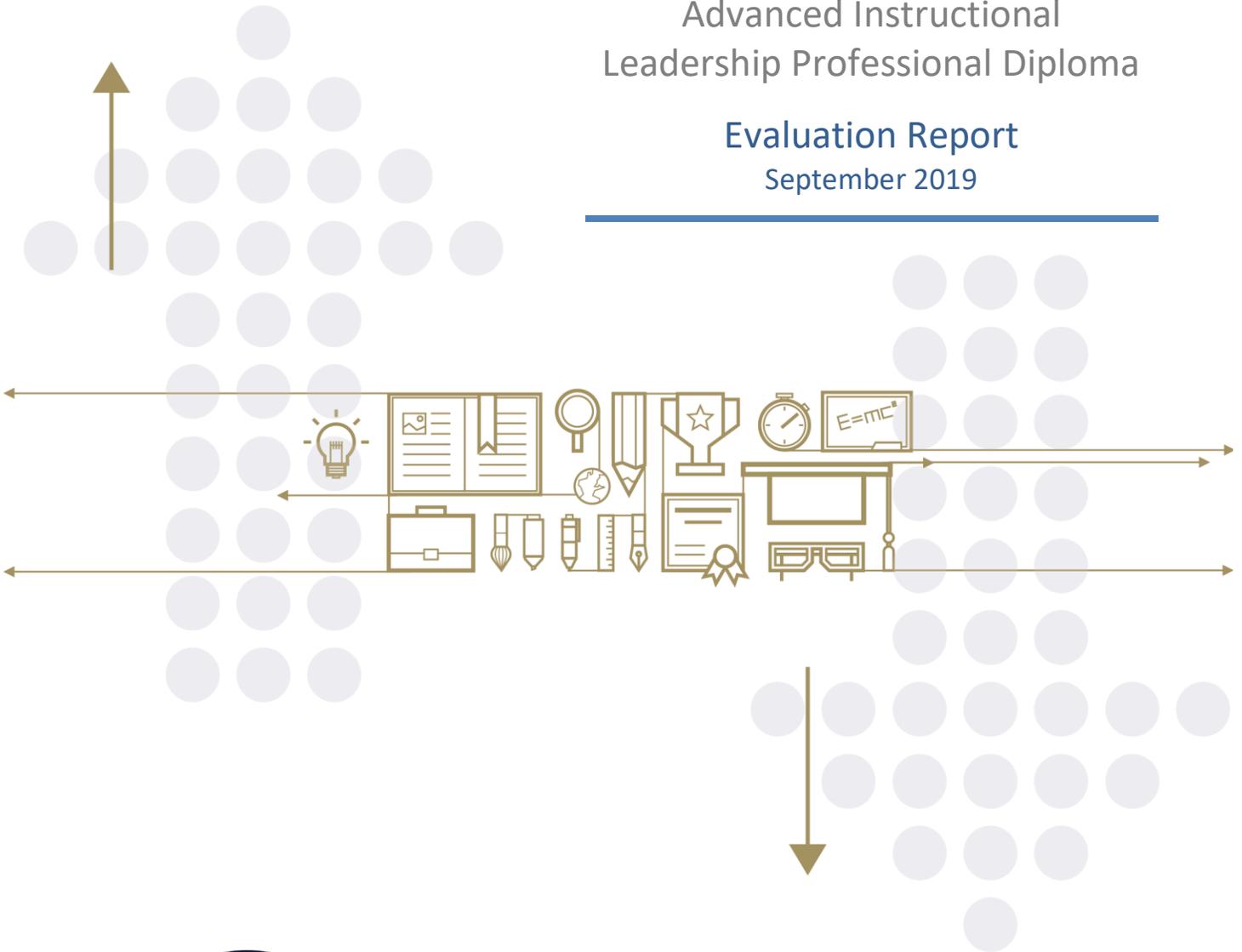


ADVANCED INSTRUCTIONAL
LEADERSHIP PROFESSIONAL DIPLOMA

QUEEN RANIA TEACHER ACADEMY
أكاديمية الملكة رانيا لتدريب المعلمين

Advanced Instructional
Leadership Professional Diploma

Evaluation Report
September 2019



Canada 

Table of Contents

Executive Summary	3
Introduction	6
Report Structure	9
Section I: Evaluating Program Impact on Participants' Knowledge and Skills	9
Methods	9
Sample	10
Data Collection and Analysis.....	13
Feedback Surveys of Module Effectiveness.....	13
Pre/Post Content Principals' Assessment.....	13
Teacher Survey of Principals' Instructional Leadership	14
Principals' Focus Groups	17
Findings.....	17
Changes to Participants' Knowledge and Skills	17
Pre/Post Content Principals' Assessment.....	17
Teacher Survey on Principals' Instructional Leadership	24
Principals' Focus Group	28
Section II: Data Collection and Analysis	29
Methods	29
Feedback Surveys of Module Effectiveness (Content and Instruction)	30
Module Content and Instruction	30
Post Module Feedback Surveys.....	30
Results of Feedback on Module Content.....	30
Principals' Focus Group	35
Results of Feedback on Instructors' Performance	35
Principals' Focus Group	40

Executive Summary

This evaluation provides evidence on the impact and effectiveness of the QRTA Advanced Leadership program by showing changes in principal knowledge, skills and instructional leadership capabilities as related to program participation. This evaluation report, prepared in collaboration by Queen Rania Teaching Academy (QRTA) Monitoring and Evaluation (M&E) team and the University of Connecticut (UConn) research team, presents cumulative findings from teams' evaluation of the effectiveness of the Advanced Leadership Training Program. There were two main issues that drove this evaluation, each with underlying research questions.

Issue 1: To assure the program's modules enhance school leaders' knowledge and skills in areas aligned with instructional leadership. For this first issue, our aligned research questions were:

1. To what degree did the program build participants' knowledge and skills associated with instructional leadership?
2. Do participants implement this new learning in effective and meaningful ways?

Issue 2: To assure that the UConn trainers and coaches develop QRTA core team members trainers to effectively deliver the modules (and modify them as needed), take full ownership of the program at the conclusion of the contract, and to maintain program quality and grow in size over time (i.e., scale up). The aligned research questions for this issue were:

1. To what extent have QRTA core team members developed the necessary content knowledge teach the modules?
2. Is their instruction of that content effective?
3. Does the program continue to thrive over time?

Methods

As the evaluation's research questions focused on both perceptual (what participants think) and behavioral (what participants and core members do) outcomes of interest, it was necessary for the evaluation team to collect multiple forms of data to understand these phenomena. Specifically, we took a mixed method approach collecting quantitative and qualitative information about the participants' views of the program, changes in their beliefs and knowledge as a result. As we describe in more detail below, we used teacher surveys, focus groups, pre/post assessments and participants' evaluation of module instruction and content as data for this evaluation. We also used a variety of analytic techniques including, in terms of statistics: simple t-tests, simple regression, and multiple regression. For the qualitative approaches, we engaged in inductive and deductive coding of open response questions and focus group responses.

For the first issue guiding this evaluation – positive changes to participants' instructional leadership – the research team took a two-pronged approach. First, we collected information on whether and to what degree module learning translated into changes in participants' knowledge

and skills. Information on changes to participants' knowledge was achieved using content-based pre-post assessments developed by UConn module instructors. We also asked participants, via focus groups with a purposeful stratified sample to create a representative group of principals, about changes they perceived to have occurred in their practice as a result of the program to positive effect. Finally, also with the use of a stratified sample, we identified a number of participants to survey their teachers regarding whether and to what degree participants' implementation of best practices in instructional leadership changed over the year. Teachers were surveyed at the beginning and end of the year (i.e., before and after their principal participated in the program) and were asked to report on his/her leadership behaviors.

For the second area of effectiveness focused on building local QRTA capacity to engage in the work, the data collection process changed over time. This is because in the first two years of implementation – the years in which the UConn instructors fully or partially taught the modules – the evaluation teams' goal was to ascertain (1) whether and to what degree core team members had gained the necessary module content knowledge and (2) their ability to think deeply about how they might instruct this material and why. As true for the principal participants, in the first year of implementation core team members' content knowledge was evaluated using the same pre/post assessment as deployed for the principal participants. For the second cohort, any new additions to the core team were also given this assessment. During this period, we also collected data via a survey on core members' views of the sessions and the degree to which they viewed the sessions as preparing them to better understand how to instruct module material.

As the UConn team pulled back on its involvement – for the second cohort the UConn team taught some of the modules and, for the third cohort the UConn team taught none – we also used participants' evaluations of the module content and instruction as a means to evaluate core members' capacity to teach the modules. In particular, we asked participants about the degree the modules were effective in terms of their content and the instructional practices deployed by the facilitators. Data sources for this portion of the evaluation came from surveys and focus group interviews of participants.

Findings

Overall, our evaluation provides robust evidence of both the positive effects of the program on participants and their schools and QRTA's capacity to sustain and grow the program overtime. Indeed, note that while the number of participants increased approximately 100% each year (33 to 60 to 127) and UConn support decreased performance improvements held or increased between the second and third cohort and was at a high level. We give a synopsis of some of the key findings below.

- Issue 1: Positive Changes to Participants' Instructional Leadership
 - Cohort 3 participants show a statistically significant positive change of approximately 17% (10 points) in their content knowledge as measured by the pre/post assessment. This is almost a 100% improvement from Cohort 1 in which the improvement was 11% (8 points) and is on par with Cohort 2 (14 points or 20% increase). There were no statistically significant differences by gender in any cohort (i.e., men and women

principals experienced the same positive results each year). This is particularly impressive given changes to the assessment which decreased the total possible point accrual by 25 points.

- Findings regarding teachers' views on participants' instructional leadership were overall quite positive. As true for Cohort 2, virtually all areas showed positive growth, and none were both statistically significant and negative. Specifically, there were three areas of instructional leadership for which teachers reported the principals, on average, made significant positive growth. These areas included, 1) whether professional development is tailored to their needs, 2) whether the principal gives regular and helpful feedback on their teaching, and 3) whether the principal takes into account the feedback the teacher provides.
- Issue 2: QRTA capacity to sustain the program effectiveness
 - Participants rated both the module content and the instruction quite highly with an average satisfaction of 87% and 89.6% respectively. These results mirror those for Cohort 2 when participants rated both the module content and the instruction quite highly with an average satisfaction of 88.9% and 88.05% respectively. Moreover, they are far higher than for Cohort 1 when the scores were 85.3% and 88.8% respectively and when the UConn instructors led the courses.
 - When disaggregating these scores by gender we find that there were sometimes small differences (always less than .4 points) between male and female participants' responses. Specifically, in Cohort 3 we found men rated all aspects of module 2 (instruction and content) higher than the female participants did and men rated the instruction for module 3 higher. There were more significant differences in cohort 2, though the differences in scores were quite small .2 or so with men rating all aspects of module 1 and 4 higher and women rating the content in module 2 significantly higher than their male colleagues.
 - These positive changes occurred in the context of a doubling of the number of participants each year and did so in a way that suggest more similar positive growth across diverse members than in prior years.

As a result of this evaluation, we conclude that the program is further improving principals' knowledge and skills relative to instructional leadership and that, in many cases, these enhancements appear to be translating to improved leadership practices. Moreover, these positive changes seem to have occurred regardless of the principals' demographic background or the nature of the school in which they teach.

Introduction

This is the culminating report on the effectiveness of the of Advanced Instructional Leadership Professional Diploma (AILPD). In particular, it focuses on the experiences and performance of the second (Cohort 2) and third cohort of participants (Cohort 3) in the program and how these outcomes compare over time. While we use the data from Cohort 1 as a baseline, we focus less on the specifics of this pilot year of the program. We do highlight the tremendous and statistically significant enhancements made to the program and its outcomes since that time. We do not provide specifics for Cohort 1 for three reasons: (1) the composition of the Cohort was different (e.g., size, composition, demographics) than the current and future Cohorts, (2) the instructors were comprised only of UConn instructors rather than QRTA core team members, and (3) there have been substantive changes to the program content, approach, and assessments since the pilot year. In essence, Cohort 1 was a true pilot year and the results from our early evaluation served their function as a formative tool to enhance implementation. As we discuss throughout this document, the pilot efforts bore substantive fruit and the program continues to produce highly effective results on multiple parameters even as it expands in scope.

This report was prepared in collaboration by the Queen Rania Teacher Academy QRTA Monitoring & Evaluation team and the University of Connecticut (UConn) research team. As we discuss in greater detail below, we utilized a variety of different data tools (e.g., teachers surveys, participants post workshop feedback surveys, focus groups, principals pre/post assessments) for the purposes of the evaluation. The evaluation is framed around two interrelated areas of effectiveness: (1) changes to principal participants' instructional leadership capabilities and (2) QRTA's evolving capacity to successfully maintain program quality and grow in size over time (i.e., scale up). As the program is now fully implemented by QRTA, participants' evaluations of the modules (instruction and content) serve as the primary and sole instruments to evaluate QRTA core team members' ability to implement the program effectively.

Additionally, while the pilot year (Cohort 1) was perhaps the most different in terms of content and delivery, it is important to remember that the program delivery systems including both who delivered the instruction (i.e., from UConn faculty to QRTA core members) and what was delivered (i.e., content and order of learning) also shifted during Cohort 1 and 2 implementation. As a result, while the overall trends can be illustrative and helpful in determining next steps, strict comparisons (e.g., this year's performance on a given module was better than last year's) are not appropriate. This is because while we are comparing how well the elements of the program work on average, the specifics regarding the composition of those elements (e.g., what was taught) shifted. The narrower our focus the more likely we are comparing things that are not fully equivalent. Indeed, as true when looking at the data from Cohort 2, program content and instruction has continued to evolve as the UConn and QRTA team received information regarding these elements' effectiveness. In other words, while the program remains the same at its core, modifications and improvements were made so that implementation for Cohort 3 was not identical to that of previous years – this includes shifts to the modules and the instructors' approach as well as the composition of the participants. As we discuss in greater detail later, there were also some changes made to at least one of the instruments (i.e., the pre/post assessment) between Cohort 2

and 3 to support better alignment between the modified curriculum and the assessment. Positively, the program continues to produce a number of positive results across our measures and, in so doing, reveals strong and growing capacity of QRTA to sustain these results over time.

To measure positive changes to participants' instructional leadership, we collected information on whether and to what degree module learning translated into changes in participants' knowledge and skills. Information on changes to participants' knowledge was again captured with content-based pre-post assessments developed by UConn module instructors. The findings were quite positive with large improvements to participants' knowledge across almost all of the content areas. These results were again far better than the first year of implementation and on pace with the statistically significant gains made from the first year of implementation. Specifically, Cohort 3 members experienced, on average, an approximate 10 total point difference from pre to post assessment versus the around 8 points of Cohort 1. This is particularly impressive given that changes to the assessment and specifically the questions focused on participants' learning relative to the Curriculum Lab module were changed such that the possible point overall accrual was cut by 25 points. As such, with fewer possible points to earn, Cohort 3 still outperformed Cohort 1 in terms of the change and had a similar change in real points (though a similar percentage point change) than for Cohort 2. Also, as we discuss later, Cohort 3 members showed more consistent positive changes in their learning. This suggests the program has enhanced its ability to meet the needs of a larger and hence more diverse group of learners and again bodes well for the program's future success as it continues to expand.

Information on changes to participants' behaviors came from surveying select group of principals' teachers about changes to the principals' instructional leadership practices before and after program participation. Findings here were also quite positive reflecting similar growth from the first implementation and explicated fully in the findings section. We also asked participants, via the focus groups, about changes they perceived to have occurred in their practice as a result of the program to positive effect.

For the second area of effectiveness focused on building local QRTA capacity to engage in the work, the analysis reflects the goal from the beginning to build core members' capacity to independently and effectively implement the program. During the pilot year and, to some degree for Cohort 2, the goal was to deepen core team members' content knowledge and support their instructional ability in regards to the modules. In this third year, core members took full ownership of the program's instructional duties. Therefore, the measures used to assess core members' teaching effectiveness were participants' reports on their experiences in the modules themselves. In particular, participants gave information regarding the effectiveness of the module content and the instructional practices deployed by the QRTA core team members. Data sources for this portion of the evaluation came from feedback surveys and focus group interviews of participants. As we discuss in detail in the findings section, despite more than doubling the number of participants while simultaneously decreasing the amount of coaching and other external supports to QRTA instructors, satisfaction with all elements of module content and instruction were significantly higher than those of Cohort 1 and mirrored or often outshone Cohort 2's positive responses. In this way, the current scores suggest a robust program poised to continue producing positive results over time.

Results showed high levels of effectiveness and that the capacity building model deployed by UConn partners was successful. QRTA core members provide high quality learning experiences for participants and do so in ways that have met the associated challenges of serving a larger and more diverse cohort.

While we will provide greater detail on the focus of this evaluation, the measures used and analytic approach later in the report, here we give an overview of the findings in both arenas. Looking across these two issues we find:

- Issue 1: Positive Changes to Participants' Instructional Leadership
 - Cohort 3 participants show a statistically significant positive change of approximately 17% (10 points) in their content knowledge as measured by pre/post assessment. This is almost a 100% improvement from Cohort 1 in which the improvement was 11% (8 points) and is on par with Cohort 2 (14 points or 20% increase). There were no statistically significant differences by gender in any cohort (i.e., men and women principals experienced the same positive results each year). This is particularly impressive given changes to the assessment which decreased the total possible point accrual by 25 points.
 - Findings regarding teachers' views on participants' instructional leadership were overall quite positive. As true for Cohort 2, virtually all areas showed positive growth and none were both statistically significant and negative. Specifically, there were three areas of instructional leadership for which teachers reported the principals, on average, made significant positive growth. These areas included, 1) whether professional development is tailored to their needs, 2) whether the principal gives regular and helpful feedback on their teaching, and 3) whether the principal takes into account the feedback the teacher provides.
- Issue 2: QRTA capacity to sustain the program effectiveness
 - Participants rated both the module content and the instruction quite highly with an average satisfaction of 87% and 89.6% respectively. These results mirror those for Cohort 2 when participants rated both the module content and the instruction quite highly with an average satisfaction of 88.9% and 88.05% respectively. Moreover, they are far higher than for Cohort 1 when the scores were 85.3% and 88.8% respectively.
 - When disaggregating these scores by gender we find that there were sometimes small differences (always less than .4 points) between male and female participants' responses. Specifically, in Cohort 3 we found men rated all aspects of module 2 (instruction and content) higher than women and the instruction for module 3 higher. There were more significant differences in Cohort 2, though the differences in scores were quite small .2 or so with men rating all aspects of module 1 and 4 higher and women rating the content in module 2 significantly higher than their male colleagues.
 - These positive changes occurred in the context of a doubling of the number of participants and did so in a way that suggest more similar positive growth across diverse members than in prior years.

In the next sections of the report we provide more information regarding each of these findings.

As a reminder, even with some statistical evidence regarding the continued improvements in program effectiveness the findings are not causal in nature. While we can show trends, we cannot pinpoint the mechanism producing these trends or its directionality. Additionally, as true for the prior reports, all of the data here was culled solely from program participants. We neither have a control group nor did we randomly assign participation in the program. As such, the findings here are correlative, not causal, in nature. This does not mean they are not valuable. To the contrary, the data here provides rich insights into our questions of interest. The different sources of data here give us opportunities to triangulate our findings and generate robust understanding of program effectiveness.

Report Structure

To address the multiple issues in this evaluation, the report is split into two sections. The first section is dedicated to the evaluation the effectiveness of the program as related to participants knowledge and skill development (i.e., Issue 1). The second section then focuses on the core team members their effectiveness in teaching the modules. Within each section there is an overview with the specific research questions guiding our approach. Next will be a methods section that includes more information on the collection methods, tools and sample. We will then shift to findings and conclude with some thoughts about potential next steps.

Section I: Evaluating Program Impact on Participants' Knowledge and Skills

The focus of this part of the evaluation was to understand how and the degree to which participants (1) gained new knowledge and skills regarding instructional leaders and then (2) implemented these leadership skills and knowledge. Therefore, we needed to know whether the modules themselves were effective, if participants learned as a result and whether those skills transferred to practice. Specifically, this part of the evaluation was driven by the following research questions:

1. Is the content of the program useful, relevant and aligned participants' needs?*
2. Do participants view the instruction of that content effective?*
3. Do participants gain new knowledge and skills as a result of participating in these modules?
4. Can participants implement their new learning in effective and meaningful ways?

* The findings for these questions will be presented in Section II, as the data used to analyze them (i.e., participants' evaluations of the module content and instruction) was also the main information used to explore the degree to which core team members' knowledge and skills to be effective instructors was appropriately developed during the course of the partnership as well as the sustainability of the program over time.

Methods

As the research questions focused on both perceptual (what participants think) and behavioral (what participants do) changes, it was necessary for the research team to collect multiple forms of data to understand these phenomena. Specifically, we took a mixed method approach

gaining quantitative and qualitative information about the participants' views of the program, changes in their beliefs and knowledge as a result.

Sample

Here we provide an overview of some of the characteristics of all of the participants in the formalized program (Cohorts 2 & 3). The second cohort of participants (n=60) included far more female than male participants (Table 1). The largest representation came from coed schools (53% of the sample). The distribution of principals from male and female schools was more evenly represented (i.e., 27% from boys' schools and 20% from girls' schools). Finally, there were more principals coming from elementary settings (62%) than from secondary settings (38%).

Table 1: Cohort 2 Participant Demographics (N=60)

Characteristic		n	%
Participants			
Sex	Female	48	80.00%
	Male	12	20.00%
Age	30-35	11	18.33%
	36-40	16	26.67%
	41-45	22	36.67%
	46-50	11	18.33%
Undergraduate Major	BA Education	21	35.00%
	BA with education coursework	38	63.33%
	BA no education coursework	1	1.67%
Graduate Degree Status	None/N/A	19	31.67%
	MA or Higher	41	68.33%
Administrative degree	MA in sch. Admin	8	13.33%
	HS Diploma in sch. Admin	8	13.33%
	One-year course in sch. admin	10	16.67%
	Professional diploma in sch. Admin	11	18.33%
	Other	23	38.33%
Previous Position(s) held	Vice Principal	8	13.33%
	Leadership Team	8	13.33%
Yrs Principal in Current Sch.	1-5 years	52	86.67%
	6-10 years	7	11.67%
School Level	Elementary	37	61.67%
	Secondary	23	38.33%
School Type	Boys	16	26.67%
	Girls	12	20.00%
	Co-Ed	32	53.33%

How often do teachers meet?	Daily	4	6.67%
	Once a week	12	20.00%
	0-1 time a month	22	36.67%
	1-2 times a month	19	31.67%
	Never	2	3.33%
	I don't know	1	1.67%

Looking now at the third cohort of participants (n=127) (Table 2). We first see that there were more than double the number of participants in Cohort 2 (n=60). When we then consider this increase in combination with the positive findings throughout this report, it suggests a program with the necessary capacity to enhance principal instructional leadership for years to come. As true for both previous cohorts, there were far more female than male participants. Indeed, there were a number of ways in which the demographics of Cohort 3 mirrored those of Cohort 2. For example, again the largest representation of principals came from coed schools (41% of the sample). The distribution of principals from male and female schools was again more evenly represented (i.e., 31% from boys' schools and 28% from girls' schools). Finally, there were more principals coming from elementary settings (53%) than from secondary settings (47%).

Table 2: Cohort 3 Participant Demographics (N=127)

Characteristic		n	%
Participants			
Sex	Female	86	67%
	Male	42	33%
Age	30-35	8	6%
	36-40	43	34%
	41-45	46	36%
	46-50	27	22%
	50+	3	2%
Undergraduate Major	BA Education	49	39%
	BA with education coursework	70	55%
	BA no education coursework	8	6%
Graduate Degree Status	None/N/A	2	2%
	MA or Higher	125	98%
Administrative degree	MA in sch. Admin	15	12%
	HS Diploma in sch. admin	26	20%
	One-year course in sch. admin	8	6%
	Professional diploma in sch. Admin	78	62%
	Other	0	0%
Previous Position(s) held	Vice Principal	84	66%
	Leadership Team	9	7%
Yrs Principal in Current Sch.	1-5 years	101	79.5%

	6-10 years	16	12.5%
	11+	10	8%
School Level	Elementary	68	53%
	Secondary	49	47%
School Type	Boys	40	31%
	Girls	36	28%
	Co-Ed	51	41%
How often do teachers meet?	Daily	4	3%
	Once a week	52	41%
	0-1 time a month	31	25%
	1-2 times a month	38	29%
	Never	2	2%
	I don't know	0	0%

Overall, the demographics suggest Cohort 3 was, on average, also similar to Cohort 2 in terms of participants' experiences in practice. Participants from Cohort 3 had educational backgrounds similar to those of Cohort 2. First, there were few participants, in this case only one cohort member, who did not take some form of undergraduate coursework in education. As true with all prior participants, most members of Cohort 3 (94%) took some kind of undergraduate courses in education and around the same portion of the group (39%, n=49) completed an education major. Slightly more of Cohort 3's participants (74%, n=93) had some form of advanced degree. Finally, in this cohort, a larger proportion of the participants (n=93 or about 73%), an almost fourfold increase, held another school leadership position prior to their current work. In contrast, in Cohort 1, 50% and in Cohort 2, 26% of the participants held such positions.

Looking at their current roles, the distribution between early career principals (i.e., those with 5 or fewer years of experience) and those in the career for longer periods of time (i.e., 6-10 years), like Cohort 2, was also more skewed than for Cohort 1, with approximately 80% of the participants working in their school for 5 or fewer years. This is contrast to Cohort 1 where the split between these two categories was about even. However, the average age of the participants was similar between the cohorts with more than half of the participants in Cohort 3 being over 41 years old.

Finally, participants' reports suggest, as of their final participation in the modules, most provided their teachers relatively frequent opportunities to meet with each other to plan. In response to the prompt, "In the past 6 months, how many times did the subject teachers in your school met to discuss teaching strategies and/or plan lessons together?" approximately 44% of participants said their teachers meet daily or once a week. When looking at how many said their teachers met more than once a month, we find 25% said this occurred. Another 29% said teachers meet only 1-2 times a month. Only 2 participants in Cohort 3 (the same number as Cohort 2) said their teachers never meet together.

Here it is worth noting that while all participants were surveyed regarding their views of the

effectiveness of each of the modules and all took the pre/post content assessment, smaller groups of participants attended the focus groups. For these two data sources we worked to identify a representative sample of the participants. We chose this approach of collecting data from a select group of participants to triangulate our data and support reliable and valid generalization. Additionally, these methods are highly resource dependent (e.g., time, money, etc.) and therefore were used somewhat more sparingly.

Data Collection and Analysis

In the following we present a description of the data collection materials we collected related to the first issue of the evaluation on participants' knowledge and skills as well as a short overview of our analytical approach. All materials are available in Appendix A.

Feedback Surveys of Module Effectiveness

At the conclusion of each module, participants were sent a survey via email asking them to evaluate its instruction and content and register their overall satisfaction with the program. All questions were rated using a Likert scale. At the conclusion of the survey, participants were given space to write additional comments regarding the module.

Results were analyzed descriptively to help better understand potential differences regarding participants' learning and growth and the degree to which the current modules and teaching was experienced positively and aligned with participants' perceived needs and learning goals. While we discuss differences in satisfaction levels between participants in Cohort 3 and 2 on all measures, we do not conduct statistical comparisons between them to assess whether the ratings were statistically significantly higher year to year. We based this decision on the fact that though the modules are the same year to year, the participants filling out the surveys were not. In this way, there might be issues related to the nature of the group that would be responsible for the differences more so than the content or instruction driving these differences. We do highlight when the numbers are higher (they are throughout) as a way to get at overall trends, in this case enhancements to module performance.

Pre/Post Content Principals' Assessment

The first data collection tool was a pre/post content assessment in the form of a survey. In addition to some demographic information, the survey was broken into sections aligned with each of the modules. The questions on the assessment came from the UConn instructors and aimed to provide insights into key learning points and underlying competencies the module was meant to convey. Prior to the start of the modules for Cohort 3, the research team asked the instructors whether revisions to the assessment were needed to help ensure alignment between the content of the survey and curricular enhancements being made. In response, the assessment for the Curriculum Lab module was substantively changed. All other items however remained the same from Cohort 2 and much higher than for Cohort 1.

In its revised state, the assessment continued to include open response and multiple-choice

questions to assess understanding of course learnings related to instructional leadership. As true with prior deployments of the survey, participants were also asked about their demographic and educational background information and how frequently teachers in their schools were given time to discuss teaching. Participants received and responded to the survey electronically through a secure server. Additionally, all participants responded at the same time at QRTA premises and under the supervision of QRTA measure and evaluation members – though their responses were not monitored in terms of their substance.

To analyze the data, a scoring document was created by the research team in collaboration with the module instructors. The QRTA and UConn team engaged in the coding of the assessments and worked together to ensure intercoder reliability. Pre/post assessments were analyzed using simple statistical methods (e.g., correlation, t-tests, ANOVA), to see whether and to what degree positive differences occurred from the first to second administration. Additionally, we ran statistical comparisons regarding whether and to what degree the growth from the pre to post assessment was different between Cohort 1 and 2 Cohort 2 and 3. We also compared whether and to what degree there were gender-based differences in performance within a given cohort. We do not compare Cohort 1 and 3 as the differences between the modules, the instructors and the participants made such comparisons nonsensical. We opt to focus on the difference (i.e., growth in performance) rather than simple comparisons in participant scores' as the Cohorts 2 and 3 the content differed across these experiences and so participants' response to a particular item may not be a valid or reliable measure on which to base true comparisons.

Teacher Survey of Principals' Instructional Leadership

We administered a survey of principal leadership to teachers in 22 schools, selected to help produce a sample representative of Jordanian schools. The goal was to show that school type, demographics or other environmental factors would not affect whether and to what degree principals would be able to effectively transfer their learning from the program to their schools. Again, we used the school size, level of students served (e.g., elementary, middle, etc.), school location, and the gender of the students to determine an appropriate sample. We used the same survey with the teachers each year. Over the three years of the study, we collected information from a total of 50 schools with 1,207 teachers responding.

This survey was based on surveys designed and validated by the Connecticut State Department of Education. It was translated into Arabic (as were all collection tools used with participants) and administered in paper form with data collectors overseeing the process. The purpose of this survey was to estimate school principals' gains in instructional leadership, as measured through teachers' perceptions, during their participation in the program. In addition to items about principal instructional leadership, the survey asked for teachers' demographic and educational background information and the frequency of their meetings with other teachers to discuss teaching. Descriptive statistics and t-tests were used to assess changes in teachers' perceptions of participants' behaviors over time.

For Cohort 2 the survey was administered in 20 schools. As shown in Table 3 sample participants were predominantly female, between 31 and 35 years old (29%), had a B.A. in education

(41%), taught in their current school for 1-5 years (46%), and work in secondary schools (54%) and girls' schools (52%). This sample was relatively similar to the sample for Cohort 1 in terms of sample size (approximately 367) and respondents' sex, age, education, and experiences level. However; the samples differed regarding school type. The majority of teachers work in secondary and girls' schools; for Cohort 1, the majority of teachers worked in elementary schools and about 37% worked in girls' and the same proportion was in co-ed schools. In addition, 20 schools participated in this year's administration while only eight participated in last year's surveys. While these differences in school type are not to be overlooked, the larger number of schools in this year's sample should also be considered in terms of the validity and potential generalizability of the findings.

Table 3: Cohort 2 Participant Demographics (Teachers), post-test (n=367)

Characteristic	%
Sex	
Female	82%
Male	18%
Age	
20-25	4%
26-30	8%
31-35	29%
36-40	28%
41-45	20%
46-50	8%
over 50	3%
Undergraduate major	
BA in education	41%
BA in other subject with coursework in education	30%
BA in other fields	27%
Diploma in education	2%
Diploma in other field with coursework in education	0%
Years in teaching at current school	
1-5 years	46%
6-10 years	25%
10+ years	29%
School Level	
Elementary	46%
Secondary	54%
School type	
Boys	20%
Girls	52%
co-ed	28%

As illustrated in Table 4, the sample for Cohort 3 was both similar departed in some ways from previous rounds. In terms of similarities, as true with the prior cohorts, the sample was predominantly female (67%) with most of the participants being between the ages of 31-35 (30%).

The proportion of teachers with a BA in education (37%) was also on par with the prior cohorts as was their degree of teaching experience (49% with 1-5 years).

In terms of differences, first, it was a larger sample. In this round, the teacher survey was administered in 22 schools and 472 teachers completed the post-test compared to 367 for Cohort 2. Second, the composition of the schools in which they worked varied in that almost three-quarters (67%) worked in secondary schools – in the prior sample approximately 54% worked in secondary schools. Last, teachers were more likely than last round to work at boys' schools (33% vs. 20%). The more comprehensive nature of the sample across the three cohorts, and as we discuss next, the positive effects we found across the 42 schools comprising the C2 and C3 samples¹, suggests the program is effective in support participant principals in enhancing their instructional leadership across a broad swath of school types.

Table 4: Cohort 3 Participant Demographics (Teachers), post-test (n=472)

Characteristic	%
Sex	
Female	67%
Male	33%
Age	
20-25	9%
26-30	17%
31-35	30%
36-40	22%
41-45	14%
46-50	5%
over 50	2%
Undergraduate major	
BA in education	37%
BA in other subject with coursework in education	34%
BA in other fields	21%
Diploma in education	3%
Diploma in other field with coursework in education	4%
Years in teaching at current school	
1-5 years	49%
6-10 years	24%
10+ years	26%
School Level	
Elementary	27%
Secondary	73%
School type	
Boys	33%

¹ The teacher survey results from the pilot year (Cohort 1) were far less positive and, in some cases negative. This information was used to enhance the program. The results here seem to indicate these changes were successful and continue to produce positive results in terms of participants' implementation of their new programmatic learning.

Girls	42%
co-ed	22%

Principals' Focus Groups

At the conclusion of each year of the program, the QRTA research team conducted focus groups with a random sample of participants (n=81 participants over the three years, Cohort 1 n=21, Cohort 2 n=20, Cohort 3 n=40). Questions within these focus groups were aimed at assessing participants' ability to link to module content with leadership standards and school needs, to gather feedback on the quality of instruction and alignment between module offerings and current needs and to learn about the ways in which they were applying their new learning in the context of their schools. Data from these focus groups were analyzed using thematic analysis.

Findings

As a reminder our work regarding the impact of the program on participants was guided by the following research questions:

1. Is the content of the program useful, relevant and aligned participants' needs?
2. Do participants view the instruction of that content effective?
3. Do participants gain new knowledge and skills as a result of participating in program modules?
4. Can participants implement their new learning in effective and meaningful ways?

In the following, we answer questions 3 & 4 utilizing our analysis of the aforementioned data. Findings for questions 1 & 2 will be presented in Section II, as the data used to analyze these questions (i.e., participants' evaluations of the module content and instruction) was also the main information analyzed to explore the degree to which core team members' knowledge and skills to be effective instructors was appropriately developed during the course of the partnership.

Changes to Participants' Knowledge and Skills

To assess the degree to which the program impacted participants' knowledge and skills, as already stated, we utilized a pre-post survey focused on program content, the focus group transcripts as highlighted above, and teacher surveys of their principals' instructional practices. As a reminder, the teacher surveys came from a select group of participants' schools in which teachers were surveyed at the beginning and end of the program to see whether they felt the principal had changed his/her instructional leadership practice over the course of the program. We present the findings for each element below.

Pre/Post Content Principals' Assessment

In this section, we present the data from Cohort 2 and then Cohort 3 drawing focus to changes over time between them and from the pilot year. As we discuss in more detail later, there have been substantive gains since the pilot year and these gains have been sustained over time.

Cohort 3 members showed equal or better growth in their knowledge gains than Cohort 2 and Cohort 2’s results were far more positive than those in the pilot year. Again, given the larger size of each subsequent cohort and QRTA’s growing independent implementation of the program (i.e., for Cohort 3 all the instructors were QRTA core members and UConn provided coaching was suspended) these results should be viewed quite positively indeed.

In terms of participants’ new knowledge as a result of the program, the pre-post surveys for Cohort 2 showed, on average, a positive, statistically significant improvement of over 20 percent (or an approximately 14-point improvement from the pre to the post test). There were no statistically significant differences between female and male participants’ improvements. The 14-point improvement was almost twice the percent improvement that occurred for the first cohort, suggesting the modifications to the program were indeed effective. It is also worth mentioning Cohort 2 had a lower average baseline score (71.50) relative to Cohort 1 (74.58 baseline score). These lower starting scores suggests that whether individuals, on average, gain knowledge in the program is not contingent on the degree of knowledge of instructional leadership they enter the program with. Indeed, looking across the cohort, we find that, of the 60 participants, only 3 made no growth on their overall score. The average growth per participant was about 20% or about 14 points and reflects the highest growth rate for Cohort 1 participants. Additionally, 12 individuals increased their scores by more than 27% or 20 points a result in sharp contrast to Cohort 1 participants’ where again 20% growth was the greatest amount across all members.

Additionally, as shown in Table 5, for every section of the assessment, enhancements to participants’ knowledge of the module content existed and were statistically significant. As a reminder, each of the sections represents the focal elements of a given module with aligned questions created by the module instructors. Overall, there were only 2 questions for which we did not see a statistically significant positive improvement in scores. One of these was the only negative result associated with the survey (but not statistically significant). As we discuss later, the results for this question were similar last time and suggest, given the two years of data, the problem may be a function of misalignment between content and the assessment rather than a failure of participants to learn the material. On this note, as the content has been modified so too will the pre/post assessment for Cohort 3 making it more aligned with current offerings.

Table 5: Cohort 2 Pre-Post Content Survey Means Differences

Question	Potential Score	Mean Pre	Mean Post	Difference	Difference vs. Cohort 1
<i>School Culture (M2 in C2, M1 in C1)</i>					
Key attributes of a positive school culture?	7	1.56	2.65	1.09***	
Methods to evaluate a school’s culture?	6	1.72	2.22	0.5**	
Agreement with each statement	65	41.86	43.43	1.57*	
Name 1 or 2 effective strategies to enhance school culture	2	0.15	0.93	0.78***	
Name 1 or 2 effective strategies to enhance	2	0.63	1.28	0.65***	

parent and/or community engagement					
Total Section	82	45.93	50.5	4.57***	1.18
<i>Curriculum Lab (M3 in C2, M3 in C1)</i>					
What is a curriculum?	1	0.22	0.58	0.36***	
How much do you agree with each statement?	30	20.58	21.55	0.97*	
What are some tools to assess whether a curriculum is effective?	5	0.93	1.36	0.43**	
Total Section	36	21.73	23.5	1.77**	-.93
<i>Supervision and Evaluation (M4 in C2, M4 in C1)</i>					
What information should guide school wide goals?	6	0.42	1.8	1.38***	
Ways teacher's professional goals should be established?	6	0.4	1.6	1.2***	
Essential elements of effective professional development for teachers?	2	0.82	0.73	-0.09	
Key elements of an effective teacher evaluation plan?	6	0.36	2.32	1.96***	
Total Section	20	2	6.45	4.45***	3.00***
<i>Organizational Effectiveness (M1 in C2, M2 in C1)</i>					
What metaphor best describes the organizational structure of a school?	1	0.42	0.53	0.11	
Describe your plan to improve teaching and learning in your school.	5	0.75	2.46	1.71***	
What can be used to evaluate your success as a leader?	4	0.67	1.98	1.31***	
Total Section	10	1.83	4.98	3.15***	2.21***
Cumulative	148	71.5	85.43	13.93***	5.66**

Note: * p < .05, ** p < .001, *** p < .0001

In looking more deeply at the survey results, we notice, Cohort 2 participant scores remain somewhat low relative to the total possible point accrual (i.e., there was an approximate 63-point gap between the average score on the post assessment and the total possible score). Again, there is no reason to believe participants should come to the program with higher levels of content knowledge. Therefore, while the scores suggest more there is more to learn in terms of the program content participants also clearly made great strides in their knowledge over the course of the modules.

Moreover, we note this cohort showed greater growth than in Cohort 1. Indeed, the final column in the table serves to highlight whether the change in participant knowledge was significantly larger for Cohort 2 than Cohort 1 for each section and on the assessment overall. Cohort 2 made larger gains in every section, on average, other than for the module focused on Curriculum Lab which was also the lowest rated for this cohort. It is worth noting this difference was not

significant and hence may be attributable to simple error. In contrast, we can see that for the modules dedicated to Organizational Effectiveness and Supervision and Evaluation as well as for the overall score, Cohort 2's participants showed greater growth in their learning than their colleagues in Cohort 1. This suggests both that participants learned quite a lot in the program as aligned with its goals and the program was improving over time.

Going through each section of the survey, we find for the questions dedicated to the module on issues of school culture; all showed a statistically significant positive change. This is in contrast with Cohort 1 in which one of the items did not show a positive increase. Additionally, while not statistically significant, the average growth in this section was larger for Cohort 2 than Cohort 1. Specifically, the enhanced responses led to the mean total score for this section being 4.57 points higher for the second administration and 1.17 points higher than for Cohort 1. In particular, Cohort 1 participants' responses reflected greater knowledge in strategies to diagnose and then address issues of school culture and parent and community engagement.

For the second section dedicated to the module content on curriculum, we again see a positive, statistically significant difference for all of the questions and for the entire section. In this case about 1.77 points on average in total – a statistically significant difference. It is worth noting that although the overall difference in scores was not larger for Cohort 2 than for Cohort 1, Cohort 2 showed consistent growth in this section.

The third section focused on supervision and evaluation and again, showed a statistically significant improvement in the overall mean score (4.45 points), an approximately 3 point bigger difference than that for Cohort 1. As already mentioned above and true for Cohort 1, all but one of the questions showed improvements. This question focused on the effective elements of professional development for teachers, was the only question for which there was a slight decrease (not significant) in the score. Again, this may be a function of the question rather than the content as this question was an issue with the prior Cohort as well.

Finally, for the section on organizational effectiveness, in sharp contrast to Cohort 1's results in which only one question within the section had a statistically significant result, Cohort 2 participants showed significant gains in their scores for all items other than one. The one that did not have a statistically significant result was still positive. The overall growth for this section was over 3 points and was an approximately 2 point larger positive difference in average scores as that acquired by members of Cohort 1. In this way, Cohort 2 far exceeded Cohort 1's performance on this section of the assessment.

Shifting now to Cohort 3 participants' new knowledge as a result of the program (Table 6), the pre-post surveys shows, on average, a positive, statistically significant improvement of approximately 17 percent (or an approximately 10 point improvement from the pre to the post test). While the growth is slightly smaller than for Cohort 2 (worth noting the larger number of participants in Cohort 2 as well), it remains far larger than for Cohort 1. Again, there were no statistically significant differences in performance by participant gender. Given efforts to make the curriculum more rigorous and the assessment better aligned while correspondingly decreasing the amount of UConn-based support to the team, the continued growth in participant knowledge and skills is to be

lauded. This is particularly true given, as we discuss later, the apparent positive transference of these skills to principals' instructional leadership practice in their schools as reported by their teachers.

Table 6: Cohort 3 Pre-Post Content Survey Means Differences (n=127)

Question	Potential Score	Mean Pre	Mean Post	Difference	Difference vs. Cohort 2
<i>School Culture (M2 in C2 & C3, M1 in C1)</i>					
Key attributes of a positive school culture?	7	1.63	2.57	0.94***	
Methods to evaluate a school's culture?	6	1.71	2.31	0.60***	
Agreement with each statement	65	41.32	43.19	1.87***	
Name 1 or 2 effective strategies to enhance school culture	2	0.52	0.68	0.16*	
Name 1 or 2 effective strategies to enhance parent and/or community engagement	2	1.04	1.31	0.27***	
Total Section	82	46.22	50.03	3.81***	0.71
<i>Curriculum Lab (new questions for C3)</i>					
What is the relationship among curriculum, instruction, and assessment?	3	0.23	0.81	0.58***	
How do curriculum, instruction, and assessment support learning goals?	3	0.71	1.50	0.79***	
What is the difference between assessment of learning and assessment for learning?	3	0.99	2.19	1.20***	
Total Section	9	1.93	4.50	2.57***	
<i>Supervision and Evaluation (M4 in C2 & C3, M4 in C1)</i>					
What information should guide school wide goals?	6	1.48	1.80	0.31***	
Ways teacher's professional goals should be established?	6	1.28	2.24	0.97***	
Essential elements of effective professional development for teachers?	2	0.64	0.55	-0.09	
Key elements of an effective teacher evaluation plan?	6	1.31	2.11	0.80***	
Total Section	20	4.70	6.70	2.00***	0.29
<i>Organizational Effectiveness (M1 in C2 & C3, M2 in C1)</i>					
What metaphor best describes the organizational structure of a school?	1	0.10	0.27	0.17***	
Describe your plan to improve teaching and learning in your school.	5	1.94	2.48	0.54***	
What can be used to evaluate your success as a leader?	4	1.81	2.31	0.50***	
Total Section	10	3.85	5.06	1.21***	-3.26***

Cumulative	121	56.70	66.28	9.58***
------------	-----	-------	-------	---------

Note: * $p < .05$, ** $p < .001$, *** $p < .0001$

Cohort 3 had a lower average baseline score (57) relative to either of the prior cohorts (both had scores in the 70s). However, because of the shift in the Curriculum Lab assessment the possible score accrual in that section went from a potential 36 points to 9 points. As such, we might actually consider this cohort to have the highest proportional scores relative to the total potential points. Indeed, looking across the cohort, we find that, of the 127 participants (more than twice the number as in Cohort 2), only 10% made no growth on their overall score (5% failed to make growth in Cohort 2). The average growth per participant was about 17% or about 10 points and was on par for the Cohort 2 participants. Additionally, 14 individuals increased their scores by more than 35% or 20 points, a number higher than Cohort 2 and still a result in sharp contrast to Cohort 1 participants' where few members achieved this level of growth.

As true across all the cohorts, were no correlations between participants' demographic characteristics and performance on either the pre or post survey. This means any differences in participants' scores relative to their backgrounds, gender, educational training, etc. can simply be attributed to error. There were no systematic differences in scores by any of these factors. In essence, these findings suggest that background factors do not play a role in how much learning we might anticipate for any given participant in the program, on average. This finding as well as the larger and slightly different composition of Cohort 3 relative to Cohort 1 and 2 again serves as confirmatory evidence that the program seems to work equally well for participants of various backgrounds coming from different types of schools in different parts of the country.

Additionally, as shown in Table 6, for every section of the assessment, enhancements to participants' knowledge of the module content existed and were statistically significant. As a reminder, each of the sections represents the focal elements of a given module with aligned questions created by the module instructors. Overall, there were only 2 questions for which we did not see a statistically significant positive improvement in scores. One of these was the only negative result associated with the survey (but not statistically significant). As we discuss later, the results for this question were similar last time and suggest, given the two years of data, the problem may be a function of misalignment between content and the assessment rather than a failure of participants to learn the material. On this note, as the content has been modified so too will the pre/post assessment for Cohort 3 making it more aligned with current offerings.

In looking more deeply at the survey results, we notice, similar to the prior cohorts, participant scores remain somewhat low relative to the total possible point accrual (i.e., there was an approximate 82-point gap between the average score on the post assessment and the total possible score). However, while the scores suggest more there is more to learn in terms of the program content participants also clearly made great strides in their knowledge over the course of the modules. Moreover, and as indicated in other parts of this report, program officials have already recognized a need for additional opportunities within and across the modules for participants to more deeply practice applying these skills. As a result, the team is currently in the process of revamping both the curriculum and associated assessments to build greater depth, better alignment and the opportunity to – through a “change project” apply the learning in real time and supported

by instructors. Doing so will undoubtedly support even stronger learning among participants and transference of these skills to Jordan's schools.

Even with this gap, however, it is worth noting this cohort showed greater growth in terms of real scores than did Cohort 2. Additionally, with the exception of Organizational Effectiveness we also find that Cohort 3 maintained the levels of relative growth seen between Cohort 1 and 2. In other words, the improvements made to the program after the first round of implementation appear to have held with the positive changes in cohort members' knowledge growing to the same statistical degree this year as last. The final column in the table serves to highlight whether the change in participant knowledge was significantly larger for Cohort 3 than Cohort 2 for each section of the assessment. We were not able, due to the new questions and rating system for the Curriculum Lab, to offer such comparisons nor could we give an overall comparison as the total score is simply the sum of the other parts. Given that, besides Organizational Effectiveness, there were no significant differences in growth, we can conclude the positive gains persisted over time (i.e., both Cohort 2 and 3 showed significantly better growth than Cohort 1).

Going through each section of the survey, as true for Cohort 2, we find, for Cohort 3, responses to questions focused on the school culture module all showed a statistically significant positive change. Additionally, the average growth in this section was larger for Cohort 3 than Cohort 2 (or 1). Specifically, the enhanced responses led to the mean change in participants' pre/post scores for this section to be 1.18 points higher than it was for Cohort 2. Thus, it seems clear that the program is successfully moving participants to be more knowledgeable about how to create a positive culture at their respective schools.

For the second section dedicated to the module content on curriculum, remember there were changes to the assessment questions across time. As such, we cannot draw comparisons between cohorts. That said, we do see a positive, statistically significant difference for all of the questions and for the entire section. In this case about 2.57 points on average in total – a statistically significant difference. In this way, it seems that the new assessment is appropriately aligned to participants' learning and that program instructors should feel confident they are building substantive knowledge regarding curriculum, instruction and assessments and the relationship among the three.

The third section focused on supervision and evaluation and again, showed a statistically significant improvement in the overall mean score (2.00 points). Again, though the average difference in pre/post test scores for Cohort 2 and 3 is not significant, this finding can be understood as positive in that the gains from the first to second year of implementation were maintained. As already mentioned above and true for Cohort 1 and 2, all but one of the questions showed improvements. This question focused on the effective elements of professional development for teachers, was the only question for which there was a slight decrease (not significant) in the score. Again, and given the consistent response pattern in which all of the other items show growth in this section, we argue this neutral result is likely due to the question and potential misalignment to skills taught than participants' true knowledge and skills.

Finally, for the section on organizational effectiveness, Cohort 3 outperformed prior cohorts

in terms of producing meaningful gains (statistically significant) on all the items in this section. That said, the overall growth for this section was lower than for Cohort 2 and seems to be primarily attributable to smaller (though still statistically significant) point differentials for items 2 and 3. Additionally, we note that there was less variability in scores for Cohort 3. As a result, we can infer that most of the cohort showed growth on these items. This is in contrast to Cohort 2 members for which the variance in responses was far larger thus suggesting that the change in average score might have been more a function of a smaller group of participants growing quite a bit in their scores with others showing far less or no growth on the same items. Given this and Cohort 3's substantively larger size, these findings should again serve to bolster our conclusion that the program continues to make substantive enhancements to participants' knowledge and skills.

Taken together then we *can conclude that participants did indeed enhance their knowledge of instructional leadership elements as presented in the modules. Moreover, this learning seemed to have occurred across the modules and generally in ways that maintained the gains Cohort 2 members made relative to those in Cohort 1. In this way, it seems clear program staff has the necessary capacity to ensure participants' continued growth and development over time.*

Teacher Survey on Principals' Instructional Leadership

We move now from changes to participants' knowledge to whether those shifts stimulated changes in their practice and specifically - their instructional leadership. To do so, we present findings from teacher survey and focus groups, beginning with the teacher surveys.

As presented in Appendix A, the 25 items on the teacher survey examined their perceptions of school leadership. Questions ranged from global assessments of conditions associated with leadership (e.g. "There is a clear academic vision for this school") to items that asked teachers to comment on their principals' activities (e.g. "The principal at this school lets staff know what is expected of them").

Before providing the results of the teacher surveys, it is worth noting the larger positive trajectory of the results. In the first, pilot year of implementation, survey responses suggested that while teachers felt, on average, their principals exhibited relatively high levels of instructional leadership both prior to and after participating in the Advanced Leadership Program, it was unclear whether and to what degree the program enhanced principals' work. Indeed, we found no positive effects the first year of implementation. Since that time however, and with numerous revisions to the program to improve its relevancy and effectiveness, we find teachers reporting positive increases across all of the principals' behaviors after their participation, with some of these differences being statistically significant. Moreover, this has occurred with no noticeable decreases in the overall scores. If anything, each year teachers seem to provide higher baseline scores making our continued ability to find significant positive results even more remarkable.

For Cohort 2, teachers gave high ratings for leadership in their buildings on both the pre- and post- surveys. However, means still increased for all but two items from the time principals began the program to after its conclusion. In addition, there were significant increases in the extent to which teachers agreed on seven items, including 1) their beliefs their principals were making

adjustments to improve the school, 2) designing professional development to respond to students' needs, 3) including teachers in decision-making, 4) encouraging collaboration, and 5) taking their feedback. There was also growth in teachers' reports that 6) expectations are high and the 7) school is caring and nurturing. Unlike last year's survey results, there were no significant decreases from the pre-survey to the post-survey. These findings are very encouraging as they serve to suggest principals were able to effectively translate their module learning into practice.

On the surveys at the beginning and end of their principal's participation in the modules, teachers tended to report they agreed or strongly agreed with the presence of positive leadership behaviors and outcomes in their school. On the post-survey, all items had means of greater than 3.67 on a scale from 1 to 5 in which 1=don't know, 2=strongly disagree, 3=disagree, 4= agree, and 5=strongly agree (See Table 6). On the initial survey, 16 items had means over 4.0 (agree) and on the post-survey 20 items were above this number. The statement "I feel my contributions at this school are important" (item 5) had the highest mean on the initial survey (4.34) – though it was higher on the post-module survey. On the post survey, teachers reported the highest mean levels of agreement on items measured the degree to which they felt that their contributions to the school were important (4.36) and they felt respected by their principal (4.3), and that the principal worked to make the vision of the school a reality (4.3).

The lowest means on the post-module survey corresponded with items that measured the degree to which teachers felt that the school is a caring and nurturing place (item 11) with a score of 3.83 and that parents take responsibility for student achievement (item 14) with a score of 3.67. However, despite these being the lowest among the items, they are still relatively high on the overarching scale and there was still improvement from the initial survey as both means are significantly higher than the corresponding pre-test means.

Table 7: Cohort 2 Initial and Post Program Means on Teacher Survey (n=364)

Statement	Initial	Post	Difference
There is a clear academic vision for this school	4.13	4.23	0.1
The principal at this school lets staff know what is expected of them	4.21	4.24	0.03
I understand the vision of my school	4.22	4.30	0.08
The principal at this school is an inspiring leader	4.06	4.17	0.11
I feel my contributions at this school are important	4.34	4.36	0.02
The principal makes adjustments when things aren't working at this school	4.02	4.19	0.17*
The principal work to make the vision of this school a reality	4.25	4.30	0.05
My professional development is tailored to my students' needs	3.69	4.07	0.38**
Expectations are high at this school	3.49	3.85	0.36**
Teachers take responsibility for student achievement at this school	3.91	3.93	0.02
The school is a caring and nurturing place	3.64	3.83	0.19
I like working at this school	4.15	4.23	0.08

The principal at this school has confidence in the expertise of the teachers	4.08	4.18	0.1
Parents take responsibility for student achievement at this school	3.29	3.67	0.38**
I have an opportunity to participate in leadership roles at this school	3.87	3.98	0.11
The principal at this school is open to constructive feedback	3.98	4.08	0.1
The principal at this school invites teachers to play a meaningful role in setting goals for the school	3.93	4.22	0.29**
The principal at this school give regular and helpful feedback about my teaching	4.03	4.12	0.09
The principal at this school encourages collaboration among teachers to increase student learning	4.15	4.29	0.14*
The principal takes into account the feedback I give	3.98	4.12	0.14~
The principal inspires optimism for the future of this school	4.06	4.14	0.08
The principal encourages my career development	4.18	4.23	0.05
I feel respected by the school principal	4.36	4.30	-0.06
I feel supported by the school principal	4.15	4.14	-0.01
The principal deals with daily tasks and daily problems in an effective and efficient manner	4.12	4.17	0.05

Note: ~p<.01, * p < .05, ** p < .001, *** p < .0001

Shifting now to teacher's ratings of Cohort 3 principals, mean responses on both the pre- and post-test were high. In this case, the scores were even a bit higher than for the prior group, 20 items had means over 4.0 on the pre-test on a scale from 1 to 5 in which 1=don't know, 2=strongly disagree, 3=disagree, 4= agree, and 5=strongly agree (See Table 8). 21 items had a score of 4 or higher on the post-test. This means teachers in these schools generally agree or strongly agree with most items on the survey. Additionally, and as already mentioned and true with the prior cohort, these high initial numbers did not preclude positive changes, means on the post-test were higher than those on the pre-test for all but two items. Additionally, and in keeping with our prior positive results for cohort 2, we found significant gains between the pre-test and post-test on four of the 25 items. We found no statistically significant negative results. As we discuss next, there was mixed overlap between last year's and this year's surveys in terms of which of the improved items resulted in statistically significant changes. However, again, there were no declines in performance suggesting some of the reason for changes in significance may be due to fluctuations in averages rather than true differences in teachers' positive views.

It is interesting to note, and as shown in Table 8, the items receiving the highest means on the pre-test also registered the highest means on the post-test. Specifically, teachers registered the highest level of agreement with "The principal works to make the vision of this school a reality" (4.53 pretest, 4.57 post-test) and "I feel respected by my principal" (4.60 pretest, 4.63 post-test). "I feel supported by my principal" also received high mean ratings. Thus, teachers appear to feel respected and supported by their school leader and recognize that he or she is working to implement a vision

in their school. These scores are also higher (pre and post) than for those of the prior cohort. Again, given the high baseline scores, the fact we were able to detect positive results suggests that the change was substantive and a large number of teachers believed this improvement to have occurred.

In terms of those items with lower scores, we find consistency regarding those which previously identified as low, though, with one exception, the means were still higher for cohort 3 than for the prior cohorts. Specifically, the same items received the lowest mean scores on the pre- and post-assessments as with the prior cohort. Teachers' responses hovered between agree (4.0) and disagree (3.0) on the following items:

- 9. "Expectations are high at this school"
- 10. "Teachers take responsibility for student achievement at this school"
- 11. "The school is a caring and nurturing place"
- 14. "Parents take responsibility for student achievement at this school"

These items address the extent to which adults (teachers and parents) hold high expectations for students and take responsibility for helping them achieve (9, 10, 14) and the degree to which the school is caring (11). However, while these items had low initial means, all had positive changes, with one item's mean (9) being statistically significant. Thus, according to teachers, the areas in which there was most room for improvement were also those in which growth occurred.

In addition to item 9, we found significant gains between the pre-test and post-test on three of the other 25 items. Specifically, teachers' responses significantly improved on items asking whether professional development is tailored to their needs (8) and on two items related to feedback: one on whether the principal gives regular and helpful feedback on their teaching (18) and a second, at the marginal level, on whether the principal takes into account the feedback the teacher provides (20). These findings suggest principals were attempting to open channels of communication between themselves and teachers regarding how to improve instruction and leadership.

As with data collection for Cohort 2, for Cohort 3 we found significant differences among schools on the survey. There was particularly large variation among schools on items 4 (The principal at this school is an inspiring leader), item 10 (Teachers take responsibility for student achievement at this school), item 11 (The school is a caring and nurturing place) and item 24 (I feel supported by the school principal).

In sum, the survey results show that teachers view leaders' work in their schools very positively. Teachers' mean ratings of principals' leadership increased on all but two items from pre- to post-test, with significant increases on four items. The two decreases were not significant. Overall, teachers agreed to strongly agreed that the 25 indicators of effective leadership were present in their schools.

Table 8: Cohort 3 Initial (n=461) and Post Program Means (n=472) on Teacher Survey.

Statement	Initial	Post	Difference
------------------	----------------	-------------	-------------------

There is a clear academic vision for this school	4.29	4.36	0.07
The principal at this school lets staff know what is expected of them	4.47	4.60	0.13
I understand the vision of my school	4.35	4.41	0.06
The principal at this school is an inspiring leader	4.49	4.49	0.00
I feel my contributions at this school are important	4.47	4.41	-0.06
The principal makes adjustments when things aren't working at this school	4.36	4.37	0.01
The principal work to make the vision of this school a reality	4.53	4.57	0.04
My professional development is tailored to my students' needs	3.99	4.19	0.20**
Expectations are high at this school	3.82	3.94	0.12*
Teachers take responsibility for student achievement at this school	3.86	3.94	0.08
The school is a caring and nurturing place	3.92	3.97	0.05
I like working at this school	4.34	4.39	0.05
The principal at this school has confidence in the expertise of the teachers	4.44	4.47	0.03
Parents take responsibility for student achievement at this school	3.58	3.62	0.04
I have an opportunity to participate in leadership roles at this school	4.09	4.12	0.03
The principal at this school is open to constructive feedback	4.32	4.35	0.03
The principal at this school invites teachers to play a meaningful role in setting goals for the school	4.26	4.34	0.08
The principal at this school give regular and helpful feedback about my teaching	4.15	4.33	0.18**
The principal at this school encourages collaboration among teachers to increase student learning	4.41	4.47	0.06
The principal takes into account the feedback I give	4.24	4.33	0.09~
The principal inspires optimism for the future of this school	4.38	4.41	0.03
The principal encourages my career development	4.44	4.45	0.01
I feel respected by the school principal	4.60	4.63	0.03
I feel supported by the school principal	4.48	4.51	0.03
The principal deals with daily tasks and daily problems in an effective and efficient manner	4.41	4.46	0.05

Note: ~ p < .01, * p < .05, ** p < .001, *** p < .0001

Principals' Focus Group

Focus group results consistently showed that participants got a great deal out of the program

and that they used their new knowledge in a variety of ways at their sites. In particular, participants highlighted their work with their teachers and their increased ability to create and sustain a positive culture. Looking at Cohort 2 specifically, participants talked about how they gained the skills to empower their teachers to engage in improvement as well as how they were more oriented towards building strong teams among teachers and administrators to support academic achievement. Other areas of implementation included engaging more productively with families and the larger community. Participants also mentioned how the program experience enhanced their interpersonal skills saying they had become better listeners and were now better able to deal with critical and/or stressful situations.

Cohort 3 participants picked up on these threads and also emphasized their enhanced leadership skills as a result of their participation in the Advanced Leadership Program. For example, one of the participants said, “One of the skills I have earned from the first module is ‘discover your style’ and my style was political leader. I started using this aspect to manage the school and keep it stable without addressing penalties. I started finding support team for projects and take advantage of the existing capabilities at the school.” Many participants commented too on how their newly acquired behaviors served to help them shift the larger culture of their school, and specifically changed how they worked with teachers to engage in continuous improvement. One participant said, “I learned skills to identify patterns. This helped me to delegate responsibilities in my school and ability to analysis data and planning that helped me make the right decisions based on the available data. I am proud of school teachers, how they understand the strategy of collecting data and evidence for any problem they faced. As a result, our planning skills became better and we know each day in the semester what to do.”

Effectively working with teachers also came up in numerous comments related to motivation, instruction, and classroom resources. Others commented on relationships with students, beginning a student leadership program, and interacting with parents in a more effective manner. Finally, it was clear from participants comments that the program would have long lasting positive effects on their practice. As one participant explained, “I still refer to my material and training notes and I felt the impact as a leader in the field. Many things I learned reflected my personality and I believe in these practices transferred to teachers.”

Section II: Data Collection and Analysis

This section focuses on Issue 2 – that of developing core team members’ knowledge and skills to be effective instructors as well as ensuring the ability of the program to thrive over time. The research questions guiding this focus were:

1. Do core members provide effective instruction of module content?

Methods

While we collected information regarding core members content knowledge and views of the coaching they received for the pilot year and for new core team members for Cohort 2, for Cohort 3 and with the end of the coaching and core members taking full responsibility for instruction, we used only participants’ ratings of the content and instruction in each module. As we

discuss here, results from these items show core members to be increasingly well trained and highly knowledgeable with participants rating their instruction, on average, as high as or higher than the UConn instructional team and better in each year of implementation.

Feedback Surveys of Module Effectiveness (Content and Instruction)

At the conclusion of each module, participants were sent a survey via email asking them to evaluate its instruction and content and register their overall satisfaction with the program. All questions were rated using a Likert scale. At the conclusion of the survey, participants were given space to write additional comments regarding the module.

Results were analyzed descriptively to help better understand potential differences regarding participants' learning and growth and the degree to which the current modules and teaching was experienced positively and aligned with participants' perceived needs and learning goals. While we discuss differences in satisfaction levels between participants in Cohort 3 and Cohort 1 & 2 on all measures, we do not conduct statistical comparisons between them to assess whether the ratings were statistically significantly higher year to year. We based this decision on the fact that though the modules are the same year to year, neither the participants filling out the surveys nor the instructors were the same across all three years. In this way, there might be issues related to the nature of the group, the instructor and/or these elements in combination that would be responsible for the differences more so than the content or instructional practices driving these differences. We do highlight when the numbers are higher as a way to get at overall trends, in this case enhancements to module performance. We also analyzed, for Cohort 2 and 3 whether there existed differences in how female and male participants viewed the modules' content and the instructors' teaching.

Module Content and Instruction

Data for this part of the evaluation came from two sources: (1) post module, participant surveys on their level of satisfaction with the content and instruction of the module and (2) focus groups who were asked to reflect on the program and its impact on their practice.

Post Module Feedback Surveys

At the conclusion of each module participants were asked, via a survey, to reflect on the structure and content of the module and the instructional practices deployed by the instructor. With an approximately 93% average response rate across the modules for both Cohort 2 and Cohort 3, and 89% for Cohort 1, these surveys provided valuable information for the instructional and leadership team to modify and enhance the modules in real time. Indeed, comprehensive reports for each module given to QRTA supported modifications to the content and instruction which provided (as seen in the results here and in the participant's pre/post assessment) positive results to participant knowledge and skills.

Results of Feedback on Module Content

When we look across the results from these surveys, we find an upward trend with satisfaction rates in Cohort 2 and 3 being much higher than in the pilot year (i.e., for Cohort 1) and on par with one another. Again, given the increased size of the cohort and move for full instructional ownership to the QRTA core team these results are quite positive and should be celebrated.

Beginning with Cohort 2, participants' reports on their satisfaction with the content of the modules (See Table 9), indicated a relatively high level of overall satisfaction - 88.5%. This was an almost 8 percentage point increase in satisfaction from Cohort 1. The overall range of scores was also quite tight and positive with the lowest score on any aspect of any module being 82.6% and the highest being 92.8%. These findings point to the success of the content modifications made prior to the beginning of the second round of modules. Indeed, across the board, with the exception again of Module 2 which showed a less than 2 percentage point decline in participant satisfaction relative to Cohort 1 members' responses, Cohort 2 members were much more satisfied with the content of the modules than in the pilot year. Indeed, there were positive differences in virtually all of the average scores ranging from an increase of 5 percentage points to over 15 percentage points from Cohort 1 to Cohort 2.

In terms of areas for growth, there were only two items for which participants, on average, expressed satisfaction under 88.5% and that these scores were substantively higher than they were for Cohort 1. As true for Cohort 1, these included how time was used during the module and whether it was appropriate to cover the content (83.75% satisfaction, up almost 8 percentage points from last time) and whether the activities were applicable, and content based 84.45% satisfaction up 3 percentage points).

Now shifting to the highest rated items, we find two of the highest rated items, on average, were the same for Cohort 2 as for Cohort 1 and each was again rated higher than last time. The two repeat items were that objectives were clearly explained at the beginning of the session (91.05%) and that the materials made a valuable contribution to participant learning (90.5%). The third item also at a 90.5% average satisfaction rating was that the module material made a valuable contribution to the participants' learning. As true with Cohort 1, these findings for Cohort 2 suggest participants found the content relevant, clear and useful.

Finally, when disaggregating participants' various scores by gender we find there were sometimes small differences (always less than .4 points) between male and female participants' responses. Specifically, men rating all aspects of module 1 and 4 slightly higher than their female colleagues did (about .2 points) and women rated the content in module 2 significantly higher (also an average of .2 points) than their male colleagues.

For Cohort 3 (See Table 10), we again see a relatively high level of overall satisfaction - 87%. This is on par with the satisfaction rate from Cohort 2 (88.5%) and still far higher than that of Cohort 1. The overall range of scores was tighter than that of Cohort 2, suggesting again that there was greater consistency in participants responses within and across the modules. The lowest score on any aspect of any module being 83% and the highest being 92%. These findings point to the continued success of the content modifications made prior to the beginning of the second and third round of modules and in direct response to participants' feedback and the evaluation team's

analysis. As true with the overall satisfaction rate, Cohort 3 members' satisfaction closely mirrored those reported by Cohort 2 members and were much better than Cohort 1's reports. There were also a number of instances where Cohort 3 satisfaction rates were higher than Cohort 2's – again, a remarkable finding given the increased size of the cohort, their corresponding learning needs, and the core team's responsibility in terms of leading the modules with a lesser degree of UConn support.

In terms of areas for growth, it is worth noting and as true with Cohort 2, there were only two items for which participants, on average, expressed satisfaction under 88.5% and that these scores were substantively higher than they were for Cohort 1. As true for Cohort 1 and 2, one of these items was how time was used during the module and whether it was appropriate to cover the content (86.4% satisfaction, still up almost 3 percentage points from Cohort 2 and 11 percentage points from Cohort 1). Varying from the other cohorts the second lowest rated item was related to logistics and whether supports were appropriate. The average satisfaction across the modules was 86.3% and approximately 2 percentage point difference from Cohort 2 but still substantively up from Cohort 1. Indeed, given the relatively high level of satisfaction even among these relatively lower items, it appears that the program is quite strong in terms of module content providing a strong foundation for further improvements moving forwards.

Now shifting to the highest rated items, we find two of the highest rated items, on average, were the same for Cohort 3 as for Cohort 1 and 2. The two repeat items were that objectives were clearly explained at the beginning of the session (90.02%) and that the materials made a valuable contribution to participant learning (90%). As true with Cohort 1 and 2, these findings for Cohort 3 suggest participants found the content relevant, clear and useful. Such findings continue to bode well for the program as it moves into its second phase of implementation.

Cohort 3 participants' varied less by gender in regards to their scores for the modules than in Cohort 2 though these differences were similarly small (on average .2 points). Specifically, men rated all aspects of module 2 (instruction and content) higher than women and the instruction for module 3 higher. Given the very small average differences between these scores, these findings may be more a function of the differing sizes of the groups (many more women) than true differences in perceptions.

Table 9: Cohort 2 Participants' Satisfaction with Module Content

Statements	Module 1		Module 2		Module 3		Module 4		Overall Satisfaction
	Mean/ All Participants	Level of Satisfaction							
The objectives were explained clearly at the beginning of the module	4.63	92.5%	4.47	89.3%	4.54	90.7%	4.57	91.4%	91.05%
The module objectives were clear	4.64	92.8%	4.38	87.5%	4.43	88.7%	4.63	92.7%	90.4%
The module objectives were achieved by its conclusion	4.49	89.7%	4.25	85%	4.3	85.9%	4.43	88.7%	87.35%
The module materials made a valuable contribution to my learning	4.65	93.1%	4.47	89.3%	4.41	88.3%	4.57	91.5%	90.5%
The module activities were applicable, and content based	4.5	90.1%	4.48	89.5%	4.27	85.4%	4.44	88.8%	88.45%
The module added to my knowledge and skills in the topic	4.62	92.4%	4.49	89.8%	4.42	88.5%	4.51	90.4%	90.2%
The module assignments were applicable, and content aligned	4.57	91.4%	4.44	88.7%	4.31	86.3%	4.49	89.3%	89.05%
The module content added to the skills I can use in my work	4.61	92.2%	4.52	90.4%	4.44	88.7%	4.55	91%	90.6%
Time allocated for the module was appropriate to cover its content sufficiently	4.49	92.2%	4.13	82.6%	4.03	80.5%	4.1	82%	83.75%
The module was well-organized	4.64	89.8%	4.38	87.5%	4.23	84.7%	4.45	88.9%	88.5%
Logistics support was appropriate	4.62	92.8%	4.4	88%	4.23	84.7%	4.52	90.5%	88.85%
Overall Satisfaction	4.5	92.4%	4.33	86.7%	4.23	84.7%	4.55	91%	88.05%

Table 10: Cohort 3 Participants' Satisfaction with Module Content

Statements	Module 1		Module 2		Module 3		Module 4		Overall Satisfaction
	Mean/ All Participants	Level of Satisfaction							
The objectives were explained clearly at the beginning of the module	4.52	90.40%	4.52	90.40%	4.35	87.10%	4.59	92%	89.9%
The module objectives were clear	4.53	90.60%	4.57	91.50%	4.34	86.90%	4.6	92%	90.2%
The module objectives were achieved by its conclusion	4.45	89.10%	4.37	87.40%	4.33	86.70%	4.49	90%	88.20%
The module materials made a valuable contribution to my learning	4.56	91.30%	4.49	89.80%	4.33	86.60%	4.62	92%	90.00%
The module activities were applicable, and content based	4.51	90.30%	4.45	89%	4.32	86.40%	4.55	91%	89.00%
The module added to my knowledge and skills in the topic	4.56	91.30%	4.49	89.80%	4.31	86.30%	4.58	92%	89.70%
The module assignments were applicable, and content aligned	4.49	89.80%	4.48	89.60%	4.29	86%	4.5	90%	88.80%
The module content added to the skills I can use in my work	4.48	89.60%	4.44	88.80%	4.28	85.70%	4.53	91%	88.65%
Time allocated for the module was appropriate to cover its content sufficiently	4.39	87.70%	4.19	83.80%	4.25	85.20%	4.45	89%	86.40%
The module was well-organized	4.59	91.70%	4.33	86.60%	4.23	84.80%	4.58	92%	88.65%
Logistics support was appropriate	4.33	86.70%	4.35	87%	4.15	83%	4.43	89%	86.30%
Overall Satisfaction	4.38	87.60%	4.29	85.80%	4.19	83.90%	4.47	89%	87.00%

Principals' Focus Group

As highlighted in the findings from the surveys, focus group participants were generally very positive regarding the module content across Cohorts. Moreover, they grew more positive each year as QRTA worked to enhance the content to better serve the needs of the participants and to include specific and clear references to the Jordanian context. In Cohort 2 for example, though virtually all focus group participants said the modules' content was excellent and that they appreciated the focus on the practical use of the content, they all also asked for additional Jordanian-based examples in how module learning could be translated to daily practice.

In contrast, while some of the Cohort 3 participants also asked for Jordanian examples, these requests were framed as additional resources. In other words, it appeared both by the frequency and the nature of these requests that the program had done a good job of adding a good number of such examples into the program. Beyond these comments, Cohort 3 participants were even more positive in terms of highlighting how the module content was directly applicable to their work and how they had changed the way they viewed their role, their school, and staff interactions as a result of the Advanced Leadership Program. As one participant related, "I earned different quality experiences and the program opened new horizons for our thinking and aspiration to achieve the best we can. The program met the purpose in an excellent way." Another participant talked specifically about how the concept of the instructional core (i.e., the relationship between the teacher, the curriculum and the student) and introduced and reinforced throughout the modules had impacted their entire way of thinking about instructional leadership. "The educational core concept made me understand that anything outside these three items (student, teacher, curriculum) is not an obstacle. We started using the four frames lens and use areas of strength and this gave us a holistic view of the student, teacher, community, resources and opportunities. Previously, we only looked at the student." Taken together, it is clear that the module content continues to meet the needs of participants and does so increasingly effectively over time.

Results of Feedback on Instructors' Performance

Before moving to the findings, it is important note that there were shifts in instructors each year of the program. In the pilot year (Cohort 1), the modules were taught solely by UConn faculty. For Cohort 2 and 3 the modules were taught by QRTA core members. Additionally, and due to the increased number of participants in the program for Cohort 3, there were multiple sections of each module. Each section was taught by a different instructor. Therefore, given changes in the program and the differences in structure and substance of the program over the years (e.g., change to who teaches a module, change to the content) straight comparisons between the modules as experienced by Cohort 3 and prior cohorts is difficult and perhaps inappropriate. However, as this was the third year of the program, we are able to compare the rating from the prior years to see whether and to what degree the module instruction for a given module in the aggregate and content adjustments were effective over time. As we discuss in more detail later, we find the overall average level of satisfaction with module instruction grew higher each year, shifting from 88.9% for Cohort 2 to 89.6% for Cohort 3. Moreover, the overall

positive changes in scores from Cohort 1 to 2 were maintained. For the three other modules were higher for Cohort 2. For example, for Module 3 which was rated at 69.4% satisfaction for Cohort 1, received a score of 86% satisfaction for Cohort 2 and 88.6% for Cohort 3. We discuss all these results and more below.

As seen in Table 11, for Cohort 2, the overall average level of satisfaction with module instruction was very high at 88.9% – almost 5 percentage points higher overall than it was for Cohort 1. Moreover, with the exception of module 2 for which the overall satisfaction was slightly lower than it was for Cohort 1 members, the overall scores for the three other modules were higher for Cohort 2. This was particularly true for Module 3 which was rated at 69.4% satisfaction last time for Cohort 1 and at 86% satisfaction for Cohort 2 - an almost 17 percentage point increase between the two cohorts.

The range of scores across the modules was far tighter and higher for Cohort 2 than for Cohort 1 where the instruction ratings for Module 3 were much lower than the other modules – so much so that in the report for Cohort 1 we presented the data with and without the data from that module in our final averages. In this case, for Cohort 2, the range of satisfaction scores for instruction across the modules was 81.3% to 98%. As one can see, these numbers are very encouraging with the lowest rating items still showing that participants were approximately 81% satisfied (higher than the prior cohort even when Module 3 was excluded). In this way, these numbers suggest Cohort 2 members experienced better instruction than their colleagues the prior year.

Looking a bit more closely at specific questions, as true for Cohort 1, two of the highest rated items included participants' view that the instructor treated all participants with respect (96.95%) and the instructor showed interest in helping participants learn (93.65% satisfaction). The third highest item was that the instructor managed discussions in ways that facilitated discussion (90.15% satisfaction). Across the board, these average ratings were higher than for Cohort 1 and again signal that participants felt the instructors were respectful and student centered in terms of approach and engaged with participants in ways that invited participation and discussion.

There were only two items for which, on average, Cohort 2 participants said they were, on average, less than 88% satisfied and these areas were the same as for Cohort 1. First, the lowest rated item for instruction on average was whether they provided real life examples (87.5% satisfaction). The next item was whether the instructor responded to participants' question appropriately (87.6% satisfaction). It is worth noting that for each of these items, the scores for Cohort 2 were nearly 10 percentage points higher than for those for Cohort 1. In this way, though these remain the lowest rated items it is also important to recognize the tremendous growth in both arenas and that the satisfaction rates are still over 87% suggesting few weaknesses on average in the program.

Shifting now to Cohort 3 it is worth noting the range of scores across the modules was even tighter than for either Cohort 1 or 2. In this case, for Cohort 3, the range of satisfaction

scores for instruction across the modules was 86.9% to 96%. This approximate 9 percentage point spread is almost half of that for Cohort 2 and again suggests greater agreement among Cohort 3 members regarding their ratings. It also hints at the fact that despite the larger and perhaps more diverse needs of Cohort 3 – a greater proportion felt adequately served by the program. As one can see, these numbers are very encouraging with the lowest rating items still showing that Cohort 3 participants were approximately 87% satisfied, a higher floor score than for either prior cohort. These numbers suggest Cohort 3 members experienced better instruction than their colleagues in either of the prior years. In this way, QRTA has much to celebrate in terms of the preparedness and capacity of their instructional team to engage in high quality instruction for many cohorts of participants to come.

Looking more closely at the specific questions highlighted in this section, as true for the prior cohorts, two of the highest rated items included participants' view that the instructor treated all participants with respect (93.4%) and the instructor showed interest in helping participants learn (91.05% satisfaction). The third highest item was different for Cohort 3, it was that the instructor was well prepared for the module (89.95% satisfaction). Across the board, these average ratings were still higher than for Cohort 1 and on par with Cohort 2 and again signal that participants felt the instructors were respectful and student centered in terms of approach and were ready, willing, and able to effectively lead module learning.

There was only one item for which, on average, participants said they were, on average, less than 88% satisfied and one for which they were 88.05% satisfied. These areas were the same as for Cohort 1 and 2. First, the lowest rated item for instruction on average was whether they provided real life examples (88.05% satisfaction). The next item was whether the instructor responded to participants' question appropriately (87.75% satisfaction). It is worth noting that for each of these items, the scores were higher than those given by either Cohort 1 or Cohort 2 members and very high overall. Indeed, though these remain the lowest rated items it is also important to recognize the tremendous growth from the first implementation of the modules and that the satisfaction rates are still at or above 88% suggesting few weaknesses on average in the program.

Table 11: Cohort 2 Participants' Satisfaction with Module Instruction

Statements	Module 1		Module 2		Module 3		Module 4		Overall Satisfaction
	Mean/ All Participants	Level of Satisfaction							
The instructor presented the module material clearly	4.63	92.6%	4.22	84.3%	4.24	84.8%	4.75	95%	89.2%
The instructor was well prepared for the module	4.75	95%	4.22	84.5%	4.33	86.7%	4.68	93.7%	89.9%
The instructor responded to questions adequately	4.61	92%	4.07	81.3%	4.35	87%	4.49	89.8%	87.6%
The instructor stimulated interest in the subject	4.76	95.2%	4.18	83.6%	4.32	86.4%	4.65	93%	89.55%
The instructor showed interest in helping participants learn	4.86	97.2%	4.52	90.4%	4.63	92.6%	4.72	94.4%	93.65%
The instructor gave clear and relevant assignments during the module	4.69	93.8%	4.33	86.7%	4.3	86.1%	4.55	91%	89.35%
The instructor managed discussions in ways that facilitated participation	4.61	92.2%	4.41	88.2%	4.4	88.1%	4.61	92.2%	90.15%
The instructor managed instructional time effectively	4.8	95.7%	4.26	85.1%	4.34	86.8%	4.54	90.9%	89.7%
The instructor treated all participants with respect	4.9	98%	4.8	95.9%	4.8	96.1%	4.89	97.8%	96.95%
The instructor's teaching methods promoted participants' learning	4.62	92.4%	4.34	86.8%	4.39	87.8%	4.63	92.7%	89.9%
The instructor provided real-life examples to explain topics	4.49	89.8%	4.27	85.4%	4.31	86.2%	4.43	88.8%	87.5%
Overall Rating of Instructor	4.75	95.1%	4.1	82.1%	4.34	86.8%	4.59	91.9%	88.9%

Table 12: Cohort 3 Participants' Satisfaction with Module Instruction

Statements	Module 1		Module 2		Module 3		Module 4		Overall Satisfaction
	Mean/ All Participants	Level of Satisfaction							
The instructor presented the module material clearly	4.47	89.40%	4.51	90%	4.45	89.10%	4.54	90%	89.85
The instructor was well prepared for the module	4.53	90.50%	4.48	90%	4.47	89.40%	4.51	90%	89.95
The instructor responded to questions adequately	4.35	86.90%	4.42	88%	4.37	87.60%	4.41	88%	87.75
The instructor stimulated interest in the subject	4.47	89.50%	4.4	88%	4.38	87.10%	4.45	89%	88.5
The instructor showed interest in helping participants learn	4.64	92.80%	4.51	90%	4.49	90%	4.57	91%	91.05
The instructor gave clear and relevant assignments during the module	4.5	90%	4.49	90%	4.43	88.70%	4.51	90%	89.65
The instructor managed discussions in ways that facilitated participation	4.45	89.10%	4.46	89%	4.47	89.50%	4.5	90%	89.4
The instructor managed instructional time effectively	4.5	90.10%	4.51	90%	4.41	88.30%	4.49	89%	89.55
The instructor treated all participants with respect	4.8	96%	4.64	89%	4.55	91.10%	4.69	93%	93.4
The instructor's teaching methods promoted participants' learning	4.4	87.90%	4.45	89%	4.46	89.30%	4.52	93%	89.15
The instructor provided real-life examples to explain topics	4.41	88.20%	4.41	88%	4.34	86.90%	4.45	89%	88.05
Overall Rating of Instructor	4.5	90%	4.5	90%	4.43	88.60%	4.49	89%	89.6

Principals' Focus Group

Across all Cohorts, focus group members reinforced much of the survey results in that they felt that the instruction was strong and effective and participants felt that the instructors really listened to their concerns and facilitated group participant and the building of ideas. Participants also appreciated the practical orientation of the instructors and their experience in leading schools and school systems. There was some variation in terms of whom the participants saw as the most effective instructor but there were few if any negative comments regarding any particular instructor's capabilities or approach. Views on the assignments was somewhat mixed for Cohort 2 and 3. Some participants felt the assignments were overly burdensome or they were not particularly well explicated. While it is somewhat difficult to assess whether the current balance between the assignment rigor and demand and participant views of these assignments is appropriate, cohort members' comments provide opportunities to consider this balance and whether modifications are useful and appropriate.

With that said, and focusing in on Cohort 3's focus group members' responses a bit more closely, it was clear that they experienced the program extremely positively and that the instructors were key to these feelings. As one participant put it, "I have 13 years of experience, but after finishing this diploma I felt that I was reborn as a principal." Another remarked on how the program and the facilitation of the modules created opportunities to experience new ideas and expand understanding of Jordan, its schools and, its students, "the beautiful thing in the diploma is that the principals were from all regions of the Kingdom in the same hall, there was a lot of exchange of experiences, we have learned a lot from the experiences, even each school is different but we took these experiences and fit them to our schools. It is not only what we have learned from the diploma that has raised our efficiency, but also the huge amount of experiences we have gained." Once again, it is clear that the program has created a number of positive results over the year and has the capacity to do so far into the future.