LEADERSHIP AND TEACHER DEVELOPMENT PROGRAM

مستقبل أفضل

تعلم عالي الجودة

قيادة تربوية

Is LTD's Professional Development Program Contributing to Improved Student Learning as Measured by Standardized Tests?

Report of an Impact Evaluation Study Conducted by the Leadership and Teacher Development (LTD) Program in Cooperation with The Assessment and Evaluation Department, MoEHE

September 2017

Table of Contents

Figures

Tables

Acknowledgements

Executive Summary

Background and Purpose

The Leadership and Teacher Development (LTD) program focuses on educational reform in the West Bank and Gaza (WBG) and aims to have a significant impact on teaching and learning practices, and on individuals and institutions involved in the education process at large. A chief goal of the program is to upgrade the qualifications, competencies and skills of teachers and improve the tools for teaching and learning, thereby improving the quality of primary and secondary education. Since 2013, three cohorts of educators in the West Bank comprising 304 principals and 1454 teachers have completed comprehensive, school-based training in leadership and teaching practices. As of this writing, a fourth cohort of principals and teachers from 151 schools is currently midway through its training cycle.

> This report presents multiple sources of quantitative and contextual evidence indicating that LTD's model of school-based reform and teacher professional development is contributing to improved student learning.
>
> - Students of LTD teachers scored higher on standardized tests of achievement in Arabic, English, mathematics, and science/biology than did students of non-LTD teachers.
>
> - LTD students scored higher than did their non-LTD peers on test items measuring the cognitive skills of knowing, applying, and reasoning across the subjects of Arabic, Math, and Science/Biology.
>
> - LTD students rated their classrooms as more "learner-centered" than did their non-LTD peers.

Central Research Question

With the LTD Program now in its final year, AMIDEAST, in collaboration with the Assessment and Evaluation Department of the Ministry of Education and Higher Education (MoEHE), undertook an impact evaluation study to examine the following question:

Is there empirical evidence indicating that LTD's teacher professional development program is contributing to improved student learning as measured by scores on standardized tests of academic achievement?

Methods, Samples, and Tools

To approach this question, LTD, in coordination with the Assessment and Evaluation Department (AED) of the Ministry of Education and Higher Education (MoEHE), administered standardized tests of achievement in May 2017 to a random sample of 3054 students of LTD teachers in four core subjects: Arabic (8th grade), mathematics (6th grade), English (7th grade), and science/biology (10th grade). The sample was stratified by cohort so that students of teachers from each of the three completed cohorts were represented. The student is thus the main unit of analysis in the study.

For the purpose of comparison, the study included a control group of 991 randomly selected students who were taught by a sample of teachers randomly selected from non-LTD schools from all districts in the West Bank. These teachers were similar to their LTD counterparts in key ways: they taught the same curriculum; used the same textbooks; taught to comparable class sizes and hours per week; and, were supervised by the same district supervisors.

In sharp contrast, however, the non-LTD teachers differed from their LTD peers in two important ways. First, they had not participated in any trainings or workshops comparable in scale or duration to LTD's yearlong learner-centered and school-based interventions supported by LTD-trained district leadership. Second, they had far less access to educational resources like those supplied by AMIDEAST to LTD schools, such as full Internet connectivity and school-wide Wi-Fi access, laptops, LCD projectors, and interactive whiteboards.

Contextual factors such as these are thus critical to interpreting the differences in exams scores between LTD and non-LTD students. For this reason, the study also captured data about the students' perspectives on the classroom context. Both groups of students filled out a questionnaire for the "Classroom Context Survey" comprised of a 20-item scale (pre-tested for reliability) designed to measure indicators of learner-centered instruction in the classroom.

Research Questions

In light of AMIDEAST's robust and sustained training and technical interventions provided to LTD teachers, schools and districts, our study predicted (hypothesized) that:

- Scores on standardized exams for LTD students would be higher than scores of students taught by non-LTD teachers.

- Scores on standardized exams for LTD students would correlate more positively with scores on the contextual survey compared to non-LTD students.

- Scores on standardized exams for LTD students would increase in relation to the duration of time that teachers had to transfer and develop their teaching competencies and strategies after their training had ended. (For this reason, we stratified the sample of LTD students by cohorts of teachers: 2013-2014; 2014-2015; and, 2015-2016.)

Summary of Key Findings

From our analysis of multiple sources of data—exam scores and contextual data—we find sufficient evidence, as shown in the tables below, to support the claim that LTD's model of teacher professional development is contributing to improved student learning. This is a consequence, we argue, of LTD's comprehensive school-based approach to developing teachers' competencies and their use of LTD-provided technology to foster learner-centered teaching and learning.

1. **Content Knowledge and Thinking Skills:** Compared to students of non-LTD teachers, **Figure 1** shows that students of LTD teachers scored higher on standardized tests of achievement in Arabic, English, mathematics, and science/Biology. There is a 5 to 6 percent difference between the scores of LTD and Non-LTD students (**Table 1**).

Although the variance of the means between the LTD and Non-LTD scores were statistically significant for the English exam only, additional contextual evidence presented below strongly suggests that the differences observed in the other exams are, practically speaking, meaningful.
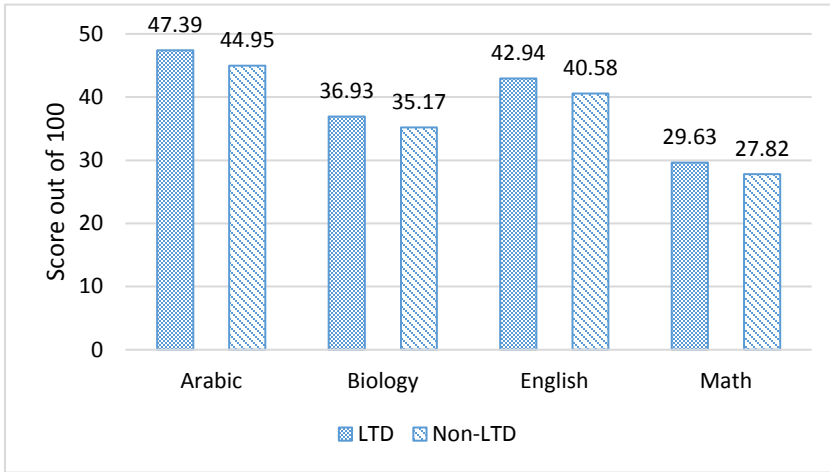
Figure 1. Results of exams for LTD and non-LTD students

Table 1. Percentage difference between LTD and Non-LTD scores

Exam	LTD	Non-LTD	% Difference
Arabic	47.39	44.95	5.28%
Biology	36.93	35.17	4.88%
English	42.94	40.58	5.65%
Math	29.63	27.82	6.30%

The exam items tested not only students' content knowledge, but also their thinking skills. Each test item was designed to measure one of three cognitive domains: knowing, applying, and reasoning. The analysis revealed that LTD students scored higher than their non-LTD counterparts did in each of these domains for Arabic, Math, and Science (**Table 2**).

Table 2. Results of analysis of cognitive domains

	Arabic			Math			Science		
	LTD	Non-LTD	Difference	LTD	Non-LTD	Difference	LTD	Non-LTD	Difference
Knowledge	47.17	43.52	8.04%	44.80	43.60	2.71%	51.14	51.56	0.83%
Application	54.25	52.13	3.98%	23.65	21.11	11.35%	39.89	38.40	3.81%
Reasoning	41.25	39.25	4.97%	15.42	13.98	9.78%	22.14	18.28	19.08%

2. **Quality of Learner-Centered Instruction:** Even though the variance in the means of the two groups—LTD and non-LTD—were not statistically significant, with the exception of English exam scores, the fact that LTD students rated their classrooms as more "learner-centered" than their non-LTD peers, as seen in **Figure 2**, suggests that the differences are in fact meaningful. The contextual questionnaire used a 5-point scale, with 5 being the highest.
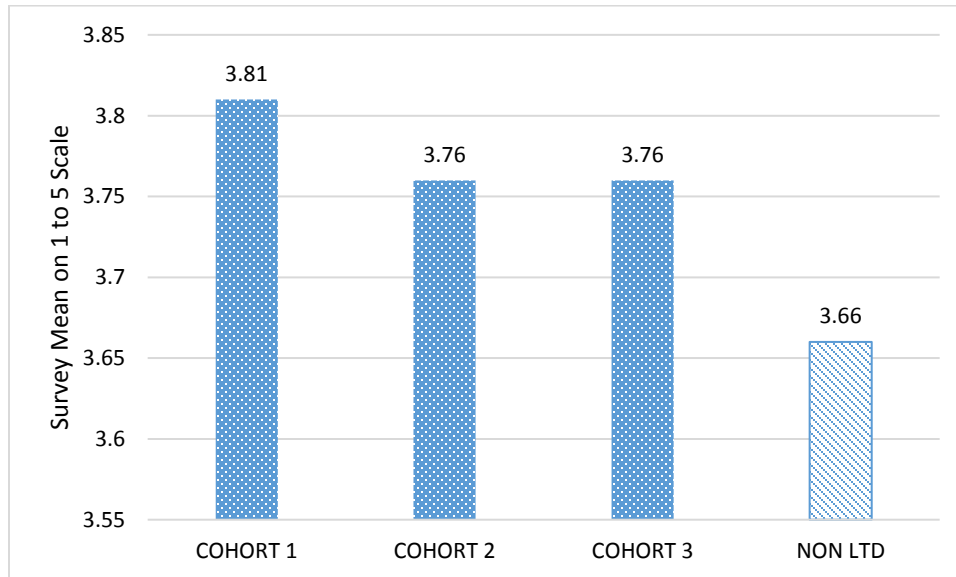
Figure 2. Results of Classroom Engagement Survey by LTD cohort and non-LTD

Additionally, the analysis found a higher positive correlation (r = .373) between the exam scores and the contextual survey for LTD students compared to their non-LTD peers (r = .273). That is, student achievement appears to increase as classrooms become more learner-centered, which clearly appears to be the case for students taught be LTD teachers (**Figure 3**).
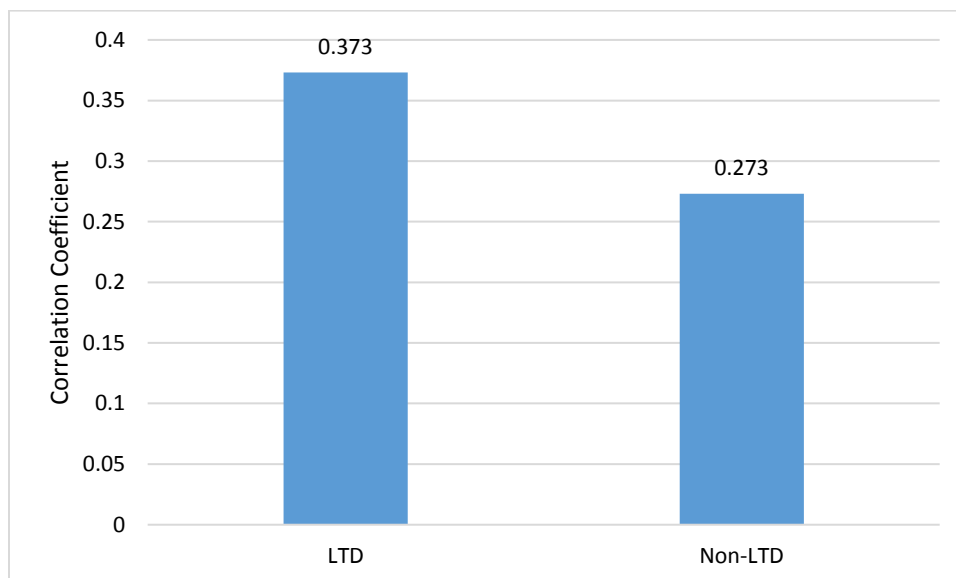
Figure 3. Correlation between Classroom Engagement Survey and exam results

3. **Post-LTD Time Span and Student Achievement**: Finally, we found promising evidence to support our assumption that student learning improves when teachers have sufficient time (and support) to transfer and develop skills acquired in LTD professional development experience. As shown in **Figure 4**, Cohort 1 students scored higher in English and Math than students in Cohorts 2 and 3. We also observe that students in Cohort 2—whose teachers had two years to develop their LTD skills—scored higher than did students of Cohort 3 in Arabic, Biology, and Math.   In no case were the scores of Cohort 3 higher than the scores of Cohort 1, though the means for the Arabic were identical.
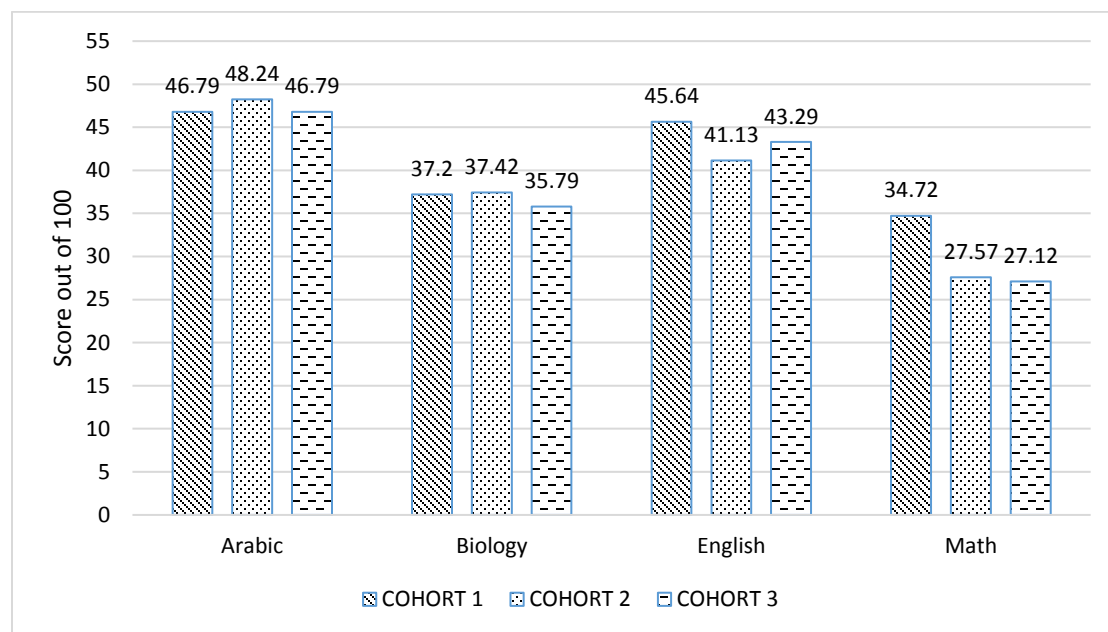
Figure 4. Exam scores by cohort

Conclusion

In light of the rigor of the study's research design, these findings lead us to conclude confidently that LTD's model of teacher professional development is making a positive contribution to student learning. The LTD model is a learner-centered approach that empowers teachers to work both independently and in professional learning communities to strengthen their competencies, self-confidence, and self-efficacy as professional educators. In this way, LTD's approach optimizes opportunities for teachers to engage in ongoing professional development and to take greater responsibility for their own growth as effective teachers. As is well documented in the research literature, effective teachers are the most important factor contributing to student achievement (Barber & Mourshed, 2007; Goddard, Hoy, & Hoy, 2000; Gutierrez, Weinberger, & Engberg, 2016; Heck, 2009; Kyriakides, Christoforou, & Charalambous, 2013; Vescio, Ross, & Adams, 2008). Aligned with comprehensive school-based reform, LTD's model of teacher professional development is producing effective teachers.

Purpose and Background

Purpose

The purpose of the study is to determine if there is empirical evidence to validate LTD's model of school-based professional development and its impact on improving the quality of teaching and learning. Results of the study are important in helping policymakers in the Ministry of Education and Higher Education and international donors to make evidence-based decisions on planning sustainable educational reforms that will improve the quality education for all children in the West Bank and Gaza. In short, the immediate aim of the study is to answer the following question:

Is there empirical evidence that LTD's teacher professional development program is contributing to improved student learning as measured by scores on standardized tests of academic achievement?

Background

Three cohorts of school principals and teachers since 2013 have completed LTD's yearlong diploma programs in school leadership and teacher qualification. Framed by standards and competences for effective schools established by the Ministry of Education and Higher Education, the crux of LTD's model of school-based reform rests on the principles of shared leadership and learner-centered instruction. As of this writing, LTD's fourth and final cohort is in progress and will be completed by December 2017.

Based on quarterly and annual reporting by LTD and its key ministry partners, the National Institute for Educational Training (NIET), the Directorate for Field Follow-Up (DFF), and the Directorate for Supervision and Qualification (DSQ), the MoEHE and USAID are confident that LTD's model of compressive school-based reform for leadership and instruction is a successful model for supporting the goals of the Ministry's five-year strategic planning focused on improving the quality of learning for all students in the West Bank and Gaza's public schools (MoEHE, 2008, 2014). Additionally, several research publications describing education reforms in the West Bank and Gaza associated with AMIDEAST/USAID programs offer both anecdotal and empirical evidence to support LTD's belief that it is making a difference in the quality of school leadership and instruction (Cristillo, Iter, & Assali, 2016; Khales, 2016; Shinn, 2012).

Whereas most evidence to date rely on self-reported data from school principals, teachers, and students, this present study uses student achievement based on standardized tests to evaluate LTD's impact on improving teaching and learning.

Research Design and Methods

Using a quasi-experimental design, LTD, in cooperation with AED and school directorates, administered standardized exams for Arabic (8th grade), mathematics (6th grade), English (7th grade), and science/biology (10th grade) to a random sample of 3054 students of LTD teachers and a random selection of 991 students taught by teachers in non-LTD schools. The exams were administered over two days, May 2-3, 2017.

Sample

Treatment Group: The unit of analysis is the student. Specifically, the study sampled students of LTD teachers who graduated from the yearlong Teacher Diploma Program delivered by NIET between 2013 and 2016. Students were selected using a double random selection process involving the random selection of teachers followed by a random selection of 10 students per class.

- **Sample Frame**: The sample frame comprised all teachers of Arabic, English, Mathematics, and Science (Biology section) who completed the LTD Teacher Diploma Program from 189 schools (Cohorts 1, 2, and 3) located in 14 directorates. The three cohorts completed their Diploma training in 2014, 2015, and 2016 respectively. The sample frame excluded teachers of Technology, as well as teachers from 61 LTD schools whose LTD-trained principals had transferred to other schools. The total number of teachers in the frame for the treatment group was 1271.
- **Random Selection:** A confidence level of 95% and a margin of error 5% produced an ideal sample size of 296 teachers. Every teacher by subject was assigned a random number and then ranked from smallest to lowest. LTD phoned every school to verify the availability of the selected teachers until the targeted size for each subject and cohort was obtained.

Actual sample sizes varied slightly from the target numbers because teachers either had transferred to other schools or were not teaching the same grade level of the test that was to be administered to students. The number of confirmed teachers was 331 (**Table 3**); due to budgeting constraints, this number was adjusted to 300, which generated a student sample of approximately 3000.

Table 3. Actual sample sizes of randomly selected LTD teachers after schools were contacted

Cohort Number	LTD Teachers by Subject				
	Arabic (8th)	English (7th)	Math (6th)	Science (10th)	Total
1	31	20	18	30	99
2	45	32	23	39	139
3	33	17	21	22	93
Total	109	69	62	91	331

Control Group: The sample frame for the control group comprised all West Bank public schools with grades 6, 7, 8, and 10, minus 394 schools because they were either LTD schools or schools of LTD's sister project, the School Support Program (SSP), or schools from a previous AMIDEAST program, the Model Schools Network (MSN). The list was further narrowed by eliminating schools that had fewer than three of the targeted grade levels—a logistical decision to facilitate field visits to schools by testing coordinators. Due to budgetary constraints, the number of teachers chosen for each subject was 25. In all, a total of 100 teachers and 1000 students were chosen by random selection.

Strategy for Randomly Selecting Students: Students were selected using the attendance sheet for each participating class. If a teacher taught more than one section, then "section A" was chosen by default. Each list was then randomized, from which the first 10 students from the top were selected. To ensure the integrity of the selection process, neither the schools nor the test administrators were allowed to select students. A small number of substitutions were necessary, however, on the day of the testing because the selected students were absent.

Table 4 and **Table 5** show the demographics of students whose completed exams were returned to AED for scoring and analysis.

Table 4. Cross tabulation of School type and student gender

SCHOOL TYPE	STUDENT GENDER		Total
	MALE	FEMALE	
NON LTD	468	523	991
	47.2%	52.8%	100.0%
COHORT 1	375	517	892
	42.0%	58.0%	100.0%
COHORT 2	524	777	1301
	40.3%	59.7%	100.0%
COHORT 3	288	573	861
	33.4%	66.6%	100.0%
Total	1655	2390	4045
	40.9%	59.1%	100.0%

Table 5. Cross tabulation by school type and subject

SCHOOL TYPE	SUBJECT				Total
	Arabic	Biology	English	Math	
NON LTD	246	237	260	248	991
	24.8%	23.9%	26.2%	25.0%	100.0%
COHORT 1	280	247	185	180	892
	31.4%	27.7%	20.7%	20.2%	100.0%
COHORT 2	408	359	306	229	1302
	31.3%	27.6%	23.5%	17.6%	100.0%
COHORT 3	305	213	166	177	861
	35.4%	24.7%	19.3%	20.6%	100.0%
Total	1239	1056	917	834	4046
	30.6%	26.1%	22.7%	20.6%	100.0%

Research Hypotheses

Given the robust nature of LTD's yearlong school-based system of professional development and technical support provided to teachers by LTD-trained principals and district leadership, we predicted (hypothesized) three likely outcomes from the study.

1. It was expected that students taught by LTD-trained teachers would score higher on standardized tests of achievement compared to students whose teachers were not LTD-trained.

2. It was expected that students taught by LTD-trained teachers would rate their classroom contexts as more learner-centered compared to students of non-LTD teachers.

3. It was expected that exam scores of students taught by teachers who graduated from Cohort 1 of LTD would be higher than the scores of students whose teachers were from Cohort 2 and Cohort 3.

Instrumentation

Data collection involved the administration of standardized tests of academic achievement in four core academic subjects and grade levels (**Table 6**). The exams for mathematics, Arabic, and science/biology were selected by AED for their reliability and validity. The English exam was developed by AED in 2014 specifically for use by LTD; it has been used twice before and is known to be a reliable tool.

Table 6. Standardized tests and grade levels

GRADE	SUBJECT	SOURCE
6	Math	UNRWA
7	English (Reading, writing, and listening)	AED
8	Arabic (Reading and writing)	UNRWA
10	Science (Biology)	National Exam, 2016

Analytical Methods

Both descriptive and inferential statistics were used. Relationships among the variables—exam scores, classroom context survey, school type (LTD, Non-LTD)—were explored using bivariate correlations, independent-samples t-tests, and one-way analysis of variance (ANOVA).[1]

Data Collection and Data Management

As a joint project of LTD and its MoEHE partner, the Assessment and Evaluation Department (AED), LTD's M&E staff worked closely with the director and senior staff of AED to coordinate all technical and logistical planning and implementation of the research.

- On April 18, 2017, several staff members from AED conducted an orientation for 16 individuals delegated by the participating school directorates to serve as test coordinators. Each coordinator received a list of the selected students and schools in his or her district. Emphasis was placed on

---

[1] Statistical tables for all variables used in the analysis can be found in Annexes 1 to 5 at the end of the report.

following a strict testing protocol to ensure the integrity of the test administration. Following this meeting, each coordinator delegated employees from each participating school to serve as test administrators.

- On the day of the test administration, teachers of students taking the exams were prohibited from entering the testing rooms. Once completed, the test booklets were resealed in secure envelopes and signed by both the school principal and the test administrator. Each district-level test coordinator collected the envelopes and mailed them to AED.

- LTD prepared an Excel codebook that facilitated data entry and merging of data from the exam booklets and questionnaires for the Classroom Context Survey. Five short-term technical assistants hired by LTD worked at AED to score and then re-score 10% of the exam booklets for validation. After the assistants finished keying in the two sets of data, they validated their work by repeating data entry for 10% of the booklets and questionnaires.

- Finally, AED's senior staff filtered and cleaned the datasets for accuracy and then generated the descriptive statistics for the exams by subject, cohort (including non-LTD schools), and gender. Before conducting its statistical analyses, the M&E staff of LTD re-screened the AED datafile for verification. No errors were found.

Limitations

Although the method of randomized controlled trial (RCT) is considered the gold standard for research design, it was fundamentally impossible for budgetary and practical reasons for LTD to *randomize the allocation* of students into "treatment" and "control" groups, and to ensure "blindness" by guaranteeing that the exam administrators would not know which students were in treatment or control groups. For this reason, LTD opted for the next best thing by using a quasi-experimental design that allowed for the randomized selection of students in treatment and control groups. Furthermore, in selecting the control schools, LTD was mindful of contextual factors (e.g., gender of school, size, and location) so as to eliminate selection bias as much as possible (Bamberger, Rugh, & Mabry, 2011).

The methodology of the present study improves on two earlier attempts to use student achievement data to assess program impact. In the previous iterations done in 2014 and 2015, each cohort of teachers had only just completed their yearlong training when their students were tested. Thus, there was insufficient time for teachers to adapt and develop their new knowledge and skills in learner-centered instruction to their classroom settings, a gradual process that usually requires several years (Guskey, 2016, p. 8). In light of this limitation, this current study, as described above, took advantage of the fact that there is now a large pool of teachers who, over three successive cohorts since 2013, have completed LTD's Teacher Diploma Program. In other words, teachers of Cohort 1 have had three years of post-training experience to adapt and develop their competencies, while Cohort 2 teachers have had two years, and Cohort 3 one year.

Results

First Hypothesis: Scores on standardized exams for LTD students will be higher than scores of students taught by non-LTD teachers.

The study found that scores for LTD students on the four exams measured higher than scores earned by students of non-LTD teachers, as seen in **Table 7**.

While the variance of the means between LTD and non-LTD students is not statistically significant for the Arabic, Biology, and Math exams, which have a p value > .05, the p value of .052 for the English exam is close enough to indicate a trend toward statistical significance. Furthermore, results from the Classroom Context Survey (discussed below) offer additional evidence that point to LTD's positive impact on student learning.

Table 7. Subject exam scores and t-tests

SUBJECT	SCORE of 100		% Difference	t-test for Equality of Means
	LTD	Non-LTD		Sig. (2-tailed)
Arabic	44.95	47.39	5.28%	.071
Biology	35.17	36.93	4.88%	.146
English	40.58	42.94	5.65%	.052
Math	27.82	29.63	6.30%	.162

The exam items tested not only students' content knowledge, but also their thinking skills. Each test item was designed to measure one of three cognitive domains: knowing, applying, and reasoning. The analysis revealed that LTD students scored higher than their non-LTD counterparts did in each of these domains for Arabic, Math, and Science (**Table 9**).

Table 8. Results of analysis of cognitive domains

Cognitive Domain	Arabic			Math			Science		
	LTD	Non-LTD	Difference	LTD	Non-LTD	Difference	LTD	Non-LTD	Difference
Knowledge	47.17	43.52	8.04%	44.80	43.60	2.71%	51.14	51.56	0.83%
Application	54.25	52.13	3.98%	23.65	21.11	11.35%	39.89	38.40	3.81%
Reasoning	41.25	39.25	4.97%	15.42	13.98	9.78%	22.14	18.28	19.08%

The results of the Classroom Context Survey (**Table 9**) indicate that students of LTD teachers rated their classrooms as more learner-centered compared to their non-LTD peers. The Classroom Context Survey is a 20-item scale that measures the level of engagement of students in a learner-centered classroom. Furthermore, the differences in the means between the two groups (non-LTD and LTD) is statistically significant (p ≤ .05) with the exception of English, where the two means are nearly the same.

Table 9. Exam scores for LTD and Non-LTD students by subject, with results of t-tests

Classroom Context by Subject	SCHOOL TYPE		t-test for Equality of Means
	LTD	NON LTD	Sig. (2-tailed)
	Mean	Mean	
Arabic	3.87	3.75	0.010
Biology	3.84	3.76	0.023
English	3.87	3.82	0.168
Math	3.42	3.33	0.005

Taken together, the results from both the exams and the contextual survey suggest that LTD teachers, as we would expect, appear to be creating learner-centered classroom environments and are having a more positive influence on student learning than their non-LTD peers. Additional statistical testing corroborates this claim even further, as explained by the use of correlations.

The correlation between the Classroom Context Survey and exam results was investigated using the Pearson product-moment correlation coefficient. **Table 10** shows that for LTD schools, there was a moderate positive correlation between the two variables, $r = .373$, $n = 2835$, $p < .01$. For the non-LTD schools, by contrast, there was a weaker positive correlation between the two variables, $r = .273$, $n = 943$, $p < .01$.[2]

In other words, as a classroom becomes more learner-centered, student learning (measured by exam results) appears to improve correspondingly. In this regard, LTD teachers seem to be more effective in creating learner-centered classrooms than their non-LTD peers.

Table 10. Correlations between classroom context and student achievement for all subjects

SCHOOL TYPE			CONTEXT	SCOREOF100
NON LTD	CONTEXT	Pearson Correlation	1	.273**
		Sig. (2-tailed)		.000
		N	943	943
	SCOREOF100	Pearson Correlation	.273**	1
		Sig. (2-tailed)	.000	
		N	943	991
LTD	CONTEXT	Pearson Correlation	1	.373**
		Sig. (2-tailed)		.000
		N	2835	2835
	SCOREOF100	Pearson Correlation	.373**	1
		Sig. (2-tailed)	.000	
		N	2835	3055

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

---

[2] Preliminary analyses were performed to ensure no violation of the assumptions of normality or linearity.

We see this same trend occurring in the correlation between the Classroom Context Survey and the results on the individual exams (**Table 11**). The correlations are higher for LTD students relative to non-LTD students in three out of the four subjects.

Table 11. Pearson Correlation between Classroom Context Survey and exam results

Subject	Non-LTD	LTD
Arabic	.154*	.367**
Biology	.067	.290**
English	.237**	.314**
Math	.290**	.257**

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Before we consider the second hypothesis, it is worth asking whether gender had any influence on the correlations between exams and the Classroom Context Survey. As reported above in **Table 10**, we find there is a moderate positive correlation between the exams and classroom context. At the same time, however, the results for girls on the exams and the Classroom Context Survey are higher overall than those of boys (**Table 12**). Thus, could the correlation we saw between exams and Classroom Engagement Survey be the result of gender?

To answer this question, we performed a statistical test that "controls" for gender (i.e., removes it statistically for the purpose of analysis). As seen in **Table 13**, the grey arrow points to the correlation before controlling for gender (.353) and the black arrow shows the correlation after controlling for gender (.332). Since the difference between the two correlations is small—a mere 6%--we can safely conclude that gender was not a confounding factor.

Table 12. Exam and Classroom Engagement results by gender (for all subjects)

	SCHOOL COHORT							
	NON LTD		COHORT 1		COHORT 2		COHORT 3	
	STUDENT GENDER		STUDENT GENDER		STUDENT GENDER		STUDENT GENDER	
	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE
	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean
SCOREOF100 (all exams)	34.46	39.61	36.19	45.28	32.59	44.89	33.02	42.53
CONTEXT	3.60	3.71	3.73	3.86	3.63	3.85	3.67	3.80

Table 13. Partial correlation, controlling for gender

			ENGAGEMENT	SCOREOF100	GENDER
Not controlling for GENDER	CONTEXT	Correlation	1.000	.353	.153
		Significance (2-tailed)	.	.000	.000
		df	0	3775	3775
	SCOREOF100	Correlation	.353	1.000	.221
		Significance (2-tailed)	.000	.	.000
		df	3775	0	3775
Controlling for GENDER	CONTEXT	Correlation	1.000	.332	
		Significance (2-tailed)	.	.000	
		df	0	3774	
	SCOREOF100	Correlation	.332	1.000	
		Significance (2-tailed)	.000	.	
		df	3774	0	

Second Hypothesis: **Students taught by LTD-trained teachers will rate their classroom context as more learner-centered compared to students of non-LTD teachers.**

Logically, we would predict that Cohort 1 classrooms would score higher on the Classroom Context Survey on the assumption that, over time, LTD teachers would develop and improve their capacity to create learner-centered classrooms. As seen in **Table 14**, Cohort 1 does in fact score better than the other two; moreover, the variance in the means between three groups (p = .057) is close enough to the conventional cut off point of p ≤ .05 to indicate a trend toward statistical significance.

Table 14. Results of Classroom Context Survey across all cohorts

Classroom Context Survey			
COHORT 1 N=828	**COHORT 2 N=1194**	**COHORT 3 N=813**	**ANOVA**
Mean	Mean	Mean	Sig.
3.81	3.76	3.76	.057

Furthermore, when we compare the results of the Classroom Context Survey by subject and cohort, as shown in **Table 15**, a similar trend emerges. Scores for Cohort 1 are higher than are those for the other two cohorts in three of the four subjects, Arabic, Biology, and English.

Classroom Context Survey				
Subject	COHORT 1	COHORT 2	COHORT 3	ANOVA
	Mean	Mean	Mean	Sig.
Arabic	3.89	3.86	3.86	.644
Biology	3.89	3.83	3.80	.093
English	3.98	3.80	3.89	.001
Math	3.41	3.41	3.44	.648

Third Hypothesis: Exam scores of students taught by teachers who graduated from Cohort 1 of LTD will be higher than the scores of students whose teachers graduated from Cohort 2 and Cohort 3.

As explained earlier, we hypothesized that since teachers of Cohort 1 had the most time—three years—to adapt and develop their knowledge and skills acquired through LTD professional development, we would expect their students to perform better than students of teachers from the later cohorts. **Table 16** shows encouraging evidence to support this assumption.

Cohort 1 students scored higher in English and Math than students from Cohorts 2 and 3. We also find that students of Cohort 2 teachers—who had two years to adapt and develop their LTD skills—scored higher than students of Cohort 3 did across all four exams. Furthermore, in no instance were the scores of Cohort 3 higher than those of Cohort 1, though the means for the Arabic exam for Cohort 1 and 3 were identical.

Table 16. Exam scores by subject and cohort

SUBJECT	SCHOOL COHORT			ANOVA
	COHORT 1	COHORT 2	COHORT 3	
	Mean	Mean	Mean	Sig.
Arabic	46.79	48.24	46.79	0.505
Biology	37.20	37.42	35.79	0.550
English	45.64	41.13	43.29	0.027
Math	34.72	27.57	27.12	0.000

Conclusions

The multiple sources of evidence presented in the above analysis of data from the exam scores and the Classroom Context Survey allow us to conclude that LTD's model of school-based reform and teacher professional development is contributing to improved student learning.   Compared to students of non-LTD teachers, students of LTD teachers scored higher on standardized tests of achievement in Arabic, English, mathematics, and science/biology. Furthermore, LTD students scored higher than did their non-LTD peers on test items measuring the cognitive skills of knowing, applying, and reasoning across the subjects of Arabic, mathematics, and science/biology.

Even though the statistical variance in the means of the two groups of students were significant for English only, the fact that LTD students across all four subjects rated their classrooms as more "learner-centered" than their non-LTD peers suggests that the differences between the two groups are meaningful, a conclusion supported by the fact that the correlations between exam scores and the Classroom Context Survey are stronger for LTD classrooms.

These differences are important because learner-centered instruction, as reported in the research literature, creates a classroom environment conducive to positive affective and cognitive development in students which, in turn, are associated with better learning and improved student outcomes (Cornelius-White, 2007; Polly et al., 2015; Salinas & Garr, 2009). At the same time, however, learner-centered classrooms are only possible with the support of school leadership committed to comprehensive learner-centered change in classroom instruction (Hargreaves & Fullan, 2013; McCombs & Quiat, 2002; Sheppard & Brown, 2014).

To sum up, we believe there are two key reasons why the results of this study point to LTD's positive contribution to improved student learning both inside and outside the classroom: the structure of LTD's learner-centered model of teacher professional development, and, LTD's systemic model of educational leadership.

Learner-Centered Professional Development: LTD's model of teacher professional development is a key reason why the study points to LTD's positive contribution to improved student learning. Unlike the typical one-off workshop or training that may last a few days or less, LTD teachers undertook a yearlong experience in which they were clustered into school-based professional learning communities that met in monthly face-to-face sessions and bi-weekly learning circles. LTD and its partner NIET (National Institute for Educational Training) employed an inquiry-based learning approach that allowed teachers to work independently and in small groups to identify problems of practice, conduct action research, and reflect critically on their teaching practices. Moreover, and in addition to the ICT resources that LTD supplied to schools and classrooms, each teacher received a laptop to enhance their capacity to integrate technology into the teaching and learning process.

In sum, the LTD's learner-centered model for professional development empowers teachers to work both independently and in professional learning communities to strengthen their competencies, self-confidence, and self-efficacy as professional educators. In this way, LTD optimizes opportunities for teachers to engage in ongoing professional development and to take greater responsibility for their own growth as effective teachers. As effective teachers, they are the most important factor contributing to

student achievement (Barber & Mourshed, 2007; Goddard et al., 2000; Gutierrez et al., 2016; Heck, 2009; Kyriakides et al., 2013; Vescio et al., 2008)

Systemic Model of Educational Leadership: LTD's positive contribution to improved student learning is also a result of the program's strategy of aligning policies, structures, and processes for improving student learning that extend from the classroom to the school to the local district and to the central Ministry of Education and Higher Education (MoEHE). Informed by several decades of research on best practices in school-based reform (DuFour & Marzano, 2011; Eaker & Keating, 2011; Huber & Conway, 2015; Lezotte & Bancroft, 1985; Sammons, 1995), LTD takes a systems approach to effective schools that empowers school principals to enact a shared-leadership and learner-centered model for school improvement.

Guided by the MoEHE's Standards for Effective Schools, principals lead a School Improvement Team (SIT) of key stakeholders—teachers, parents, support staff and often students—in developing a school's annual School Improvement Plan (SIP). The District Leadership Team (DLT), a specialized unit created by LTD and formally sanctioned by the MoEHE, provides technical support to facilitate the flow of information and resources to schools. Furthermore, LTD's yearlong Leadership Diploma training emphasizes the role principals—working both independently and in communities of practice—to provide supportive instructional supervision to teachers. In sum, LTD's alignment of policies, structures, and processes creates overlapping and sustainable systems of support that enable the success of LTD's model of teacher professional development.

Annex 1: Categorical Variables

TYPE

		Value	Count	Percent
Standard Attributes	Label	SCHOOL TYPE		
Valid Values	0	NON LTD	991	24.5%
	1	LTD	3055	75.5%

Cohort

		Value	Count	Percent
Standard Attributes	Label	SCHOOL COHORT		
Valid Values	0	NON LTD	991	24.5%
	1	COHORT 1	892	22.0%
	2	COHORT 2	1302	32.2%
	3	COHORT 3	861	21.3%

SUBJECT

		Value	Count	Percent
Standard Attributes	Label	SUBJECT		
Valid Values	1	Arabic	1239	30.6%
	2	Biology	1056	26.1%
	3	English	917	22.7%
	4	Math	834	20.6%

STUGENDER

		Value	Count	Percent
Standard Attributes	Label	STUDENT GENDER		
Valid Values	1	MALE	1655	40.9%
	2	FEMALE	2390	59.1%
Missing Values	System		1	0.0%

Annex 2: Dependent Variables

Means of Exams

Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
SCOREOF100	991	.00	94.12	37.1787	17.69765
Valid N (listwise)	991				

a. SCHOOL COHORT = NON LTD

Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
SCOREOF100	892	.00	94.12	41.4588	19.52457
Valid N (listwise)	892				

a. SCHOOL COHORT = COHORT 1

Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
SCOREOF100	1302	.00	93.10	39.9512	19.89339
Valid N (listwise)	1302				

a. SCHOOL COHORT = COHORT 2

Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
SCOREOF100	861	.00	88.61	39.3519	19.08783
Valid N (listwise)	861				

a. SCHOOL COHORT = COHORT 3

Annex 3: Independent Variables

Means of Classroom Context Survey by School Type

Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
Classroom Engagement	943	1.65	5.00	3.6634	.51287
Valid N (listwise)	943				

a. SCHOOL COHORT = NON LTD

Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
Classroom Engagement	828	1.00	5.00	3.8095	.48249
Valid N (listwise)	828				

a. SCHOOL COHORT = COHORT 1

Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
Classroom Engagement	1194	1.50	4.95	3.7589	.49981
Valid N (listwise)	1194				

a. SCHOOL COHORT = COHORT 2

Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
Classroom Engagement	813	1.75	5.00	3.7630	.50377
Valid N (listwise)	813				

a. SCHOOL COHORT = COHORT 3

Means of Classroom Context Survey by Exam (Non-LTD Schools)

Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
Classroom Engagement	227	1.80	4.75	3.7452	.53612
Valid N (listwise)	227				

a. SUBJECT = Arabic

Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
Classroom Engagement	233	2.05	4.79	3.7593	.46940
Valid N (listwise)	233				

a. SUBJECT = Biology

Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
Classroom Engagement	244	2.00	5.00	3.8225	.46313
Valid N (listwise)	244				

a. SUBJECT = English

Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
Classroom Engagement	239	1.65	4.20	3.3300	.42592
Valid N (listwise)	239				

a. SUBJECT = Math

Means of Classroom Context Survey by Exam (LTD Schools)

Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
Classroom Engagement	908	1.75	5.00	3.8671	.48478
Valid N (listwise)	908				

a. SUBJECT = Arabic

Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
Classroom Engagement	784	2.00	5.00	3.8385	.46475
Valid N (listwise)	784				

a. SUBJECT = Biology

Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
Classroom Engagement	603	1.50	5.00	3.8731	.49239
Valid N (listwise)	603				

a. SUBJECT = English

Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
Classroom Engagement	540	1.00	4.20	3.4176	.39515
Valid N (listwise)	540				

a. SUBJECT = Math

Annex 4: Independent-Samples t-Tests

Exams: LTD and Non-LTD

Independent Samples Test

SUBJECT			Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
			F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Arabic	SCORE_100	Equal variances assumed	2.227	.136	-1.808	1237	.071	-2.44015	1.34970	-5.08810	.20781
		Equal variances not assumed			-1.897	399.673	.059	-2.44015	1.28651	-4.96932	.08903
Biology	SCORE_100	Equal variances assumed	6.663	.010	-1.369	1054	.171	-1.76168	1.28724	-4.28753	.76417
		Equal variances not assumed			-1.457	422.359	.146	-1.76168	1.20948	-4.13904	.61568
English	SCORE_100	Equal variances assumed	8.631	.003	-1.838	915	.066	-2.35934	1.28393	-4.87913	.16045
		Equal variances not assumed			-1.946	538.712	.052	-2.35934	1.21243	-4.74102	.02233
Math	SCORE_100	Equal variances assumed	5.009	.025	-1.343	832	.180	-1.80820	1.34636	-4.45086	.83447
		Equal variances not assumed			-1.402	513.581	.162	-1.80820	1.28996	-4.34244	.72604

Annex 5: One-Way ANOVA

Classroom Engagement Survey (Cohorts 1, 2, and 3)

ANOVA

Classroom Engagement

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.414	2	.707	2.874	.057
Within Groups	696.614	2832	.246		
Total	698.028	2834			

References Cited

Bamberger, M., Rugh, J., & Mabry, L. (2011). *RealWorld evaluation: Working under budget, time, data, and political constraints*: sage.

Barber, M., & Mourshed, M. (2007). *How the world's best-performing schools systems come out on top*: McKinsey & Company.

Cornelius-White, J. (2007). Learner-Centered Teacher-Student Relationships Are Effective: A Meta-Analysis. *Review of Educational Research, 77*(1), 113-143.

Cristillo, L., Iter, N., & Assali, A. (2016). Sustainable Leadership: Impact of an Innovative Leadership Development Program for School Principals in the West Bank and Gaza. *American Journal of Educational Research, 4*(2A), 37-42.

DuFour, R., & Marzano, R. J. (2011). *Leaders of learning: How district, school, and classroom leaders improve student achievement*: Solution Tree Press.

Eaker, R., & Keating, J. (2011). *Every School, Every Team, Every Classroom: District Leadership for Growing Professional Learning Communities at Work TM*: Solution Tree Press.

Goddard, R. D., Hoy, W. K., & Hoy, A. W. (2000). Collective Teacher Efficacy: Its Meaning, Measure, and Impact on Student Achievement. *American Educational Research Journal, 37*(2), 479-507. doi:10.3102/00028312037002479

Guskey, T. (2016). Gauge impact with 5 levels of data. *SMEC2016 Organising Committee*, 6.

Gutierrez, I., Weinberger, G., & Engberg, J. (2016). *Improving Teaching Effectiveness: Impact on Student Outcomes*: RAND Corporation.

Hargreaves, A., & Fullan, M. (2013). THE POWER of PROFESSIONAL CAPITAL. *Journal of staff development, 34*(3), 36-39.

Heck, R. H. (2009). Teacher effectiveness and student achievement. *Journal of Educational Administration, 47*(2), 227-249. doi:10.1108/09578230910941066

Huber, D. J., & Conway, J. M. (2015). THE EFFECT OF SCHOOL IMPROVEMENT PLANNING ON STUDENT ACHIEVEMENT. *Planning and Changing, 46*(1/2), 56-70.

Khales, B. (2016). The Impact of a Teacher Training Program on Mathematics Teaching Methodologies: Using Student-centered Learning. *American Journal of Educational Research, 4*(14), 992-998.

Kyriakides, L., Christoforou, C., & Charalambous, C. Y. (2013). What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective teaching. *Teaching and Teacher Education, 36*, 143-152. doi:http://dx.doi.org/10.1016/j.tate.2013.07.010

Lezotte, L. W., & Bancroft, B. A. (1985). School Improvement Based on Effective Schools Research: A Promising Approach for Economically Disadvantaged and Minority Students. *The Journal of Negro Education, 54*(3), 301-312. doi:10.2307/2295066

McCombs, B. L., & Quiat, M. (2002). What Makes a Comprehensive School Reform Model Learner Centered? *Urban Education, 37*(4), 476-496. doi:doi:10.1177/0042085902374002

MoEHE. (2008). *Education Development Strategic Plan: 2008-2012 Towards Quality Education for Development*. Ramallah.

MoEHE. (2014). *the West Bank and Gaza 2020: a learning nation, Education Development Strategic Plan, EDSP 2014-2019*. Ramallah: Ministry of Education and Higher Education.

Polly, D., McGee, J., Wang, C., Martin, C., Lambert, R., & Pugalee, D. K. (2015). Linking professional development, teacher outcomes, and student achievement: The case of a learner-centered mathematics program for elementary school teachers. *International Journal of Educational Research, 72*, 26-37. doi:http://dx.doi.org/10.1016/j.ijer.2015.04.002

Salinas, M. F., & Garr, J. (2009). Effect of learner-centered education on the academic outcomes of minority groups. *Journal of Instructional Psychology, 36*, 226+.

Sammons, P. (1995). *Key characteristics of effective schools: A review of school effectiveness research*: ERIC.

Sheppard, B., & Brown, J. (2014). Leadership for a new vision of public school classrooms. *Journal of Educational Administration, 52*(1), 84-96. doi:http://dx.doi.org/10.1108/JEA-03-2012-0027

Shinn, C. (2012). Teacher education reform in the West Bank and Gaza: Policy challenges amid donor expectations. *Comparative Education Review, 56*(4), 608-633.

Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education, 24*(1), 80-91. doi:http://dx.doi.org/10.1016/j.tate.2007.01.004