EdData II

Education Data for Decision Making (EdData II): National Early Grade Literacy and Numeracy Survey–Jordan

Responses to MoE Comments on the Intervention Impact Analysis Report

EdData II Technical and Managerial Assistance, Task Order Number 16
Contract Number: AID-278-BC-00019
March 2015

This publication was produced for review by the United States Agency for International Development. It was prepared by RTI International.

Education Data for Decision Making (EdData II): National Early Grade Literacy and Numeracy Survey–Jordan

Responses to MoE Comments on the Intervention Impact Analysis Report

Second draft

EdData II
Task Order No. 16
Period Ending November 30, 2015

Prepared for
USAID/Jordan
Angie Haddad, COR

Prepared by Aarnout Brombacher
RTI International
3040 Cornwallis Road
Post Office Box 12194
Research Triangle Park, NC 27709-2194

RTI International is a trade name of Research Triangle Institute.

The authors' views expressed in this publication do not necessarily reflect the views of the United States Agency for International Development or the United States Government

Comments on the National Early Grade literacy and Numeracy Intervention Pilot Report

The Ministry of Education would like to thank the National Survey team for the tremendous efforts made during all the national survey stages. The following are the key comments on the first draft of the report the DCU received from MOE directorates and NCHRD.

National Center for Human Resources Development

The report has addressed comprehensive key issues of education in Jordan, in addition to attempting to measure the impact using the pilot design. The following comments aim to improve future quality of the study.

1. The study sample should cover all schools in the country, as well as all authorities (the UNRWA, military schools and private schools). **The study included the MoE schools; however, it did not take into account the school location variable (urban and rural areas).**

الإجابة: هذه ملاحظة مهمة ومثيرة للاهتمام. لم تشتمل عينة الدراسة على مدارس أخرى غير مدارس وزارة التربية والتعليم لأن الدراسة انبثقت من المسح الوطني لعام 2012، وفي المسح الوطني لعام 2012 تقرر استهداف مدارس وزارة التربية والتعليم فقط وعدم استخدام الخصائص الحضرية والريفية كمعيار لأخذ العينات.

Response: This is an important and interesting observation. The study sample did not include schools other than MoE schools because the study evolved from the 2012 National Survey, and in the 2012 National Survey, it was decided to target MoE schools only and not to use urban and rural characteristics as a sampling criteria.

2. The study design needs a higher control level, as the treatment group was purposively selected, whereas it should be randomly distributed to the treatment and control groups.

الإجابة: إن طريقة أخذ العينات الهادفة هي من الأساليب الصحيحة في عملية أخذ العينات لموضوع بحث من هذا النوع. فمن خلال طريقة أخذ العينات الهادفة، يتم اختيار مواضيع البحث (المدارس) بسبب بعض الخصائص. في حالة هذا الموضوع البحثي لمشروع التدخل، اتاحت طريقة أخذ العينات الهادفة وجود مجموعات فرعية معينة وسهلت إجراء مقارنات بين هذه المجموعات الفرعية. وقد كانت المجموعات الفرعية في هذه الحالة هي مناطق مشروع التدخل التي يوجد لها مشرف واحد ومناطق مشروع التدخل التي لها مشرفين اثنين أو أكثر، وفي جميع الحالات مدرستين على الأقل لكل مشرف.

Response: Purposive sampling is a valid sampling technique for a research activity of this nature. With purposive sampling, subjects (schools) are selected because of some characteristic. In the case of this intervention research activity, purposive sampling allowed for particular subgroups and facilitated comparisons between these subgroups. The subgroups in this case were intervention districts with a single supervisor and intervention districts with two or more supervisors, and in all cases at least two schools per supervisor.

3. It would be better if both groups (treatment and control) are equal, as the current distribution lacks this feature

الإجابة: على الرغم من أن هذا الأمر صحيح عادة، إلا أنه في هذه الدراسة كانت الموارد الوحيدة التي يمكن أن يتم شملها هي 45 مدرسة كحد أقصى في المجموعة التجريبية. في المسح الوطني لعام 2014 ودراسة نهاية المدة، تم اتخاذ قرار الإبقاء على جميع المدارس التي تم تضمينها في المسح الوطني لعام 2012، لتمكين إجراء مقارنة أكثر فعالية. لهذا السبب، كان عدد مدارس المجموعة الضابطة كبيراً على نحو غير معهود.

Response: While typically this is correct, in this study we only had the resources to include, at most, 45 schools in the treatment group. In the 2014 National Survey and Endline Study, the decision was made to retain all the schools of the 2012 National Survey, to enable more effective comparison. For this reason, the number of control schools was uncharacteristically high.

4. It is suggested that the Project/Intervention is reconsidered to be dependent from the measurement tool in order to be rolled out in the school when it succeeds, especially that the findings of the study showed that the Project/Intervention was successful in the girls' schools, while it had no positive impact in the boys schools.

الإجابة: لقد كان الغرض من إجراء مشروع التدخل هو البحث فيما إذا كانت "الممارسة اليومية للمهارات الأساسية من خلال أنشطة متعمدة, ومنظمة، ومناسبة تنموياً تدعم الأطفال حتى يتمكنوا من القراءة مع الفهم وإجراء العمليات الحسابية مع الفهم". وهذا يعني أن الموضوع البحثي المراد هو التقصي فيما إذا كان من الممكن استهداف مهارات القراءة والحساب التي حددها المسح الوطني لعام 2012 بأنها ليست متطورة بما فيه الكفاية. لهذا السبب، كان من المناسب استخدام أدوات القياس نفسها التي تم استخدامها في المسح الوطني لعام 2014 وفي دراسة نهاية المدة. وهذا يعني بأن الباحثون استكشفوا أيضاً خصائص أداء مدارس المجموعة التجريبية والمجموعة الضابطة في التقييم الوطني لوزارة التربية والتعليم، التي أجريت في نهاية العام الدراسي 2014/2013. فقد أشار تحليل الأداء للتقييم الوطني لوزارة التربية والتعليم أن: متوسط نسبة النجاح في امتحان اللغة العربية كان 75% للمدارس المشمولة في المسح الوطني. وأظهرت مدارس المجموعة التجريبية نسبة نجاح أعلى بمقدار 7% من مدارس المجموعة الضابطة (79% لمدارس المجموعة التجريبية مقابل 72% لمدارس المجموعة الضابطة). في تقييم مادة الحساب، كان متوسط نسبة النجاح هو 65%. وتفوقت مدارس المجموعة التجريبية على مدارس المجموعة الضابطة بحوالي 10.5% (71% لمدارس المجموعة التجريبية مقابل 60.6% لمدارس المجموعة الضابطة). وكانت الفروقات في اللغة العربية والحساب ذات دلالة إحصائية عند قيمة P < مستوى 0.05. وهذه النتائج مشجعة، لأنها تشير إلى أنه يمكن مشاهدة أثر مشروع التدخل ليس فقط في أدوات المسح (البيانات التعليمية) وإنما أيضاً في تقييم لا علاقة له بذلك (وزارة التربية والتعليم).

Response: The purpose of the intervention was to research whether or not “daily practice of foundational skills through deliberate, structured, and developmentally appropriate activities support children to be able to read with comprehension and do mathematics with understanding.” That is, the research activity wanted to investigate whether it was possible to target the reading and mathematics skills that the 2012 National Survey had identified as underdeveloped. For this reason, it was appropriate to use the same measurement tools in the 2014 National Survey and Endline Study. That said, the researchers also explored the performance characteristics of the treatment and control schools on the MoE National Assessment, conducted at the end of the 2013/2014 school year. The analysis of the performance on the MoE National Assessment indicated that: the average passing rate for the Arabic language test was 75% for the schools included in the National Survey. Treatment schools showed a 7% higher passing rate than control schools (79% treatment schools versus 72% control schools). For the mathematics assessment, the average passing rate was 65%. Treatment schools outperformed control schools by approximately 10.5% (71% treatment schools versus 60.6% control schools). The differences both for Arabic language and for

mathematics are statistically significant at the $p < 0.05$ level. These results are encouraging, as they indicate that the impact of the intervention was seen not only on the survey tools (EdData) but also on an unrelated assessment (MoE).

5. The training program accompanying the Intervention was not similarly implemented in terms of the time allocated for supervisors and teachers.

الإجابة: من الصحيح أن نقول بأن مشروع التدخل لم يتم تنفيذه على نحو متماثل في جميع المدارس، حيث قام بعض المشرفين بتقديم دعماً أكثر من غيرهم. وقد كان هذا متوقعاً، وأتاح للباحثين استكشاف سؤال البحث الثاني: "ما هي الشروط التي تساعد المعلمين على تنفيذ الإجراءات الروتينية اليومية والأنشطة المرتبطة بها بثقة وإخلاصٍ ؟" من حيث صلته بتكرار دعم المشرف. وأظهرت النتائج بوضوح أن "إن الزيارة بواقع 16 مرة أو أكثر خلال فترة مشروع التدخل يرتبط مع زيادة مقدارها 27% في النسبة المئوية للطلبة الذين يجيدون القراءة وزيادة مقدارها 15% في النسبة المئوية للطلبة الذين يجيدون الحساب في صفوف هؤلاء المعلمين".

Response: It is correct to say that the intervention was not similarly implemented in all schools, with some supervisors providing more support than others. This was anticipated, and allowed the researchers to explore the second research question: “What are the conditions that help teachers to implement the daily routine and the associated activities with fidelity and confidence?” as it relates to the frequency of supervisor support. The results clearly showed that “Being visited as frequently as 16 or more times over the period of the intervention was associated with a 27% increase in the percentage of readers and a 15% increase in the percentage of mathematicians in the classrooms of those teachers.”

Development Coordination Unit

1. How can the improvement in students' performance be attributed to the interventions activities rather than to other factors such as the close follow up that were undertaken by the supervisors for the targeted groups, in addition to that the school management was not included in the study so how we can be assured that the improvement is not resulted from school management factors.

الإجابة: من الصحيح تماماً الإشارة إلى وجود مجموعة من العوامل التي ساهمت في أداء مدارس المجموعة التجريبية في المسح الوطني لعام 2014 وفي دراسة نهاية المدة. بالنسبة لدور المشرفين، كان الموضوع البحثي قادراً على قياس مدى مشاركتهم (انظر الإجابة السابقة أعلاه). أما بالنسبة للإدارة المدرسية، لم يتمكن موضوع البحث من استكشاف مدى أو أثر مشاركة الإدارة المدرسية. وهذا يعني بأنه لا شك في أن الإدارة المدرسية يجب أن تكون قد لعبت دوراً، ودور أكبر في بعض المدارس بالمقارنة مع غيرها. وحتى نكون منصفين، من غير المرجح أن المدى الكامل لأثر مشروع التدخل يمكن أن يُعزى إلى دور الإدارة المدرسية والمشرفين فقط. هناك دليل حقيقي جداً أن مواد ومنهجية مشروع التدخل قد ساهمت أيضاً في تحسين النتائج في مدارس المجموعة التجريبية.

Response: It is quite correct to indicate that there were a range of factors that contributed to the performance of the treatment schools on the 2014 National Survey and Endline Study. For the role of supervisors, the research activity was able to quantify the extent of their involvement (see previous response above). For school management, the research activity did not explore either the extent or the impact of school management's involvement. That said, there is no doubt that the school management must have played a role, and a greater role in some schools than in others. In fairness, it is unlikely that the full extent of the intervention impact can be attributed to the role of the school management and supervisors only. There is very real evidence that the materials and methodology of the intervention also contributed to the improved results in the treatment schools.

2. The findings of the 2014 National Survey indicated that there is a statistically significant difference of gender for the favor of females, while this difference was not statistically significant in the National Survey/2012. Moreover, this finding was not analyzed in the 2014 report to highlight the reasons behind the differences between 2012 results and 2014.

الإجابة: يحتاج هذا السؤال، بشكل عام، إلى المزيد من البحث لأن المسألة شائكة ومعقدة. مع ذلك، فإن تقرير عام 2014 لا يشير إلى الفروق في الأداء حسب النوع الاجتماعي عموما، وإنما إلى الفروق حسب النوع الاجتماعي في الاستجابة لمشروع لتدخل. في حالة مشروع التدخل، فقد كانت المشاهدة هي أن الإناث قد استفدن من مشروع التدخل أكثر من الذكور. وتشير النتائج أيضاً إلى أن الأطفال في مدارس الإناث قد استفادوا من مشروع التدخل أكثر من الأطفال في المدارس المختلطة وأن الأطفال في مدارس الذكور كانوا أقل المستفيدين.

Response: This is a question that, in general, needs more research because the issue is complex. However, the 2014 report does not indicate a difference in performance by gender in general, but rather the difference by gender in response to the intervention. In the case of the intervention, the observation is that girls appear to have benefited more from the intervention than boys did. The results also suggest that children in all-girls schools benefited more from the intervention than children in mixed schools and that children in all-boys schools appear to have benefited the least.

3. Regarding the assessment of students' skills in Math (EGMA), it is noted that some of the sub skills in the 2012 National Survey 2012 were time-limited, while they were not time limited in the 2014 National Survey. This raises several questions regarding attributing the effect of the intervention considering the difference related to the time factor, so this requires further articulation.

الإجابة: هذه ملاحظة هامة. من الواضح أنه لهذا السبب، وكذلك للحد من تأثير التحيز في الاختيار، قد تم تحديد تأثير التدخل باستخدام منهجية الفرق في الفروقات (DID). فمن خلال مقارنة (1) متوسط التغير مع الزمن في متغير النتيجة للمجموعة التجريبية مع (2) متوسط التغير مع الزمن في النتيجة للمجموعة الضابطة، يتم احتساب دور التغييرات في مهمة تقييم و/ أو التحيز غير المقصود في الاختيار.

Response: This is an important observation. It is expressly for this reason, as well as to limit the effect of selection bias, that the intervention impact was determined using difference-in-differences (DID) methodology. By comparing (1) the average change over time in the outcome variable for the treatment group with (2) the average change over time in the outcome for the control group, the role of changes in the assessment task and or unintended selection bias are accounted for.

4. The executive summary should be more focused on the results and findings of the study giving that the details are already explained in the related sections.

الإجابة: تمت إضافة ملخص تنفيذي للتقرير على المسودة الأولى التي تستند إليها هذه التعليقات.

Response: An executive summary has been added to the report since the first draft on which these comments are based.

5. We assume that the second draft will include the recommendations as discussed in the workshop so as to be considered by the Ministry.

الإجابة: تم إضافة هذه التوصيات التي تم وضعها كجزء من أنشطة ورشة عمل نشر التقرير إلى المسودة الأولى التي تستند إليها هذه التعليقات.

Response: The recommendations developed as a part of the dissemination workshop activities have been added to the report since the first draft on which these comments are based.

6. It would be much helpful to indicate in a summary the key indicators resulted from 2014 study, and identify which indicators could be considered baseline indicators so as to measure the improvement resulted from the future investment and initiatives, the MOE has already started using new curriculum for instance, and will also start a new initiative supported by USAID (RAMP).

الإجابة: لقد تم إضافة مقاييس الأداء والأهداف المخطط لها لمدة خمس سنوات، والتي تم وضعها كجزء من أنشطة ورشة عمل نشر التقرير، إلى المسودة الأولى التي تستند إليها هذه التعليقات. نعتقد بأن مقياس الأداء والأهداف هذه تعالج القضية المطروحة.

Response: The benchmarks and five-year targets developed as a part of the dissemination workshop activities have been added to the report since the first draft on which these comments are based. We believe that these benchmarks and targets address the issue raised in the question.

7. The study assumed that this intervention is the only one which will achieve the best results in a short period, how this could be justified and clarified without giving solid examples on other interventions results or at least experiences from other countries.

الإجابة: على الرغم من أن التقرير يفيد (يدعي) بأن مشروع التدخل قد حقق أهداف موضوع البحث، إلا أنه لم يتم الإفادة (الإدعاء) فيما إذا كان أي شكل آخر للتدخل يمكن أن يكون أكثر أو أقل فعالية. ولا يمكن لكل من موضوع البحث أو البيانات أن تتيح لنا القيام بمثل هذه الإفادات (الإدعاءات).

Response: Although the report does make the claim that the intervention achieved the objectives of the research activity, no claim is made about whether or not another form of the intervention would have been more or less effective. Neither the research activity nor the data would allow us to make such claims.

Managing Directorate of Educational Planning and Research (DEP&R).

1. The Interventions were implemented on a sample of student different from the one that was surveyed in 2012, for example, 2nd grade students who are now in grade 4, and 3rd grade students are now in grade 5, which might affect the study reliability.

الإجابة: إن المشاهدة المتعلقة بالطلبة صحيحة تماماً. عموما، سيكون من الصعب إجراء دراسة طولية على الطلبة كما هو موضح أعلاه. وفي حالة هذه الدراسة والإطار الزمني المرتبط بها، كان هذا الأمر مستحيلاً. هذا هو السبب في أن الصفوف المدرسية كانت محور تركيز التحليل في الدراسة، وليس الطلبة أنفسهم.

Response: The observation about students is completely correct. It would, in general, be difficult to conduct a longitudinal study with students as described above. In the case of this study and the associated timeframe, this was impossible. It is for this reason that the classroom was the focus of analysis in the study and not the students themselves.

2. On (Page14 & 21) there is contradiction between the figures; as it was noted on page 14 that the intervention was implemented on (12000) students, while on page 21 it was noted that the intervention was implemented on (2935) students.

الإجابة: يشير الرقم 12000 إلى عدد الطلبة الذين كانوا في الصفوف التي تم تنفيذ مشروع التدخل فيها. ويشير الرقم 2935 إلى عدد الطلبة في عينة الطلبة التي تم تقييمها كجزء من المسح الوطني لعام 2014 و في دراسة نهاية المدة.

Response: The 12,000 refers to the number of students who were in classes where the intervention was implemented. The 2,935 refers to the number of students in the sample of students that were assessed as part of the 2014 National Survey and Endline Study

3. Editing the translation using educational terminologies.

الإجابة: تم تحرير وتدقيق المسودة الأولى على النحو المطلوب. ومن المأمول أن تكون المسودة الثانية أفضل فيما يتعلق بهذا المعيار.

Response: The first draft has been edited as requested. It is hoped that the second draft will be better with regard to this criterion.

4. Clarifying the type of the sample and methodology used in the selecting the sample.

الإجابة: لقد تم وصف طريقة أخذ العينات للمسح الوطني لعام 2012 في القسم 2.2.1. وقد تم وصف معايير أخذ العينات لاختيار مدارس مشروع التدخل في القسم 2.2.2. وقد اشتملت عينة المسح الوطني لعام 2014، كونها دراسة نهاية المدة، على المدارس التي تم اختيارها للمسح الوطني لعام 2012.

Response: The sampling methodology for the 2012 National Survey is described in section 2.2.1. The sampling criteria for choosing the intervention schools is described in section 2.2.2. The sample for the 2014 National Survey, being an endline study, consisted of the schools selected for the 2012 National Survey.

5. There are some queries regarding the abilities of grade 2 and 3 students to orally respond to the questionnaire.

الإجابة: إن المنهجية المعيارية التي تم استخدامها في تقييم مهارات القراءة للصفوف الأولى (EGRA) وتقييم مهارات الحساب للصفوف الأولى (EGMA) في المسوحات الوطنية لعامي 2012 و2014 هي إعطاء هذه الامتحانات شفوياً. وقد تم ذلك على نحو متعمد للتأكد من أن قدرة (أو عجز) الطلبة على قراءة التعليمات أو كتابة الأجوبة لا تربك قدرتهم على المشاركة في التقييم، والذي يؤدي إن حصل، إلى إخفاء النتائج. على سبيل المثال، هل كان سبب عدم قدرة الطلبة على الإجابة بشكل صحيح هو عدم تمكنهم من قراءة التعليمات أو عدم تمكنهم من كتابة الإجابات، أو أنهم لم يستطيعوا الإجابة عن السؤال؟ ونتيجةً لاستخدام نفس المنهجية في المسوحات الوطنية لعامي 2012 و2014، فإن الطريقة التي تم تقديمها بها لا تتضارب مع تحديد أثر مشروع التدخل.

Response: The standard methodology for the Early Grade Reading Assessment (EGRA) and the Early Grade Math Assessment (EGMA) that was used in both the 2012 and 2014 National Surveys is to administer these tests orally. This is deliberately done to ensure that the ability (or inability) of students to either read instructions or to write answers does not confound their ability to participate in the assessment, and in so doing, to mask the findings. For example, did students not respond correctly because they could not read the instructions or could not write the answers, or because they could not answer the question? Because the same methodology was used both in the 2012 and 2014 National Surveys, the manner in which it is administered does not interfere with determining the impact of the intervention.

Educational Training Center

1. The material presented to the students do not commensurate with the time allocated to implementing the tasks, which affects students' performance and character.

الإجابة: لقد أقر التقرير بأن المعلمين قد عانوا (ولا سيما في البداية) لإكمال أنشطة مشروع التدخل خلال الـ 15 دقيقة المُخصصة لذلك.

Response The: report acknowledges that teachers struggled (especially at first) to complete the intervention activities in the allocated 15 minutes.

2. The material presented to the students are enriching activities rather than remedial, which affected the realization of the project's objective, which is addressing the weak acquisition of skills at students revealed through the National Survey 2012.

الإجابة: هذه ملاحظة مثيرة للاهتمام. بالنسبة للجزء الاكبر، فقد كان القصد من الأنشطة أن تكون تأسيسية أكثر من أن تكون إثرائية، وبالتأكيد تشير الأدلة إلى أن الأنشطة تدعم تنمية المهارات التأسيسية التي تم تحديدها على أنها غير متطورة بما فيه الكفاية في المسح الوطني لعام 2012. لذلك، وجد المعلمين بأن المسائل الحسابية في أنشطة الحساب خارج نطاق المنهاج الدراسي. وقد كان هذا بمثابة سوء فهم لما كان هو مقصود من المشروع، وسيتم مناقشة هذا الأمر بشكلٍ مطول في القسم 1.4 من التقرير.

Response: This is an interesting observation. For the most part, the activities were intended to be more foundational than enriching, and certainly the evidence is that the activities supported the development of foundational skills that were identified as underdeveloped in

the 2012 National Survey. That said, teachers did experience the word problems in the mathematics activities to be beyond the scope of the curriculum. This was a misunderstanding of the project's intent and is discussed at length in section 1.4 of the report.

3. The material included learning outcomes that are higher than the outcomes identified to be acquired by students in those grades.

الإجابة: ينطبق هذا التعليق غالبا على المسائل الكلامية في أنشطة الحساب ويمثل سوء فهم لما هو مقصود من المشروع، كما أنه يشكل أحد التحديات التي تواجه عملية التنفيذ. وستتم مناقشة هذا الأمر بإسهاب في القسم 1.4 من التقرير.

Response: This comment applies mostly to the word problems in the mathematics activities and represents a misunderstanding of the project's intent and one of the implementation challenges. It is discussed at length in section 1.4 of the report.

4. Individual differences were not taken into account when preparing the material that was limited to enriching activities that measure higher levels, which negatively affected the students' performance.

الإجابة: لا بأس من القول أن هذه المواد لم تعمل على دعم عمليات التعليم المتمايز. فقد كانت المواد نفسها بالنسبة لجميع الطلبة. فقد كان هدف البحث من مشروع التدخل وضع برنامج من شأنه أن يدعم المعلمين في توفير ممارسة يومية متعمدة, ومنظمة، ومناسبة تنموياً في المهارات الأساسية للقراءة والحساب ومعرفة فيما إذا كان من شأنها أن تعالج الفجوات في المهارات الأساسية التي تم تحديدها في المسح الوطني لعام 2012. يتناول موضوع البحث العلاجي الحالي عمليات التعليم المتمايز. وقد تم بالفعل تناول التعليق الذي يتضمن أنه تم اختبار المواد على أنها إثرائية في الإجابة على السؤال 2.

Response: It is correct to say that these materials did not support differentiated teaching. The materials were the same for all the students. The research interest of the intervention was to develop a program that would support teachers in providing deliberate, structured, and developmentally appropriate daily practice in foundational skills for reading and mathematics and to see if that would address the gaps in foundational skills that were identified in the 2012 National Survey. The current remedial research activity addresses differentiated teaching. The comment that the materials were experienced as enriching has already been addressed in the response to question 2.

5. The material did not include varied reading activities; despite that reading is one of the key elements of the project. In addition, reading the invented words subtask was measured, which does not commensurate with the students' development characteristics.

الإجابة: في برنامج القراءة، تم تقديم مواد القراءة في الفصل الدراسي الثاني. حيث تعمل المهارة الفرعية لاكتشاف الكلمات المخترعة على تقييم قدرة الطلبة على فك رموز النص. إن فك رموز النص يعتبر مهارة أساسية وضرورية في حال أردنا تحسين قدرة الطلبة على القراءة بطلاقة. وتهدف مهمة اكتشاف الكلمات المخترعة إلى تقييم فيما إذا كان الطلبة يطبقون معرفتهم للعلاقات ما بين الأصوات والرموز، بدلا من قراءة الكلمات من الذاكرة. إن الكلمات المخترعة القابلة للفظ في المهمة الفرعية تتبع جميعها أنماط إملائية جائزة في اللغة العربية.

Response: In the reading program, reading materials were introduced in the second semester. The invented word subtask assesses the ability of students to decode text. Decoding text is a foundational skill and critical if students are to improve in their ability to read fluently. The invented word task assesses whether students are applying their knowledge of the relationships between sounds and symbols, rather than reading words from memory. The pronounceable invented words in the subtask all follow legal spelling patterns in Arabic.

6. Lack of alignment between the material content presented to the students with the content of the textbooks.

الإجابة: بإعتبار أن الغرض من مواد مشروع التدخل هو تطوير المهارات في القراءة والحساب والتي وجد المسح الوطني لعام 2012 بأنها غير متطورة بما فيه الكفاية كما أن تحليل المناهج الدراسية لمنهاج عام 2012 كشف بأن العديد من هذه المهارات لم يتم تناولها في الصفين الثاني والثالث، جاءت مواد مشروع التدخل لتتناول هذه المهارات. لذلك، نتج عدم تطابق متوقع ما بين مواد مشروع التدخل ومحتويات الكتب المدرسية. وستتم مناقشة هذه المسألة أيضا في القسم 1.4 من التقرير.

Response: Given that the purpose of the intervention materials was to develop skills in reading and mathematics that the 2012 National Survey had found to be underdeveloped and that the curriculum analysis of the 2012 curriculum had revealed that many of these skills were not addressed in grades 2 and 3, the intervention materials had to address these skills.

Therefore, a predictable mismatch resulted between the intervention materials and the contents of the textbook. This issue is also discussed in section 1.4 of the report.

7. The teachers had an impression that the routine activities are additional burdens rather than activities contributing to improving students' performance.

الإجابة: بالفعل، شعر المعلمون بأن مشروع التدخل هو بمثابة عبء، وقد تمت مناقشة الأمر على نطاق واسع في التقرير. سيكون لهذه التجربة آثار ومضامين بالنسبة لأي دراسات تدخل مستقبلية ستعمل فقط مع عينة من المدارس، لأن المعلمين في إحدى المدارس سيقومون بتنفيذ أنشطة التدخل، في حين لن يقوم المعلمون في المدرسة المجاورة بذلك. وهذا الأمر يزعج المعلمين. لا يدعم البحث الادعاء القائل بأن المعلمين لم يشعروا بأن المواد تساهم في تحسين أداء الطلبة. على العكس من ذلك، وافقت الغالبية العظمى من المعلمين الذين تمت مقابلتهم على أن أداء طلبتهم تحسن نتيجةً لمشروع التدخل.

Response: It is true, and it has been discussed extensively in the report, that teachers experienced the intervention as a burden. This experience has implications for future intervention studies that work with only a sample of schools, because teachers in one school will be implementing the activities of the intervention, and teachers in the neighboring school will not. This upsets teachers. The research does not support the claim that teachers did not experience the materials as contributing to improving students' performance. On the contrary, the vast majority of interviewed teachers agreed that the performance of their students improved as a result of the intervention.

8. The findings of the project showed that there was some improvement in the students' performance after implementing the intervention activities. Despite that such improvement is statistically significant, it was not equivalent to the efforts, time and cost, and some skills were not affected at all.

الإجابة: من الصعب الإجابة على الادعاء القائل بأنه "لم يكن يعادل الجهد, والوقت, والتكلفة" لأن أساس هذا الادعاء غير واضح. ومن المثير للاهتمام أن نلاحظ أن قسم المناهج قد قام بدمج العديد من منهجيات مشروع التدخل في المناهج والكتب المدرسية المنقحة والمستخدمة في المدارس خلال العام الدراسي 2014/ 2015.

Response: It is difficult to respond to the claim that “it was not equivalent to the efforts, time, and cost,” because the basis for making this claim is not clear. It is interesting to notice that the Curriculum Department has incorporated many of the methodologies of the intervention into the revised curriculum and textbooks that are used in schools for the 20142015 school year.

9. The project’s findings showed that some variables were controlled, most important of which is the economic status, whereas some other variables, which are more important, were not controlled.

الإجابة: ليس من الممكن الإجابة على هذا الادعاء، لأن "المتغيرات الأخرى، والتي هي أكثر أهمية" لم يتم تحديدها.

Response: It is not possible to respond to this claim because the “other variables, which are more important” are not identified.

10. The report did not include the recommendations.

الإجابة: تمت إضافة التوصيات التي تم تطويرها كجزء من أنشطة ورشة عمل نشر التقرير إلى المسودة الأولى التي تستند إليها هذه التعليقات.

Response: The recommendations that were developed as a part of the dissemination workshop activities have been added to the report since the first draft on which these comments are based.

Managing Directorate of Curricula and Textbooks (DCT).

1. The report did not consider the effective role of the Managing Directorate of Curricula and Textbooks at all stage of the project.

الإجابة: لقد أدرك المؤلفون أنه في المسودة الأولى من التقرير، لم يتم الاعتراف بمديرية إدارة المناهج والكتب المدرسية بشكلٍ مناسب لدورها الهام والضروري في جميع الأنشطة المرتبطة بمشروع التدخل. وقد تم تصحيح هذا الخلل، وقام المؤلفون بالاعتذار دون تحفظ على هذا الإغفال.

Response: The authors realize that in the first draft of the report, the Managing Directorate of Curricula and Textbooks was not appropriately recognized for their important and critical role throughout the activities associated with the intervention. This has been corrected, and the authors apologize without reservation for this omission.

2. Regarding the Math skills (EGMA), there are no technical comments on the implementation procedures, training materials and the results.

الإجابة: يقدم هذا التقرير لمحة عامة عن المواد, وتصميمها..الخ. ويمكن الإطلاع على المزيد من التفاصيل المتعلقة بشأن الفلسفة وإجراءات التنفيذ في ملاحظات المعلمين.

Response: The report provides an overview of the materials and their design, etc. More detail on the philosophy and implementation procedures can be found in the teachers' notes.

3. The routine activities will be repeated for students in the control and experimental groups when these activities are rolled out in the upcoming years.

الإجابة: هذا هو الأمر الأكثر تشجيعاً!

Response: This is most encouraging!

4. The study did not differentiate between the students' ability to read and the timed oral reading fluency. This means that the study has limited the students' ability to read and comprehend in a limited time with no more than one minute.

الإجابة: ترتبط الطلاقة في القراءة الشفوية (ORF) والفهم مع بعضهما[1] البعض. إن السماح للطلبة الذين لا يجيدون القراءة بالقراءة لفترة أطول لن يحسن بالضرورة من درجات الفهم. وقبل أن يتم الوصول إلى درجة عالية من الدقة، لن تتحسن الطلاقة في القراءة الشفوية (ORF). وقبل أن تتحسن الطلاقة في القراءة الشفوية (ORF)، لن يتحسن الفهم.

[1] فوكس, ل., فوكس, د, هوسب, م., ك., وجينكينز, جيه. (2001). الطلاقة في القراءة الشفوية كمؤشر على الكفاءة في القراءة: تحليل نظري, وتجريبي, وتاريخي. الدراسات العلمية في القراءة. 5 (3), 239-256.

Response: Oral reading fluency (ORF) and comprehension are linked.[2] Even allowing poor readers to read for longer will not necessarily improve the comprehension scores. Until accuracy is high, ORF will not improve. And until ORF has improved, comprehension will not improve.

5. On page 4 of the report, it was noted that the curricula are not capable of developing the students' basic reading skills, knowing that the implementation of the intervention activities coincided with the new curricula that focused on those basic skills.

الإجابة: يشير التعليق في الصفحة 4 إلى تحليل المناهج لعام 2012 وليس إلى المنهاج الجديد لعام 2014. وقد تم توضيح هذا في التقرير.

Response: The comment on page 4 referred to the analysis of the 2012 curriculum and not to the new 2014 curriculum. This has been clarified in the report.

6. On page (7) of the report, it was noted that "*it was decided in designing this intervention that teachers would not be provided with scripted lessons*". This does not apply to most of the activities implemented in the second semester that consisted of scripted lessons followed by questions and activities.

الإجابة: يهدف التقرير إلى إيصال معلومة أن تلك الدروس لم تتم كتابتها بمعنى أن الدروس لم تملِ على المعلمين ما يجب أن يقولوه بالضبط. وهذا يعني، إنه من الإنصاف أن نقول إن الأنشطة الروتينية كانت "منظمة للغاية." وقد تم تنقيح التقرير لإيصال هذه المعلومة بشكل أكثر وضوحاً.

Response: The report intended to convey that the lessons were not scripted in the sense that the lessons did not dictate to teachers exactly what to say. That said, it is fair to say that the activities of the routines were "highly structured." The report has been revised to convey this more clearly.

[2] Fuchs, L., Fuchs, D., Hosp, M. K., & Jenkins, J. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5(3)*, 239–256.

7. The report did not consider the real reasons behind the low results of students in the reading skills according to the National Survey results, most important of which are:
   a. Despite being trained, the staff implementing the National Survey test was not competent enough, as most of them had no experience in education, particularity in the listening texts and comprehension questions.
   b. Selecting staff whom the students were not familiar with has negatively affected the students' results.
   c. During the National Survey 2012, students were taught following the comprehensive way for teaching reading according to the old curricula. During the intervention stage / 2014, the curricula were reviewed and started using the syllables way for teaching reading /verbal method which has positively affected the study's results.

الإجابة:

أ. فيما يتعلق بالتعليق القائل بأن "*الموظفين الذين قاموا بتنفيذ امتحان المسح الوطني لم يمكن لديهم الكفاءة اللازمة، لأن معظمهم ليس لديه خبرة في التعليم، ولا سيما في نصوص الاستماع وأسئلة الفهم*". تتضمن إجراءات التدريب للمقيّمين امتحان صارم للموثوقية بين المقيمين (IRR) للتأكد من أن جميع المقيمين كانوا أكفاء لتقديم التقييم بشكل صحيح. وقد قام كل مقيم، أجرى التقييمات، بتسجيل درجات ملائمة بما لا يقل عن 90% مع المعيار الذهبي لتقييمات الموثوقية بين المقيمين (IRR). علاوةً على ذلك، لقد كان نصف المقيمين على الأقل في كلا المسحين من موظفي وزارة التربية والتعليم.

ب. تم إجراء مقابلات مع جميع الطلبة من قبل المقيمين الذين ليسوا من معلميهم، لضمان حصول جميع الطلبة على نفس التجربة. يمكن القول بأن الطلبة يشعرون بالخوف عندما تتم مقابلتهم من قبل أشخاص/المقيمين الذين لم يعتادوا عليهم. ومع ذلك، لم يكن ذلك ما اختبرناه.

ج. لهذا السبب تم استخدام منهجية الفرق في الفروقات (DID) لتفسير نتائج المسح. حيث تقارن منهجية الفرق في الفروقات (DID) (1) متوسط التغير مع الزمن في متغير النتيجة للمجموعة التجريبية, مع (2) متوسط التغير مع الزمن في نتيجة المجموعة الضابطة. وقد تم احتساب دور التغييرات في المناهج بشكل عام، والتغيرات في مهمة التقييم, والتحيز غير المقصود في الاختيار.

Response:

a. With regard to the comment that "*staff implementing the National Survey test was not competent enough, as most of them had no experience in education, particularity in the listening texts and comprehension questions.*" The training procedures for assessors involved rigorous inter-rater reliability (IRR) testing to ensure that all assessors were competent to administer the assessment properly. Each assessor, who conducted the assessments, scored a minimum of 90% agreement with the Gold Standard on the IRR assessments. Furthermore, at least one half of the assessors in both surveys were MoE staff.
b. All students were interviewed by assessors who were not their teachers, to ensure that all students would have the same experience. That said, it could be argued that students feel intimidated when they are interviewed by assessors who are unfamiliar. However, that has not been our experience.
c. It is for this reason that the difference-in-differences (DID) approach was used to interpret the results of the survey. The DID approach compares (1) the average change over time in the outcome variable for the treatment group with (2) the average change over time in the outcome for the control group. The role of changes in the curriculum in general, changes in the assessment task, or unintended selection bias are accounted for.

8. There is no need to roll out the routine activities materials (EGRA) for all schools, for the following reasons:
   a. The new Arabic Language textbook follows the syllables way for teaching reading/ verbal method in teaching Arabic language skills, which the intervention activities depend on.
   b. The content of the routine activities are not compatible with the content of the textbooks in terms of timeline. It is known that the warming up part that includes the intervention activities at the beginning of the lessons should be linked with the learning objectives of the lesson.
   c. Giving more attention to the in-service training of teachers.
   d. Making intensive field visits by supervisors to activate their role.
   e. Limiting the transfer of the first three grades teachers in order to retain trained and experienced teachers.

الإجابة: إن هذه الإجابة مثيرة ومشجعة على حدٍّ سواء.

Response: This response is both exciting and encouraging.

Managing Directorate of Examinations and Tests (DET).

1. Amending the title by adding “the First Three Grades”.

الإجابة: لقد تمت إضافة الحاشية التالية للتقرير "على الرغم من أن "الصفوف الأولى" تستخدم بشكلٍ عام في هذا التقرير للإشارة إلى سنوات المدرسة الثلاثة الأولى (الصفوف الأول-الثالث)، وقد كان هذا في حالة المسح الوطني 2012، ولكن فيما بعد أي في المسح الوطني لعام 2014، تم تقييم طلبة الصفين الثاني والثالث فقط. وفي أنشطة التدخل، كانت جميع الصفوف الثلاثة (الأول- الثالث) معنية".

Response: The following footnote has been added to the report “[1] Although “early grades” is used generally in this report to refer to the first three school years (grades 1–3), in the case of the 2012 National Survey, and later the 2014 National Survey, only grade 2 and grade 3 students were assessed. In the intervention activity, all three grades (1–3) were involved.”

2. The study questions are not independently well articulated in the report and should be included in a separate section.

الإجابة: تم سرد أسئلة البحث في القسم 1.2.1.

Response: The research questions are listed in section 1.2.1.

3. Clarifying Figure 1 on allocating time for three skills within the classroom daily activities.

الإجابة: تم تعديل التقرير لمعالجة هذه النقطة.

Response: The report has been adjusted to address this point.

4. Clarifying the map on page 12.

الإجابة: تم تعديل التقرير لمعالجة هذه النقطة.

Response: The report has been adjusted to address this point.

5. Separating Table 1 into two tables; one for reading and the other for Math, as the skills are different for both subjects.

الإجابة: تم تعديل التقرير لمعالجة هذه النقطة.

Response: The report has been adjusted to address this point.

6. Including all study tools in the report.

الإجابة: يمكن الاطلاع على جميع أدوات الدراسة في الملحق 2 من التقرير النهائي.

Response: All study tool will appear in Annex 2 of the final report.