Antura and the Letters

Impact and Technical Evaluation

Prepared by: Nedjma Koval-Saifi, Principal Investigator, INTEGRATED,
& Jan Plass, Co-Investigator, New York University CREATE

March 2018

ALL CHILDREN READING: A GRAND CHALLENGE FOR DEVELOPMENT
DLD DIGITAL LEARNING FOR DEVELOPMENT
unicef INNOVATION
Norad
INTEGRATED from insight to impact
create Consortium for Research and Evaluation of Advanced Technologies in Education

Integrated Services, Indigenous Solutions, 2018.

Published by
World Vision, Inc. and Foundation for Information Technology Education and Development, Inc.

World Vision, Inc.
300 I Street NE
Washington, D.C. 20002 USA
All Children Reading: A Grand Challenge for Development (ACR GCD)
www.allchildrenreading.org

Foundation for Information Technology Education and Development, Inc. (FIT-ED)
3/F Orcel II Building
1611 Quezon Avenue
Quezon City 1104 Philippines
Digital Learning for Development
www.dl4d.org

This research was conducted under the Digital Learning for Development (DL4D) project of the Foundation for Information Technology Education and Development (FIT-ED) of the Philippines, jointly funded by the International Development Research Centre (IDRC) of Canada, the Department for International Development (DFID) of the United Kingdom, and All Children Reading: A Grand Challenge for Development (ACR GCD)—a partnership of the United States Agency for International Development (USAID), World Vision, and the Australian Government. The views expressed in this work are those of the authors and do not necessarily represent those of the funders.

Recommended citation:
Koval-Saifi, N., & Plass, J. (2018). *Antura and the Letters: Impact and technical evaluation*. Washington, DC: World Vision and Foundation for Information Technology Education and Development.

Acknowledgments

The members of the evaluation team from Integrated Services, Indigenous Solutions (INTEGRATED) would like to acknowledge the contributions of individuals and institutions that supported the field work, through arranging and facilitating field work in Azraq camp. Specifically, our thanks go to Dr. Ahmad Rababa'a of Relief International (RI), who made the testing within RI's centers possible. We would also like to thank RI's volunteers, teachers, and center coordinators who facilitated the testing and implementation of sessions in the centers and provided administrative and logistical support to the evaluation team in Azraq camp. Most importantly, the evaluation team would like to thank children and parents who engaged in the impact evaluation—without their participation this evaluation would not be possible.

Finally, we would like to thank the team at All Children Reading: A Grand Challenge for Development (ACR GCD), Digital Learning for Development (DL4D), the Norwegian Agency for Development Cooperation (Norad), and United Nations Children's Fund (UNICEF) Office of Innovation who made this study possible and who worked together with us to maneuver challenges encountered in the field. We want to extend our special appreciation for Dr. John Comings' technical guidance throughout the process. In addition, we want to thank Andrew MacNamara of CREATE for his assistance in data collection, analysis and report writing. Our gratitude is also extended to the back office and field teams at INTEGRATED for their tireless efforts in tracking children and parents within this study to minimize attrition among study groups, despite extremely challenging circumstances.

Contents

Acronyms and Abbreviations

ACR GCD	All Children Reading: A Grand Challenge for Development
CLSPM	Correct Letter Sounds per Minute
CNWPM	Correct Non-Words per Minute
CREATE	Consortium for Research and Evaluation of Advanced Technologies in Education
CSPM	Correct Syllables per Minute
CWPM	Correct Words per Minute
DDL	Development Data Library
DGBL	Digital Game-Based Learning
DiD	Difference in Difference
DL4D	Digital Learning for Development
EGRA	Early Grade Reading Assessment
ELS	E-Learning Sudan
FGD	Focus Group Discussion
FTM	Feed the Monster
GSMA	Groupe Spéciale Mobile Association
ICT	Information and Communications Technology
IDRC	International Development Research Centre
INEE	Inter-Agency Network for Education in Emergencies
INTEGRATED	Integrated Services, Indigenous Solutions
IRB	Institutional Review Board
IRR	Inter-Rater Reliability
IS	Islamic State
KG	Kindergarten
KII	Key Informant Interview
LTM	Little Thinking Minds
M&E	Monitoring and Evaluation
M4R	Mobiles for Reading
MDI	Minimum Detectable Impact
MENA	Middle East and North Africa
Norad	Norwegian Agency for Development Cooperation
NTNU	Norwegian University of Science and Technology
ORF	Oral Reading Fluency

R&D	Research and Development
RAMP	Early Grade Reading and Mathematics Project
RCT	Randomized Control Trial
REACH	Regional East African Community Health
RI	Relief International
RTI	Research Triangle Institute
SDQ	Strengths and Difficulties Questionnaire
SEL	Social Emotional Learning
SIM	Subscriber Identification Module
TNO	The Netherlands Organization for Applied Scientific Research
UNHCR	United Nations High Commissioner for Refugees
UNICEF	United Nations Children's Fund
USAID	United States Agency for International Development

Photo Credit: Norad/Marit Hverven

1. Executive Summary

An estimated 2.3 million Syrian children are out of school because of violent conflict in their country. These children are primarily displaced within Syria; living as refugees in Turkey, Lebanon, Jordan, and Iraq; or in transit camps in countries like Greece and Italy. Many Syrian children have endured multiple traumas and high levels of stress, affecting their ability to learn. Moreover, some Syrian refugee children, who are attending school in a new country, are often being taught in a language they do not speak or understand. These complexities and others stress the urgency for finding innovative, scalable solutions to this education crisis.

Motivated to ensure Syrian children have an opportunity to learn to read in Arabic and continue their education, the Norwegian government funded the EduApp4Syria competition. The competition sought to develop an open source smartphone application that could build foundational literacy skills in Arabic and improve psychosocial well-being for Syrian refugee children. The competition was coordinated by the Norwegian Agency for Development Cooperation (Norad) in cooperation with the Norwegian University of Science and Technology (NTNU); All Children Reading: A Grand Challenge for Development (ACR GCD)—a partnership of the United States Agency of International Development (USAID), World Vision, and the Australian Government; mobile operator Orange; and the Inter-Agency Network for Education in Emergencies (INEE). Two games were shortlisted as winners through the two-phased competition: *Antura and the Letters* and *Feed the Monster*.

ACR GCD, Digital Learning for Development (DL4D), and the United Nations Children's Fund (UNICEF) Office of Innovation, supported an evaluation of the two apps that sought to assess the effects on literacy learning and psychosocial outcomes. The evaluation also assessed the technical and gaming aspects of the two apps and compared these against commonly accepted measures of quality and evidence-based practice.

Integrated Services, Indigenous Solutions (INTEGRATED), in partnership with Consortium for Research and Evaluation of Advanced Technologies in Education (CREATE) of New York University, conducted an impact evaluation using a longitudinal quasi-experimental design to estimate the impacts of the EduApp4Syria games on children's literacy and psychosocial outcomes over time. In this design, we compared growth in literacy outcomes for two groups of children, each using one of the games, to a group of children in matched environments who did not have access to the games. All children in the study had little or no schooling and lived in the Azraq refugee camp in Jordan. This design relies on comparing two groups that experience the same historical trends and events over time. Simultaneously, INTEGRATED and CREATE conducted a technical evaluation, working closely with ACR GCD and DL4D to refine key game evaluation questions and identify usability improvements for the next release of the games; provide feedback on the open beta versions built in Phase 2 of the competition; and inform improvements to be applied in Phase 3. The qualitative evaluation was conducted in collaboration with CREATE researchers who are among the leading experts in their specialization: assessing game use and engagement among children with digital-based learning games. The contributions of CREATE ensured depth in qualitative gaming use data used while triangulating conclusions.

Key Findings: Literacy

- Overall, the game effectively introduces children to the basics of Arabic literacy and has resulted in ***positive learning outcomes across all age groups and genders*** with a relatively low dosage of 27 hours. Absolute ***gains tended to concentrate in foundational Early Grade Reading Assessment (EGRA) subtasks*** (letter, syllable, and oral reading fluency [ORF]), gaining an additional two to three letters/syllables/words on average. ***Zero scores also decreased*** across all subtasks.
- For children with previous exposure to basic literacy, the game ***effectively reinforces previously acquired skills***. Regression analysis across subtasks confirms this, as ***variables influencing positive performance showed statistical significance with Age and Self-Reported Ability to Read Alone, meaning those with previous literacy exposure can make greater absolute gains***. Nevertheless, ***younger children showed higher rates of change*** across all subtasks. This indicates that children who sustain this rate of change over a longer dosage period could potentially make greater absolute gains, especially if the game is modified to suit their age/learning level.
- ***The lack of direct instruction of letter sound***—a key element of decoding in the game—***likely explains why early or younger learners were unable to make larger absolute gains*** in letter sound, syllable reading, invented words, and ORF. ***The game must offer targeted instruction in phonics and decoding skills***, key predictors of early reading.
- The ***levels are appropriately staggered*** such that basics are taught first—with levels building on concepts introduced in previous levels—and new concepts taught within clusters. However, many ***complex mini-games*** are introduced too early, ***do not correspond to the level of instruction***, and ***appear to be inappropriately matched*** to early learners. In this sense, the game could benefit from increased tailoring of mini-games to the level/lesson under review.
- While the ***Difference in Difference(DiD) analysis did not demonstrate statistical significance***, the learning gains and positive trends in rates of change in ***learning outcomes because of game play show promise with this low level of dosage.*** These ***upward trends promise to better serve literacy outcomes, once the game is modified*** to account for recommendations and a higher dosage and/or larger sample.
- The ***game is very long***, as demonstrated by the fact that no children completed the game. The age-appropriateness of pace, especially for younger children, should be a focus of revision.
- ***Boys tended to have much higher gains and rates of change across subtasks than girls***. While ***regression analysis did not demonstrate gender being a significant factor*** influencing subtask outcomes, it remains ***an area for further investigation*** to explain why boys outperformed girls in every subtask.
- Finally, the rate of ***smartphone penetration*** among Syrians participating in the study is ***very high***, demonstrating ***promising implications for the widespread download and use of literacy games*** such as Antura and the Letters.

Key Findings: Engagement and Psychosocial Outcomes

- Antura and the Letters ***engaged children*** regardless of literacy level, age, or gender. Antura and the Letters ***was well designed in its variety of game play and leveling of skill sets***, which enabled most children of all ages and genders to play the first two levels. While the initial interface was challenging for some, by midline most children could ***play independently and make progress***. This reflected a relative ease of use by children.
- Within the game, children ***reacted positively to Antura as a character; the off-task play space*** motivated children to continue playing and ***sustained their interest***. The mini-games were many and varied, and were randomized such that ***child trajectories were not the same***, which likely increased the level of ***sustained engagement***.
- While game play was varied and game mechanics for most mini-games were easily understood, ***objectives or pathways of the game were not immediately clear to children***, generating some level of frustration. Some mini-games either ***need revision to be better understood*** or should be dropped from the game.
- Children were ***effectively rewarded by the game, through verbal praise and in-game rewards***, and received feedback from the game. However, some children were unable to perform the tasks requested as fast or with as much precision as the older children. ***Younger children tended to be more frustrated with mini-games requesting more advanced fine motor skills, precision, and speed*** and tended to have difficulty in playing mini-games.
- At a psychosocial level, the game appears to have ***supported the development of positive social outcomes***, as demonstrated by both an assessment employing the results of the Strengths and Difficulties Questionnaire (SDQ) and parental feedback. The sense of accomplishment and learning seems to have contributed to ***children feeling happy***. The game enabled them to also feel ***a sense of ownership and attachment*** either through child profiles or through the customization of Antura. ***Peer interaction increased*** during game play, which had a positive ***impact on emotional states and social behaviors,*** while gaming engendered ***high levels of motivation to attend summer camp***. This had implications for ***motivations to learn.***
- While parental engagement with the game, albeit not part of the testing design, was limited due to a lack of parental awareness or lack of access to a device or electricity. Most ***parents expressed a willingness to download the game***, with one parent noting she used it to ***increase her own literacy skills***.

1.1 General Literacy Recommendations

Antura and the Letters could benefit from more focus on teaching letter sounds and building syllables—before building words—with a greater focus on phonics and decoding. Specifically, with respect to skills the EGRA subtasks assessed, the game in its current form lacks phonics as a key element of decoding skills that are needed for improving literacy outcomes.

Since the results by age group differed greatly, the game could also benefit from improved pacing for younger children and simplification in early mini-games that focus on shorter words, easier tasks,

more time per game, and improved matching of games to skill levels. This should include using letters/diacritics already introduced, making use of short three-letter words, and matching instruction and game play to the lesson under review.

The game in its current form is too long and may benefit from modifications to pacing, delayed introduction of more complex mini-games, and removal of games that are not understood. This could include eliminating randomization of mini-games—allocated at least in the early stages—and presentation of more challenges in later stages.

1.2 General Engagement and Psychosocial Recommendations

To ensure pathways are clear and game play is clearly understood, improvement and simplification of the game's user interface is needed. The simplification should enable younger children to play with minimal or no adult supervision. The game should offer a clearer introduction of objectives, scoring, and travel and better communicate how the child is progressing through the game. In addition, the game should be better streamlined to enable more frequent milestone achievements and completion of levels through game simplification and improved pacing.

> Antura and the Letters engaged children regardless of literacy level, age, or gender. Antura and the Letters was well designed in its variety of game play and leveling of skill sets, which enabled most children of all ages and genders to play the first two levels.

Since the Antura off-task play space effectively engaged all children, the game could benefit from increased visibility of this play space. As the game enables children to feel a sense of ownership and attachment—either through child player profiles or through the customization of Antura—the game could benefit from increased personalization of profiles and off-task spaces.

For younger children, the game could improve through differentiated pacing, are vision of mini-game mechanics, and the allocation of more time to complete tasks.

While not a requirement of the EduApp4Syria competition, there is room for capitalizing on the positive impact of social interaction in game play to increase pro-social behaviors and psychosocial outcomes. This could be accomplished through the development of sharing mechanisms to support peer interaction or friendly competition. Alternatively, sharing gaming through group game play—as practiced in classrooms in the testing of this game—can also serve this function and help maximize social interaction through the joint goal of game completion.

Photo Credit: Norad/Marit Hvervẹn

2. Introduction

2.1 Syrian Refugee Context

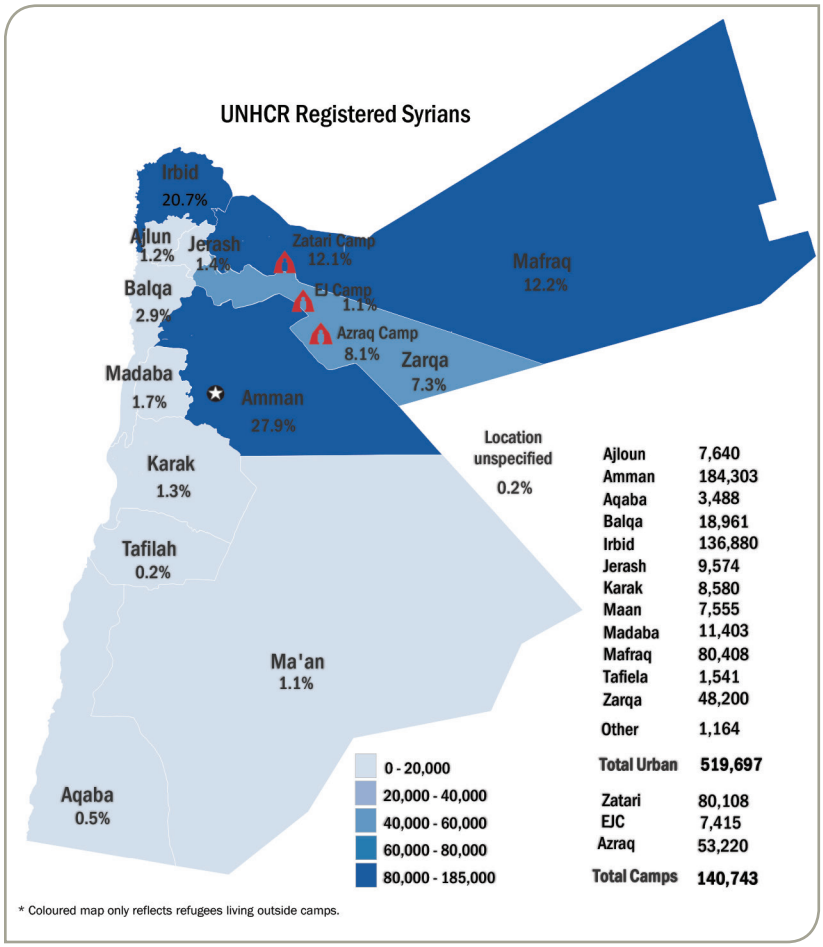


Figure 1. *Map of registered Syrian refugee populations in Jordan, courtesy of UNHCR.*

In the Levant region and around the world, perhaps the most pressing challenge in aiding Syrian refugees is how to address the educational challenges facing a wave of Syrian school-aged children, displaced within Syria, refugees predominantly residing in Jordan, Lebanon, and Turkey,[1] or those in transit camps on their way to Europe. Since the Syrian Civil War broke out in January of 2011, it has displaced over 11.3 million Syrians. At least 6.3 million of these civilians are displaced within their own country, and over 5 million have become refugees, fleeing primarily to neighboring Middle Eastern countries.[2] An estimated 2.3 million Syrian children are out of school because of violent conflict in their country.[3] Many of these children must cope with memories of multiple traumas and high levels of stress in their present living environment. The conflict has disrupted their education, and trauma and stress levels often affect their ability to learn.

As of 31 July 2017, demographic information from the United Nations High Commissioner for Refugees (UNHCR) places the number of registered Syrian refugees in Jordan at 660,440, although refugees not registered with UNHCR raise this figure to over 1 million *(see Figure 1)*. These refugees live throughout the country, with about 520,000 living in urban areas and about 140,000 living in formal UNHCR camps of Zaatari, Emarati and Azraq. As of 2013, there were nearly 300,000 Syrian children registered as refugees in Jordan,[4] with many more living in the country unregistered. Many Syrian children in Jordan have struggled to continue their studies in the face of limited education infrastructure in Jordanian cities and within refugee camps *(Table 1)*. Despite efforts to expand opportunities, 17% of the children are not enrolled in any form of schooling, and 29% are only enrolled in informal education programs.[5] These circumstances make Jordan the third largest host of Syrian refugee children and a country that reflects many of the educational challenges facing this population of children worldwide.

Table 1: *Enrollment Status of School-Aged Syrian Refugee Children in Jordan*

	School-Age Children	Enrolled in Formal Education	Enrolled in Non-Formal Education	Out of School
Registered Syrian Refugees in Jordan	232,868	125,000	67,658	40,210

2.2 EduApp4Syria

To ensure Syrian children have an opportunity to learn to read Arabic and continue their lifelong learning process, the Norwegian government funded the EduApp4Syria competition. The competition sought to develop an open source smartphone application that could build foundational literacy skills in Arabic and improve psychosocial well-being for Syrian refugee children. The competition was coordinated by the Norwegian Agency for Development Cooperation (Norad) in cooperation with the Norwegian University of Science and Technology (NTNU); All Children Reading: A Grand Challenge for Development (ACR GCD)—a partnership including the United States Agency of International Development (USAID), World Vision, and the Australian Government; mobile operator Orange; and the Inter-Agency Network for Education in Emergencies (INEE). Two games were shortlisted as winners through the two-phased competition: *Antura and the Letters* and *Feed the Monster*. This competition aimed to use Digital Game-Based Learning (DGBL) as a groundbreaking method for opening new avenues to literacy for Syrian refugee children worldwide.

DGBL, an instructional method by which educational content or learning objectives are incorporated into digital games, is touted by scholars as a leading method of learning. An innovative approach to learning, DGBL could address some of the challenges faced by educators in the Middle East and North Africa (MENA) region. Specifically, game-based learning offers potential to reduce barriers in accessing learning materials and alleviate the need for physical learning environments; it also offers the opportunity for iterative and differentiated learning and potential to improve the quality of education in places where schooling is weak. DGBL is gaining momentum in the Middle East with growing work in design and implementation. USAID's 2014 landscape review of 44 projects and studies undertaken of information and communications technology (ICT) for education (specifically reading) revealed that "use of mobile ICTs designed to help children learn to read, practice reading (reading to learn), and acquire a broader range of learning skills that support a literate society…provides a new opportunity to re-imagine traditional forms of educational design and delivery."[6] A recent global meta-analysis of 77 randomized experiments on school-based interventions on learning in developing country primary schools found the largest mean effect sizes (among school based interventions) were those that included treatments with computers or instructional technology. There is a lack of a rigorous evidence base on the impact of DGBL on learning. In the 2014 landscape review of 44 mobiles for reading (M4R) projects, "only one M4R project contained an adequate randomized control trial (RCT) impact evaluation design."[7]

In the MENA region, very few DGBL applications have undergone rigorous testing providing such an evidence base. We have found no examples of rigorously-tested DGBL solutions for Arabic literacy with applicability to the Syrian refugee context. An example with many relevant features and high complementarity to the EduApp4Syria approach is, however, Qysas (although not a game, it is a rigorously-tested Arabic literacy app developed under ACR GCD). Another relevant example is Can't Wait

to Learn (a game-based Arabic numeracy app developed under E-Learning Sudan [ELS]). These two examples were chosen as two rigorously tested digital learning tools with relevance to Arabic language instruction, psychosocial well-being, and education in crisis contexts, and they illustrate some of the challenges and possible solutions to widespread DGBL adoption. The desk review conducted within this study and attached in Annex 1 goes into more depth on the features and measured impact of these apps. One of the more immediate challenges facing DGBL solutions in crisis contexts is that of the available technological infrastructure. Within conflict or refugee settings, access to electricity can be unpredictable, and Wi-Fi or Internet connections are often weak (if they are present at all). Therefore, leveraging DGBL to serve these groups requires off-line capability.

Research has shown that specific conditions must be met for games to be effective. Games are most effective when used for subjects most learners feel unmotivated to study, or when offering engaging, contextualized learning material. Games must contain a well designed learning mechanic, which is the essential game play repeated throughout the game that leads to the desired learning outcomes. The context in which the game will eventually be used must also be considered, as should an appropriate balance between fun and learning. For more on DGBL in the region and associated approaches and challenges, please see Annex 1, Literature Review.

The winning games were intended primarily for individual use by Syrian children affected by the crisis, regardless of their current country of residence. The games are meant to be intuitive and engaging enough to be played independently on a smartphone with minimal adult supervision. The primary target group is out-of-school children. The games are also intended for use as a mother tongue learning supplement by refugee children who are attending school in a setting with a new language of instruction. The games are not primarily intended to be used in more formal or structured educational settings, although such use is encouraged by the competition partners. The testing of the game was, however, conducted in a structured manner within remedial education centers to ensure adherence to testing protocols and dosage.

Antura and the Letters was developed by a consortium that includes the Cologne Game Lab, Wixel Studios, and Video Games Without Borders. In responding to the competition requirements, the team focused on several key goals during the development process. These goals are pedagogical in that they involve teaching the Arabic language, and psychosocial in that they support the player's well-being.

The story arc of the game tasks the player with helping an old "Keeper" character watch over little Living Letters with minds of their own. With the help of her dog, Antura, the player goes through a series of mini-games that progress through a basic literacy curriculum. The game comprises six Journey Maps that correspond to the six major stages of the Syrian primary school curriculum. The game was designed with constant input from Arabic literacy experts who broke the language into Learning Blocks with specific micro-pedagogical goals to be achieved. A block of new letters, for example, requires not only recognizing the letter in a word, but also the association of the letter's shape, its sound, and how to write the letter, among other tasks.

General goal	Mini-game or activity name
Recognize the letter shape, from its sound or name	FastCrowd_letter
	HideSeek
	ThrowBalls_letters
	Tobogan_letters
	Balloons_letter
Pronunciation and sound recognition games	TakeMeHome
	Egg
	MakeFriends
Draw the shape	Maze
	ColorTickle

Figure 2. *Pedagogical goals associated with the mini-games in Antura and the Letters.*

Figure 2 above is an excerpt from game development materials, illustrating the connections between pedagogical goals and the mini-games contained in Antura and the Letters.

In addition to fostering the player's literacy capabilities, Antura and the Letters was designed to support the player mentally and emotionally. According to the developers, the core mental well-being objective of Antura and the Letters is to put the player in a mental state of "procedural flow." This means the game is meant to offer players a series of constant challenges to keep them physically and mentally engaged.

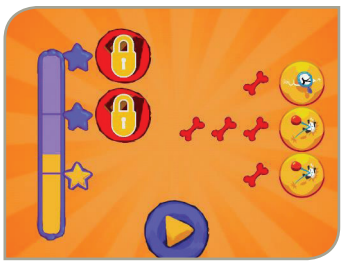


Figure 3. *Players earn bones to feed Antura (at right) and track their progress through the meter on the left.*

Some mini-games also contain emotional incentives constituted in the behavior of the Living Letters/Words. Players receive positive feedback for correct answers, and consistent achievement is incentivized in game, as exemplified by the "Egg" mini-game in which players must help Living Letters hatch. In another mini-game, the player helps reconcile two words that have quarreled by joining them together with a shared letter.

According to the developers, Antura and the Letters' pedagogical and psychosocial components are supported by strategies to keep players engaged. As children progress through the mini-games, they are rewarded new accessories to change the dog's costume. They are also able to see their progress toward the goals of progressing through the map, gaining new accessories for Antura, and gaining bones to feed Antura. At the end of a given set of mini-games, players are shown how many bones they earned in each game *(Figure 3)* and their progress fills up the meter on the left, unlocking new accessories to put on Antura.[8]

Photo Credit: Norad/Marit Hverven

3. EduApp4Syria Evaluation Goals and Objectives

ACR GCD and DL4D, solicited an evaluation of Antura and the Letters and Feed the Monster among their target end users: out of school Syrian children ages five to 10. The evaluation of both winning games assessed the games according to three objectives:

- Effects on literacy learning;
- Effects on improvements in psychosocial well-being; and
- Ease of use and ability to engage children.

INTEGRATED conducted an impact and technical evaluation to assess the impact of the EduApp4Syria games on children's literacy and psychosocial outcomes over time. In partnership with CREATE of New York University, INTEGRATED also assessed the game's user experience within the two primary focus areas of assessment: literacy and psychosocial outcomes. While the intended use of the game is home-based use by out of school children, testing of the game was conducted in a structured manner within remedial education centers to ensure adherence to testing protocols and dosage.

3.1 Literacy

Achieving reading and writing fluency (literacy) is foundational for lifelong learning. The EduApp4Syria competition criteria for literacy learning included:

- Learning that language is made up of a specific set of individual sounds and that letters and letter combinations represent those sounds;
- Developing the skill to decode letters, letter combinations, and words;
- Vocabulary;
- Oral reading fluency (speed and accuracy); and
- Comprehension.

The impact evaluation used a longitudinal quasi-experimental design to estimate the impacts of the EduApp4Syria on children's literacy through the EGRA assessment. This was supported by a qualitative evaluation of the games' impact on literacy levels and learning. The impact evaluation compared growth in literacy outcomes for children using the games to children in matched environments who did not receive the game. The technical evaluation examined the competition criteria and Arabic language acquisition best practices within the game, as well as parent, child, and teacher responses to evaluation questions. The goal was to assess strengths and weaknesses to provide the game developer with useful inputs on the perceptions and performance of games in field environments. Key evaluation questions in this domain were as follows *(Table 2):*

Table 2: *Key Literacy Evaluation Questions*

Impact of Antura and the Letters on Literacy
To what extent did the game achieve the planned objective(s) of improving literacy?
■ To what extent has the game been successful in achieving improved literacy outcomes among children in the targeted communities?
■ How do instructors/adults rate the reading abilities of each of the participating children before starting to play the games? (No previous Arabic literacy skills, Basic, Moderate, Good, Very Good, Excellent)
■ How would you evaluate the game in terms of teaching literacy to the children?
■ How would you evaluate the game in terms of improving the children's reading skills and vocabulary? (Supported by EGRA findings)
Ways in Which Use of Antura and the Letters Can Be Improved
To what extent is the game design based on the technological needs of the beneficiary population?
■ To what extent are the game's requirements (hardware—tablets, phones, Internet, and electricity accessibility requirements) based on the Syrian refugee context?
■ Is there sufficient hardware to support access?
■ Do children have access to hardware available at home? In community centers?
■ Male/Female differentials?
To what extent is the game design based on the language learning needs of the beneficiary population?
■ To what extent is the game's design based on relevant Arabic for the beneficiary group?
■ Is the game using vocabulary, storylines, and language appropriate for the beneficiary group?
■ What do you think of the game in terms of teaching literacy to the children?
■ In which ways can the game or use of the game be improved to teach literacy to children?
■ How would you evaluate the game in terms of improving the children's reading skills and vocabulary?
■ In which ways can the game or use of the game be modified to improve children's reading skills and vocabulary?
■ Do Syrian parents and/or instructors consider the game and its content appropriate for their children?

3.2 Psychosocial Outcomes

Syrian children inside or outside of school are living under the extreme stress of a protracted conflict, with elevated and prolonged stress levels. The EduApp4Syria competition criteria for psychosocial well-being included:

- Happiness;
- Ability to play;
- Cognition (concentration, memory, ability to solve a problem);
- Sadness (including depression, grief, crying);

- Stress reactions (fear, anxiety, arousal, avoidance, sleep problems, regression, etc.);
- Somatic health (self-reported stomach pains, headaches, sleep, tiredness, appetite);
- Helplessness, clinginess, independence;
- Ability to regulate emotions and communicate;
- Level of worry about everyday life or self while playing the game; and
- Emotional and visceral involvement in the game.

The evaluation used a mixed-methods approach to estimate the impacts of the EduApp4Syria game on children's psychosocial well-being through the Strengths and Difficulties Questionnaire (SDQ) assessment among sampled children's parents, supported by a qualitative evaluation of the game's impact on psychosocial well-being and engagement. In this design, we intended to compare growth in psychosocial well-being for children using the games to children in matched environments who did not receive the game over time. The technical evaluation examined the competition criteria as well as parent, child, and teacher responses to evaluation questions among the treatment group. The goal was to provide the game developer with useful inputs on the perceptions and performance of games in field environments. Key evaluation questions in this domain were as follows *(Table 3)*:

Table 3: *Key Psychosocial Evaluation Questions*

Impact of Antura and the Letters on Psychosocial Outcomes and Engagement
To what extent did the game achieve the planned objective(s) of improving psychosocial outcomes?
■ How effective was the game in engaging children?
■ Which elements/stimuli of the game were most/least engaging?
■ To what extent did the children respond positively to the rewards, and in what ways?
■ How effective was the game in enabling children to achieve intended psychosocial outcomes among early grade learners in the targeted communities?
■ Are there observed gender differentials with respect to use of the games? (Supported by SDQ findings)
Ways in Which Use of Antura and the Letters Can Be Improved
To what extent is the game design based on the language learning needs of the beneficiary population?
■ Are the game's storylines appropriate for the beneficiary group?
■ Do Syrian parents and/or instructors consider the game and its content appropriate for their children's psychosocial well-being?
In which ways can use of the game be improved?
■ Are there observed differences in levels of interest among age groups? Among testing cohorts? If so, what are they?

3.3 Gaming User Experiences

To achieve the intended impact in literacy and psychosocial wellbeing, the EduApp4Syria competition criteria for user experience included:

- Game interfaces and mechanics are fun and engaging for children in the target age group.
- Game interfaces and mechanics are simple enough for children in the target age group to use them with minimal adult supervision.
- The game provides various stimuli to hold the user's concentration over a sufficient period.
- The game matches the user's level of literacy skills and increases the challenge at an appropriate pace as the user's literacy skills improve.
- The user is rewarded appropriately for her/his efforts and skills development.
- The user feels a sense of control over the content in the game, and the application supports recovery from errors.
- The overriding goals of the game are presented early and clearly, and intermediate goals are presented at appropriate times.
- The user receives immediate feedback on her/his actions and on progress toward the goals of the game.

The qualitative evaluation was conducted in collaboration with CREATE researchers specialized in assessing game use and engagement among children. CREATE focused on assessing key elements of DGBL: rationale for game use; appropriateness of the learning mechanic; appropriate design for users and context; and balance of fun and learning. The assessment of user gaming experience as it pertains to the two primary objectives of literacy learning and psychosocial well-being were jointly assessed by INTEGRATED and CREATE through gaming observation protocols, engagement observations, and qualitative feedback from parents, children, and teachers. The goal was to provide the game developer with useful inputs on the perceptions and performance of games in field environments. Findings pertaining to gaming user experiences are embedded within the two primary objectives of the literacy and psychosocial wellbeing assessment. Key evaluation questions in this domain were as follows *(Table 4):*

> Nevertheless, serious games or simulations with clear "learning outcomes can create interactive experiences that actively engage the players in the learning process, [whereby] experimentation, graceful failure, and identification of lessons learned can result from game-based learning; where decisions and actions are chosen, consequences experienced, goals are achieved, and feedback furnished."[8]

Table 4: *Key Gaming User Experience Evaluation Questions*

Ways in Which Use of Antura and the Letters Can Be Improved
To what extent did the game achieve the planned objective(s) of improving psychosocial outcomes?
■ How effective was the game in engaging children?
■ Which elements/stimuli of the game were most/least engaging?
■ To what extent did the children respond positively to the rewards, and in what ways?
To what extent is the game design based on the technological needs of the beneficiary population?
■ To what extent does the game meet the technical/software requirements set by the EduApp4Syria competition?
■ To what extent are the interfaces and mechanics simple enough for children to use with minimal adult supervision?
How can the game be improved?
■ Are there observed gender differentials with respect to game use?
■ Are there observed differences in levels of interest among age groups? Among testing Cohorts? If so, what are they?

Answers to these evaluation questions will inform developers of the game improvements necessary to better serve the needs of Syrian refugee children aged five to 10 with little or no schooling.

3.4 Literature on Digital Game-Based Learning and Engagement

A review of relevant literature on DGBL and Arabic literacy acquisition was conducted to better inform the evaluation. DGBL is a disruptive approach to learning that holds the possibility to address some of the challenges faced by educators in the MENA region. DGBL offers the potential to reduce barriers to accessing learning materials, offers the opportunity for iterative and differentiated learning, and alleviates the need for physical learning environments. Game-based learning also offers the opportunity to improve the quality of education delivery. This is because DGBL supports children with direct instruction "targeted at their individual learning needs, and adaptive to their learning progress."[9]

DGBL is gaining momentum in the Middle East, with growing work in design and implementation. Although research based on these development activities is underway, few studies have been conducted to date on educational games, the markets for DGBL, and the impact of DGBL in the MENA region. Nevertheless, serious games or simulations with clear "learning outcomes can create interactive experiences that actively engage the players in the learning process, [whereby] experimentation, graceful failure, and identification of lessons learned can result from game-based learning; where decisions and actions are chosen, consequences experienced, goals are achieved, and feedback furnished."[10]

A growing number of early childhood scholars are advocating for the inclusion of digital technology in early childhood education. However, research is unclear about the possible effects of digital games on young learners whose brains are still maturing.

Other research shows that while connections between game play and improved cognitive ability exist, there may be differentials among boys and girls.[11] For example, some studies show that boys demonstrate "greater involvement with, liking of, and experience in computer gaming" but that "learning gains that boys and girls achieved using the game did not differ significantly, and the game was found to be equally motivational for boys and girls."[12]

Evidence for DGBL shows its potential to support the acquisition of knowledge, assist teachers in instruction, and offer mechanisms for safe training circumstances/simulations for hazardous environments.[13] These games require players to make decisions to drive progress and have been embraced by educators inside and outside the classroom.

The literature and empirical evidence reveal that game-based learning environments offer potential for increasing student engagement and motivation, which have natural ties to learning.[14] The relationship between emotional and cognitive activities is strong: "that positive affect such as engaged concentration, joy, and excitement can lead to increased learning through better strategy selection, increased persistence, and improved use of mental resources."[15]

DGBL has the potential to mimic the benefit of one-to-one tutoring through interactivity, individualized attention, and feedback. For example, games offer motivational features found to motivate learners, including "incentive structures, such as stars, points, leader-boards, badges, and trophies, as well as game mechanics and activities that learners enjoy or find interesting."[16] Learning gains are only possible if the player or child is engaged and motivated by this interaction and use.

Studies in computer-based learning environments show how students' emotional states—including boredom, confusion, frustration, and anger—can lead to decreased motivation and disengagement from the task. Disengagement can manifest itself as off-task behavior,[17] although the literature reveals further investigation is needed to assess whether off-task behavior negatively affects learning, or if off-task behavior can also be a coping mechanism for negative learning emotions. The literature also notes that certain off-task, in-game behavior may allow the player to remain engaged in the gaming environment rather than the learning content. This may engender positive feelings towards the gaming environment but lower the learning impact. Removal of these off-task in-game features may decrease positive outcomes such as engagement.[18]

A recent study on different types of instruction within DGBL revealed that, "depending on their nature, instructions [whether play-focused or goal-focused] can therefore play a key role during the cognitive processing of educational content," where goal-focused instruction elicited deeper learning than play-focused instruction, without negatively impacting motivation. DGBL can "promote learning and motivation provided its features prompt learners to actively process the educational content."[19] Still, other studies reveal that multi-player DGBL may enhance collaborative learning, group dynamics of positive interdependence, and an intergroup dynamic of competition.[20]

A handbook developed by All Children Reading (ACR) outlining DGBL's potential in developing countries notes that digital games can hold children's attention while managing to provide practice in basic literacy skills, with the ultimate result being that children grow in "automaticity" (i.e. the ability to decode words automatically). As such, DGBL enables students to more quickly reach the point where they are no longer struggling with the basics of reading text, which will improve comprehension and allow children to enjoy learning through reading.[21]

According to the handbook, the best DGBL tools allow users to build on what they have already learned and increase their knowledge by small increments.[22] Successful games also make children "active participants" in learning to read and write and take into account the three hierarchical domains of learning (cognitive, affective, and psychomotor) that feed into better educational outcomes.[23] These three domains of learning ensure the student is engaged on the levels of knowledge, attitudes, and skills, respectively.[24] In doing so, games can address the dominant theories of learning—both cognitive and constructivist.

According to the cognitive theory of learning, students learn through a sequential knowledge-building process designed and facilitated by a teacher. According to the constructivist theory of learning, children manage their own learning.[25] DGBL platforms allow for both cognitive and constructivist methods of teaching to the extent that they provide either a closed or open-ended game. Closed games guide players through a series of challenges they must overcome. Open-ended games, on the other hand, give players freedom to express themselves creatively and do not necessarily guide players linearly through a set of mini-games.

As such, game designers must balance the engaging appeal of open-endedness with the structure of a closed design when developing their products. Examples of open-ended learning in educational games include "learning by exploring," which encourages a player to find extra areas of the game world, and "learning by building," which allows players to customize elements of the game world.[26] Whatever the design choices within a game, its creators must ensure that "game activities, rules, and mechanics all help the players achieve [the goals of allowing players to learn and practice as they play]."[27]

3.5 Literature on Arabic Literacy Acquisition

Recent studies reveal that an important part of the Arabic literacy problem is the Arabic orthographic system and its multiple scripted variations.[28] The Arabic alphabet has28 basic consonant letters, each with up to three or four variant shapes dependent on their placement in the word. In total there are more than 60 base forms, which "lead to graphemic difficulty and a significant learning problem."[29] Studies reveal Arabic textbooks overwhelmingly teach reading through whole words, moving away from leveled introduction of letters, partial words, and phonetic learning and toward a desire for direct reading of unvoweled Arabic. This places complex demands on young children for language and literacy acquisition. Studies show that whole language and whole word approaches are inefficient in lower-income environments.[30] EGRAs conducted by Research Triangle Institute (RTI) in Jordan revealed that overall students' knowledge of letter-sound correspondence is low, and as a result, students' ability to decode new words is low. Those who fail to learn the sound-letter combinations or small letter units cannot go on to texts of greater complexity.

The written Arabic language can be written with diacritical markings (i.e., vowels, shaddah, hamza) or without such markings. When written without diacritical markings, the reader must be able to guess the word from memory and sentence context. An empirical research study undertaken in Abu Dhabi and Palestine on the role of diacritics for beginning readers reveals that "diacritical markings were found to significantly influence the reading of both poor and skilled readers."[31] Research also found that both skilled and poor readers improved their reading accuracy when they read with vowels."[32]

Finally, the gap between fuṣ-ha (Modern Standard Arabic-MSA) and the Arabic dialect spoken outside of the school environment seems to be a major cause of low learning achievement in schools.[33] School-aged children generally have about 2,000 to 5,000 words they can comprehend orally and use in communication, which in most other languages forms the foundation on which language acquisition is built.[34] However, in Arabic, fuṣ-ha is the language of instruction, and therefore places children at a disadvantage due to diglossia (use of high Arabic—fuṣ-ha—and low Arabic—colloquial). Proficiency in colloquial Arabic sometimes then contributes to confusion and difficulty in learning connections, since children cannot always leverage native linguistic competence in colloquial Arabic. "They cannot use their lexical familiarity with their native basic Arabic sounds, forms, structures, and syllabic and prosodic features because these are not necessarily identical with fuṣ-ha forms and structures, even though they may show important and striking similarities."[35] The literature recommends using vocabulary drawn from the 40% overlap between fuṣ-ha and colloquial Arabic to reduce diglossia challenges *(Table 5)*.

Table 5: *Challenges to Arabic Literacy Acquisition*

1	The Arabic alphabet has 28 letters, many of which have similar shapes distinguished only by dots. Letters also have three or four variations depending on their placement in the word.
2	Arabic textbooks overwhelmingly teach reading through whole words, which do not allow for gradual introduction of letter phonics and voweled variations before introduction of whole words. Textbooks overwhelmingly teach reading through whole words.
3	The gap between fuṣ-ḥa and the dialect spoken at home contributes to low literacy achievement.
4	The Arabic script is consistent and easy to read when vowels are marked; when they are not, students must predict words. Because of these changes, some students may see the vowels deleted just as they are beginning to read automatically. This may reduce reading speed and further delay comprehension, resulting in poor literacy performance.
5	If all attention must be spent on deciphering texts, there is little time left to spend on information and comprehension.
6	Comprehension is the goal of reading, but occurs only after early obstacles are overcome. To attain literacy automaticity, students need pages of legible text to practice and allocation of learning time and space for comprehension.

3.6 Azraq Camp Context

Azraq houses more of the new arrivals from Syria, and is the reception point for those Syrians coming in from the border.

Table 6: *Population of Azraq Camp by Age/Gender*[36]

Age Range	Male	Female	Total
0–17	29.6%	27.6%	57.2%
18–59	18.7%	21.9%	40.6%
60+	0.9%	1.3%	2.2%
Total	**49.2%**	**50.8%**	**100%**

Population Profile of Azraq Camp: While the registered population of Azraq Camp is 53,220, the residents in the camp number 35,529.[37] The population of this camp is even younger than the overall profile of Syrian refugees within Jordan, with 57.2% of the population below the age of 18 years old. Primary school-aged children between the ages of 5 and 11 alone represent 24.4% of the camp population.[38] As in the general refugee population, gender is balanced across age groups *(Table 6).*

The approximately 35,500 refugees currently in the camp live in a 14.7km^2 camp area, in caravan-style shelters built of sheet metal. The camp is divided into six "villages," which are population centers with a designated number of units that are separated from one another by distance or walls and additional security checkpoints. At the time of this report, only four of the six villages are occupied: Village 2, Village 3, Village 5, and Village 6.

Table 7: *Population of Syrian Refugees by Place of Origin* [39]

Place of Origin	Percentage of Azraq Camp Population	Percentage of Total Syrian Refugee Population
Aleppo	26.8%	10.3%
Homs	18.6%	15.9%
Der'a	13.2%	41.7%
Raqqa	10.1%	2.5%
Other	31.3%	29.6%

Origin of Refugees: Azraq Camp differs from the general Syrian refugee population because of the high presence of refugees who fled Raqqa, the center of Islamic State (IS) control over Syria. In contrast to the general population, of which 41.7% have their place of origin in Der'a (a rural farming province), only 13.2% of Azraq residents hail from Der'a. About 10.1% of Azraq residents come from Raqqa and 26.8% from Aleppo—both which are urban centers *(Table 7).*

Security Measures Among Villages: The high concentration of residents from Raqqa has resulted in higher security measures in the Azraq camp. Village 5 presents extra security measures and a near-complete ban on residents' travel outside of the Village perimeter fence. The Village 5 security measures are in place as a precaution in the case that some refugees from areas previously held by IS remain loyal to IS.[40] Even beyond Village 5, however, mobility is limited, with stronger restrictions on travel outside the camp than exist in the older, more established environment of Za'atari Camp. Moreover, the Azraq camp has limited or no Wi-Fi/Internet access, except weak reception in some areas near the main highway.

Early Grade Education in the Azraq Camp: One of the largest challenges in the Azraq Camp is the provision of education to the large population of children, of which 9%, or 1,084 school-aged children, were out of school as of 14 May 2017.[41] A larger proportion of school-aged children has spent a prolonged period out of school and has thereby fallen behind. In 2015, over 43% of children ages 6–17 in the camp were out of school, over 25% of the children in the camp had previously attended formal education and then dropped out of school, and over 17% of children had never attended school.[42]

Photo Credit: Norad/Marit Hverven

4. Literacy Learning

4.1 Measuring Literacy Learning: Methodological Approach

INTEGRATED employed a dual approach to evaluation: impact evaluation and technical evaluation.

4.1.1 Impact evaluation

The impact evaluation used a longitudinal quasi-experimental design to estimate the impacts of the EduApp4Syria games on children's literacy over time, comparing growth in literacy outcomes for children using the games to children in matched environments who did not receive the game.

Hypothesis: if the games are applied five to six times per week for up to 30 hours, then literacy outcomes of children using the games will improve more than those who did not use the game.

The hypothesis assumes the treatment and control groups displayed comparable traits before the games were implemented, and the conditions existed to confidently attribute observed changes in children's literacy and psychosocial well-being outcomes to use of the game(s), rather than other factors.

***Early Grade Reading Assessment (EGRA)**:* The primary instrument used to assess literacy levels was the EGRA already developed and tested by RTI in Jordan, and approved by the Ministry of Education in 2012. Using EGRA sub-tasks pertaining to basic literacy acquisition for Grades 1 to 3, the evaluation team conducted a baseline among control and treatment groups. The EGRA focused on four subtasks corresponding to the Antura game, along with description and analysis *(Table 8):*

Table 8: *EGRA Subtask Description and Analysis*

EGRA Subtask	Description and Analysis
Letter-Sound Identification	The score for this subtask is the number of letter sounds a child reads correctly in one minute, a measure known as Correct Letter Sounds per Minute (CLSPM). There is a total of 100 letters presented on the stimulus.
Syllable Reading	The score of this subtask is the number of syllables read correctly in one minute, a measure known as Correct Syllables per Minute (CSPM). There is a total of 100 syllables presented on the stimulus.
Non-Word Reading	The score for this subtest is a measure of the number of Correct Non-Words Read per Minute (CNWPM). There is a total of 50 words presented on the stimulus.
Oral Reading Fluency (ORF)	The score of this subtest is a measure of the number of Correct Words read per Minute (CWPM). There is a total of 52 words presented on the stimulus.

Sampling of Control and Treatment Group: The impact evaluation aimed to establish treatment and control group equivalency to the extent possible in the refugee camp context in Jordan. To address concerns that children within the three assigned groups (two treatment groups and one control group) have underlying differences, children were matched[43] on a range of characteristics. Characteristics of critical importance for establishment of equivalence among groups were:

- Syrian nationality
- Little or no schooling
- Ages 5 to10
- Gender balance
- Living in camp setting in Jordan
- A beneficiary of a humanitarian actor
- Access to a mobile phone (if possible)

The study first began by identifying possible test subjects meeting the above criteria. Relief International (RI), a humanitarian relief agency providing remedial education services to 30,000 Syrian children, offered its education centers as testing centers for the game. Numbers and locations of children served by RI meeting the above criteria in Azraq camp were identified (see Annex 2, Methodological Approach).

Control and treatment groups were randomly assigned to different administrative villages. The humanitarian response community offers similar basic services within each village, establishing a similar environment across villages. Parents of selected children were then invited to participate in the study to provide survey feedback on the psychosocial outcomes testing (see Methodology in Section 3.2 Psychosocial Outcomes).

Child Selection: Selection was according to age strata with the following results *(Table 9):*

Table 9: *Child Selection Results*

Baseline	5 years	6 years	7 years	8 years	9 years	10 years	Total
Antura Treatment	1.4%	20.5%	16.7%	20.2%	19.9%	21.3%	366
Control	1.4%	19.8%	18.7%	21.2%	19.8%	16.5%	283
Endline	**5 years**	**6 years**	**7 years**	**8 years**	**9 years**	**10 years**	**Total**
Antura Treatment	2%	18%	15%	19%	24%	22%	202
Control	1%	21%	19%	22%	19%	19%	200

Although the populations experienced high levels of attrition, the population stratification at endline did not differ significantly from the baseline population, especially when looking at younger (ages 5–7) versus older (ages 8-10) age groupings, performance on EGRA, and gender, enabling analysis according to plan. **There was no statistical significant difference in the baseline population compared to the endline population.**

Testing Context: While use of the game is intended for home-based use by out-of-school children, testing of the game was conducted in a structured manner within remedial education centers to ensure adherence to testing protocols and dosage. Due to the higher proportion of out of school children among newer arrivals in the Azraq camp, it was selected as the most appropriate testing site. Volunteers and teachers were trained in testing protocols and tablet/gaming operation and instructed to help the children navigate the game on their own with minimal intervention.

Relief International Centers: In response to the need to provide educational services to out of school children, RI operates six remedial education centers in Azraq camp that target children ages 5–18 and aims to provide multiple remedial pathways with learning support for children who have been out of school for several months or years. This program enables students to gain the basics of literacy and numeracy and helps them pass the official formal school exams and move forward to the next grade. Since 2013, the RI remedial education program has supported approximately 30,000 students within RI remedial education centers in Za'atari and Azraq Camps. Syrian refugees with teaching qualifications teach courses covering Grades 1–12 from a central curriculum in math, science, Arabic, and English. The educational centers double as locations for student recreation in a safe, supervised space. Syrians are also recruited in the centers to conduct case management for students, as well as community mobilization. It is from these centers that children (and parents) participating in this study were recruited.

In the summer term between academic school years, RI provides a "Summer Club" for children in Azraq Camp. The Summer Club is not part of the required informal education program, but rather serves to strengthen students' learning in preparation for the school year to come. For the 2017 summer term, 1,725 children in total registered for the Summer Club in Azraq Camp.[44] In the Summer Club, children attend an Arabic class, a math class, and two activity classes with arts, games, and recreation. For the course of the study, the children registered in the experimental group had "Tablet Class" as one of their two activities classes, whereas children in the control center in Village 5 continued with its usual curriculum. Each class ran for about 45 minutes, and the centers were operational six days a week, except for the two Eid holidays (June 25–July 2 and August 30–September 5).

Game Testing Context and Conditions: INTEGRATED conducted the impact evaluation with 366 treatment and 300 control children in Azraq Camp, within remedial education centers operated by RI in Azraq camp in Village 2 (Antura and the Letters) and Village 5 (Control). The children were provided with tablets uploaded with the game. Syrian teachers and volunteers ran the game testing sessions, trained by INTEGRATED staff in basic protocols and procedures of game play, attendance data collection, and child-tablet tracking. INTEGRATED staff members followed up twice per week with center staff administering the testing.

Antura and the Letters Testing: Within these classes, Syrian volunteers/teachers supported the rollout of the testing of Antura and the Letters, played on tablets provided by ACR and DL4D. Antura can be completed in 15 hours.[45]

Treatment and Control Settings: During the summer months in Azraq Camp, official public schools are not operational. Nevertheless, RI continues to run remedial support and recreation classes during the summer, under their summer club programming described above. These remedial support classes are Arabic literacy, math, and recreation activities. Thus, the control group was engaged in "business as usual" in the remedial centers, and treatment groups engaged in use of the game during one of the activity classes offered during the summer camp program.

Attrition: The treatment and control groups were over-sampled to about 300 children per assessment group to account for up to 25% attrition and assure a minimum endline cohort of 225 children per treatment/control group. However, attrition levels experienced were higher than planned for all three groups, reaching 45% among children and 60% among parents, caused by a variety of factors:

- *General Factors:* The testing conditions in the summer camp in Azraq camp are not easily controlled, leading to difficulty in tracking children over time. Children in Azraq camp can "shop" their summer camp and do not always commit to one summer camp program, leading to a situation where some children tested at baseline did not join the summer camp program. Others left the camp to spend the summer in other areas in Jordan, while others may have returned to Syria when the borders were reopened in August 2017.
- *Village 2 (Antura):* During the first week of game testing, RI opened another educational center in Village 2. The children who had tested into the study at baseline who lived closer to the second RI center, moved to the second RI center—effectively dropping out of the study. Moreover, in late July and early August 2017, in preparation for the arrival of additional refugees from the border, one area of Village 2 was vacated and fenced off to receive new refugees. Families in Village 2 were moved to Village 3, and the children participating in the study dropped out after moving to another village. Finally, some Village 2 families received permission to leave the camp, and as families left–albeit temporarily–participating children of those families participating dropped out. By early August, attrition levels were so high that an additional 80 children were EGRA-tested to participate in the study. The study period was also extended for an additional two weeks of testing.
- *Village 5 (Control):* In August 2017, Azraq camp administrators began to prepare for the arrival of new refugees from the border. In preparation for the new arrivals, portions of Village 5 residents were moved to Villages 2 and 3. INTEGRATED conducted endline data collection in Village 5 at the end of August, and captured 162 children in Village 5. INTEGRATED then worked closely with RI to track those refugees who had moved. Where that information was available, INTEGRATED tracked refugees to their new homes, conducting house-to-house visits in Villages 2 and 3, as well as new areas of Village 5 to locate those parents and children who had participated in the study at baseline. This house-to-house endline data collection was conducted over three weeks (August 28–September 20).

Demographic Survey: A brief and basic demographic survey accompanied the EGRA and the SDQ for children and parents, respectively. The demographic survey collected basic information pertaining to selection criteria (age, years of schooling, child gender) for children, as well as household size and educational attainment for parents. Both demographic surveys (parents and children) aimed to establish home conditions for use of technology, such as access to electricity and presence of technology such as smartphone, tablet, laptop, etc. This was intended to inform about the Syrian refugee camp context. To the extent possible, research on conditions outside the camp with respect to access to technology was supplemented from secondary data.

4.1.2 Technical (Qualitative) Evaluation

The technical evaluation focused on questions designed to assess strengths and weaknesses of the game, should the impact evaluation be inconclusive, as well as provide a qualitative interpretation to the quantitative results. INTEGRATED, in collaboration with CREATE, employed a participatory approach to the technical evaluation using mixed methods that utilized a combination of qualitative and quantitative methods to document challenges and successes in game play and learning. This gaming assessment conducted by CREATE allowed for a dimension of analysis rarely seen in studies of educational interventions.

The participatory approach supported the gathering of rigorous and credible evidence. It included focus group discussions (FGDs), observation checklists, and key informant interviews (KIIs), whereby methods and their composition were selected to enable the collection of primary and secondary data to assess qualitative aspects of literacy learning. The methods were designed to interlink, allowing the triangulation of data to produce a verifiable body of evidence to:

- Assess the impact of each game on literacy; and
- Assess the ease of use of each game, and identify the ways in which use of the game could be improved.

The evaluation methodology consisted of: a desk review of DGBL on literacy, as well as regionally-relevant apps designed to increase literacy; semi-structured FGDs with children testing the game and their parents; in-depth interviews and FGDs with RI; observation of children using the games for gaming assessment (CREATE); and play-intensive interviews (CREATE) with child beneficiaries *(Table 10)*.

Table 10: *Evaluation Stakeholders and Data Collection*

Stakeholders	Data Collection Conducted
Donors	■ 3 KIIs
Implementing Partners (RI)	■ 1 KII with RI Head of Education ■ 1 FGD volunteers/teachers
Parents of Children in Treatment	■ 10 FGDs Antura
Children in Treatment	■ 10 FGDs Antura ■ 118 engagement observation ■ 43 gaming observation ■ 22 play-focused interviews for each game
Game Designers	■ 1 KII

4.2 Game Exposure and Dosage

There is currently no literature on dosage for DGBL, nor is there any research that estimates the impact dosage for literacy instruction for children in the MENA region. A study by Comings (1995) that focused on adult literacy program in Asia and Africa is commonly cited among education practitioners for early grade literacy as well, and it estimates that 250 hours of instruction is the minimum needed to produce meaningful reading skills. This level of dosage was not possible within the two-month time frame of the evaluation.[46] Comings informally asked evaluators of three early-grade reading improvement projects about dosage and found that about 60 hours of effective instruction produced a 0.5 effect size for change in oral reading fluency. The testing period was scheduled to coincide with the summer break, whereby Syrian children could play 45 minutes daily. Based on the days and hours of operation of the RI summer camp, a maximum of 31.5 hours of dosage (45 minutes per day five to six times per week) could be allocated over the two-month testing period. Antura and the Letters game developers estimate 15 hours as the total time to complete the game.

However, attempting this dosage within this two-month testing time frame posed some risks and challenges:

- RI centers rely on generator-powered electricity. The generator in Village 6 broke down for a period of five to six days in August, preventing the charging of tablets and the loss of seven planned sessions.
- At times, other activities took precedence in the center, and planned game-testing sessions were not delivered on those days. Those amounted to two sessions.
- The centers did not always deliver the sessions six times during the week.

The total dosage for Antura and the Letters amounted to an average of 26.6 hours against the proposed 30.

Contamination	
Antura Treatment	5 children
Control	12 children

Contamination: The Azraq camp has severely restricted Internet access and connectivity, which limited the ability of non-test subjects or control subjects to download the games. In Azraq, control and treatment groups were also separated by administrative villages, with Village 5 remaining physically separated from the other villages, as Village 5 is fenced in for security purposes. To limit cross-contamination between treatment groups, INTEGRATED worked closely with RI to ensure a strong understanding among RI staff/volunteers of the importance of non-contamination for the testing period. Nevertheless, it was discovered that some children/parents had downloaded the Feed the Monster game (the other EduApp4Syria game). An exit survey was developed at endline to establish the levels of contamination. Contaminated children's EGRA data and parents' SDQ data were removed from the study.

In one focus group with parents, one mother said, "I am illiterate. When my children told me about the game, I downloaded it and started learning it". She indicated she can now identify letters, letter sounds, and read words, which has enabled her to read signs and has raised her self-esteem.

Software/Hardware: The games operated on a combination of used and newly-purchased tablets distributed and used within RI education centers. The games were uploaded offsite and given to RI for use in the testing period. While it was planned that each tablet would serve only one child, the used tablets could not support the games. As such, those tablets were removed from testing, and new tablets were used once in the morning shift and once in the afternoon shift. Children were assigned player profiles by gender, to ensure unique logins. As Antura and the Letters enables multiple player profiles, some children created multiple. The game developers limited the avatars to two: one assigned to girls and one assigned to boys. Since Internet/Wi-fi is for the most part unavailable in Azraq camp, the games operated offline.

4.3 Limitations

Attrition was higher than the planned-for rate of 25%. Parents completing SDQ were 125 per group at baseline and 50–60 respondents at endline in each group. This smaller sample impacts the analysis of the results.

Randomization was compromised due to attrition, and the addition of more children to the sample resulted in the elimination of randomization of the treatment group. As such, the study is not a randomized control trial, and results of analysis pertain only to comparisons between the two groups studied. Conclusions cannot be projected to the larger Syrian population.

Dosage was planned to be 31.5 hours, but no child completed the fully-planned dosage, as sessions were lost to electricity cuts due to fuel shortages and generator breakdowns that prevented the charging of tablets. Since the dosage was not achieved, there may be limitations in assessing the full potential impact of the game with additional dosage.

Figure 4. *Conducting an interview with a boy from the treatment group.*

Testing conditions were fluid and not easily controlled, resulting in a situation where impact evaluation design planning and associated information was not always reflective of the realities in the field. The field team relied on 2016–2017 registration information, which was collected before summer camp registration was finalized. When refugee movements occurred, the team depended on RI's tracking of children, which could only happen if the children came to RI centers in their newly-assigned area. The ***constrained time frame*** within which this impact evaluation took place was eight weeks in the field, during the summer camp roll-out period.

Self-reported data presents its own limitations, in that responses rely on the honesty and introspective ability or bias of respondents, and therefore places some limitations on the validity of the self-reported data.

Photo Credit: Norad/Marit Hverven

5. Literacy Findings

5.1 Demographic Survey Findings

To better understand the refugee population participating in the study, the team conducted a demographic survey including background information on smartphone penetration, electricity access, current reading practices, parental education status, and use of home language. This survey was previously mentioned in Section 4.1.1, Impact Evaluation. These contextual factors were collected to better understand the sample population. Some findings might help explain the contextual factors for EduApp4Syria take-up and child performance.

***Educational Status of Parents*:** Of the parents surveyed in Village 2, 41% were illiterate (as compared to 27% in the Village 5 Control Group). Less than half (43%) were literate but had not completed high school (as compared to 42% in Village 5 Control). One-sixth (15%) had completed high school or had a bachelor's degree, as compared to 28% in Village 5 Control. In one focus group with parents, one mother said, "I am illiterate. When my children told me about the game, I downloaded it and started learning it." She indicated she can now identify letters, letter sounds, and read words, which has enabled her to read signs and raised her self-esteem.

***Access to Smartphones and Awareness of Learning Games*:** The majority (77%) of parents surveyed in Village 2 had a smartphone, as compared to 86% of parents surveyed in Village 5 Control. All parents surveyed in Village 2 could charge their phone, while 97% of parents in Village 5 Control could charge their phone. Most parents (74%) were aware of mobile learning games, as compared to 57% in Village 5 Control. Of those who had smartphones, 79% said they let their children use their smartphones, as compared to 87% in Village 5 Control. One volunteer/teacher said he sent the link to his brothers in Turkey to download for their children; the game is said to be spreading among refugee communities there.

Reading: Of the parents surveyed, 20% had books at home, while 38% of parents surveyed in Village 5 Control said they have books at home. Of those who said they had books at home, 81% said they read with their children, as compared to 40% in Village 5 Control. Twenty-three percent of parents surveyed in Village 2 said they have books at home. Of the children surveyed in Village 2, 33% said they could not read on their own, as compared to 36% in Village 5 Control.

5.2 Early Grade Reading Assessment Findings

5.2.1 Overview of EGRA instrument

Table 11 summarizes the four subtasks of EGRA used in this study. The subtasks measured children's alphabet knowledge, decoding skills, and fluency skills.

Table 11: *EGRA Subtasks and Corresponding Skills*

Subtask	Skill	The child is asked to ...
Letter Sound Knowledge	Alphabet knowledge, letter-sound correspondence	...read aloud a list of 100 letters of the alphabet presented in random order on a sheet. Letters are presented as stand-alone letters or letters as they appear in the beginning, middle or end of a word. *(Timed, 1 minute)*
Syllable Reading	Alphabet knowledge and decoding	...read aloud 100 syllables presented on a sheet. Syllables show letters with diacritics, or two letters joined to form a syllable—hence assessing knowledge of letters, letters in beginning, middle, and end of word, and letters with diacritics. *(Timed,1 minute)*
Invented Word Decoding	Alphabet knowledge and decoding	...read aloud a list of 50 invented words presented in random order on a page. Words were constructed from legitimate Arabic letter combinations but were nonsensical—e.g., a legitimate word with one letter replaced to make an invented word. *(Timed,1 minute)*
Oral Reading Fluency	Automatic word recognition (fluency)	...read aloud a grade-level short story printed on a page. *(Timed, 1 minute)*

Table 12 illustrates the game play elements against the EGRA subtask.

Table 12: *Game Play Elements and Related EGRA Subtask*

Related EGRA Task	Antura Element	Issues to Consider When Comparing EGRA to Antura Game Elements
Letter Sound Knowledge	Letter shape and name	The game introduces letters with their names, but not their sounds. While the introduction of the letter and its placement in the word is clear, the game does not make letter-sound associations, therefore differing from the letter-sound subtask.
Syllable Reading	Letters with diacritics syllable segments	The game shows letters as they appear in the beginning, middle, and end of the word and associates them with their original shape. In early levels, the game first asks children to spell the word letter by letter but not yet with its connections, then with connections by Level 3. Diacritics are introduced as lessons in Level 4, but are shown in earlier levels. Syllables in EGRA tests are made up of two or more letters—often connected—or one letter with a diacritic.
Invented Word Decoding	Letters with diacritics syllable segments	As above, the game does not make letter-sound associations—required for decoding and word attack for invented word reading.
Oral Reading Fluency	New words	The game introduces many words with pictures, and whole word reading in games, and the higher levels (4–6) introduce independent reading. Few children reached these levels.

Sampling of Control and Treatment Group: The impact evaluation aimed to establish treatment and control group equivalency to the extent possible in the refugee camp context in Jordan. To address concerns that children within the two assigned groups (one treatment and one control group) have underlying differences, children were matched[47] on a range of characteristics. Characteristics of critical importance for establishment of equivalence among groups were:

- Syrian nationality
- Little or no schooling
- Ages 5–10
- Gender balance
- Living in camp setting in Jordan
- A beneficiary of a humanitarian actor
- Access to a mobile phone (if possible)

The study first began by identifying possible test subjects meeting the above criteria. A planned randomized sample of treatment children was selected from Azraq camp, however, high levels of attrition made randomization impossible when additional recruits were added to the study. In the first week of August, an additional 80 children were added to the study. The treatment group included 366 children at baseline while the control group started at 283 children. However, due to high dropout rates, the sample was 202 and 200 in the treatment and control groups, respectively, at endline.

In total, there were 159 children who dropped out of the treatment group and 71 who dropped out of the control group. Some of the children went back to Syria with their families as the borders opened during the testing period. Moreover, many families got permits to leave the camp on "vacations" where they could visit their families that live outside the camp across Jordan. Drop-outs were assessed by age, gender, and EGRA scores to assess equivalence of baseline and endline groupings *(Tables 13–15):*

Table 13: *Sample Dropout by Gender and Group*

	Male	Female	Total
Antura	76	83	159
Control	30	41	71
Total	106	124	230

Table 14: *Sample Dropout by Age and Group*

	5 Years	6 Years	7 Years	8 Years	9 Years	10 Years	Total
Antura	1	37	28	36	24	33	159
Control	3	11	12	13	18	14	71

Table 15: *EGRA Mean Scores at Baseline for Dropped-Out Children by Subtask and Group*

		Male	SD	Female	SD	Total	SD
Letter Sound Knowledge	Treatment	5.93	13.0602	11.78	16.9641	8.99	15.4568
	Control	15.03	17.9261	13.09756	19.37112	13.92	18.66758
Syllable Reading	Treatment	3.05	6.25011	4.12	7.57627	3.61	6.97274
	Control	6.53	8.24928	6.341463	11.57067	6.42	10.23253
Invented Word Decoding	Treatment	0.83	2.04214	1.06	2.97284	0.99	2.5651
	Control	2.3	4.07812	2.00	5.074446	2.13	4.650436
ORF	Treatment	2.49	5.00265	3.20	6.71503	2.86	5.95072
	Control	5.1	9.72147	4.22	7.938867	4.59	8.680977

Although the populations experienced high levels of attrition, the population stratification at endline did not differ significantly from the baseline population, especially when looking within older (ages 8–10) versus younger (ages 5–7) age groupings, gender distribution, and EGRA performance at baseline, enabling analysis as planned. When comparing the mean scores of the group of students who dropped out of the sample and the ones who stayed, the t-test shows **no statistical significant difference in the baseline population as compared to the endline population.**

5.2.2 EGRA summary results

Overall Results by Group Type—Treatment versus Control: Per the results in *Table 16.a,* the treatment group showed gains in performance from baseline to endline and greater gains in comparison to the control group. Mean scores improved by two to three words/letters for letter sound knowledge, syllable reading, and ORF subtasks. The treatment group showed higher rates of change[48] than the control group from baseline to endline. The negative scores and rate of change indicate lower scores at endline.

Table 16.a: *Subtask Results for Treatment and Control Groups*

Subtask	Group Type	Baseline		Endline		Total		Rate of Change
		Mean	SD	Mean	SD	Gain	SD	
Letter Sound Knowledge	Treatment	7.72	14.2272	10.31	16.9379	2.59	16.6053	34%
	Control	13.10	17.60739	13.12	18.64871	0.02	20.48555	0%
Syllable Reading	Treatment	4.42	8.47273	6.67	10.2905	2.25	6.87183	51%
	Control	7.85	11.90235	8.44	12.59203	0.59	11.57939	8%
Invented Word Decoding	Treatment	0.90	2.4011	1.63	3.84714	0.73	3.15526	81%
	Control	2.54	5.74136	2.25	4.827161	-0.29	5.090929	-11%
Oral Reading Fluency	Treatment	3.42	7.32108	5.84	9.93051	2.42	6.31632	71%
	Control	6.43	11.30546	6.46	11.1038	0.03	8.851034	0%

Moreover, a t-test was run to compare changes in scores (gains) between the treatment and control groups to identify the significance of difference in gains between both groups on each subtask. As noted in *Table 16.b,* the difference between the gains of the treatment and control groups was significant on the Invented Word Decoding and ORF subtasks.

Table 16.b: *Inferential comparison between the Total Gains of each group, by subtask*

Subtask	Group Type	Total		t-statistic	Pr (ITI>Itl) 0.05	Significance
		Gain	SD			
Letter Sound Knowledge	Treatment	2.59	16.6053	1.3846	0.1669	Not Significant
	Control	0.02	20.48555			
Syllable Reading	Treatment	2.25	6.87183	1.7473	0.0814	Not Significant
	Control	0.59	11.57939			
Invented Word Decoding	Treatment	0.73	3.15526	2.4470	0.0148	Significant
	Control	-0.29	5.090929			
Oral Reading Fluency	Treatment	2.42	6.31632	3.1132	0.0020	Significant
	Control	0.03	8.851034			

Effects of the Game on EGRA Scores: To estimate the impact of the game on children's reading achievement, we compare the ORF scores, measured in CWPM, across time and treatment/control status, to answer the evaluation question of the effect size of the game on children's reading abilities.

To do so, we examine changes of test scores on the ORF subtask from baseline to endline to understand the impact of Antura on children's reading achievement in this subtask across the treatment and control groups. The analysis helps to measure the impact of Antura on EGRA scores after eight weeks of implementing the game with the treatment group to test the hypothesis that children's literacy will improve with game play. ORF was identified by the early-grade reading community as the key literacy outcome, and thus the mean scores of children on this subtask were used to draw conclusions about the impact of instructional interventions on the literacy outcomes of children.

Nevertheless, given the dosage of 27 hours, the difference in gains between treatment and control demonstrates that gains can be made in literacy with relatively low dosage of gaming.

The team used panel data from 202 children in the treatment group as well as 200 children in the control group to examine Antura's treatment effects using the DiD model. This analysis helped examine whether there were differences between baseline and endline scores of children in the treatment schools (treatment effects), and whether those differences were distinct from the differences for comparison children (effect size); and to attribute the results to Antura and not to chance or measurement error.

To calculate the effect size of the treatment, we used two variables. The first "treatment" was a dummy that takes the value of one for treatment children and zero for control children. The second "post" was also a dummy variable that takes the value of one for scores taken at endline and zero for scores taken at baseline. Then, a new variable, "DTrX," was generated to function as the interaction of the two variables. The coefficient for this new variable is the DiD, or the estimation of the effect size of introducing the game on children's literacy levels.

Figure 5: *Difference in Difference Results for ORF Subtask*

Source	SS	df	MS
Model	1248.972	3	416.323999
Residual	80565.3116	800	100.706639
Total	**81814.2836**	**803**	**101.885783**

Number of obs = 804
F (3, 800) = 4.13
Prob > F = 0.0064
R-squared = 0.0153
Adj R-squared = 0.0116
Root MSE = 10.035

ORF	Coef.	Std. Err.	t	P> l t l	[95% Conf.	Interval]
Post	.03	1.003527	0.03	0.976	-1.939857	1.999857
Treatment	-3.004208	1.00104	-3.00	0.003	-4.969183	-1.039233
DTrX	2.385842	1.415684	1.69	0.092	-.3930528	5.164736
_cons	6.425	.7096007	9.05	0.000	5.032101	7.817899

As shown in *Figure 5,* the coefficient of the newly generated variable indicates that children in the treatment groups scored 2.38 more CWPM than those in the control group when compared over time. The difference between two groups (such as an experiment versus control group) is judged to be statistically significant when $p = 0.05$ or less. The DiD p-value is 0.092. If criteria for significance are raised to a 10% margin, then it is interesting to note that the gains attributable to the treatment could be considered significant at that level. However, as the impact evaluation was not randomized and the sample was underpowered due to a higher than planned rate of attrition, these results pertain only to the two groups compared. Therefore, such results cannot be extrapolated to the larger Syrian refugee population. Nevertheless, given the dosage of 27 hours, the difference in gains between treatment and control demonstrates that gains can be made in literacy with relatively low dosage of gaming. This is a noteworthy and important finding that may have implications for dosage for literacy-based gaming interventions; it also shows the potential for DGBL to positively impact literacy outcomes.

Effect Size of the Experiment, Comparison at Endline *(Table 17):* First, we ran a t-test to test the null hypothesis that the average mean score for each subtask is the same in the treatment and control groups. The significance of the t-statistic is calculated using the two-sided p-value, which indicated that the difference between mean gains of the two groups are statistically significant if less than 0.05.

To estimate the effect size, we calculated Cohen's d using a pooled standard deviation, since this is a common practice in EGRA Studies. The control SD provides a better estimate of standard deviation, however, since it consists of a representative group of the population not affected by the experimental intervention. The confidence interval for each of these coefficients is 95%.

Table 17: *Experiment and Control Groups when Comparing the Main Gains (Endline–Baseline)*

Subtask	Difference = Mean Gain Treatment - Mean Gain Control	t-statistic	Pr (ITI>Itl) 0.05	Pooled Std. Deviation	Effect Size (Cohen's d)	Significance
Letter Sound Knowledge	2.574059	1.3846	0.1669	18.65827	0.1381158	Not Significant
Syllable Reading	0.9486164	1.7473	0.0814	9.534046	0.1742977	Not Significant
Invented Word Decoding	1.032624	2.4470	0.0148	4.25668	0.244093	Significant
Oral Reading Fluency	2.385842	3.1132	0.0020	7.765411	0.3105516	Significant

Results by Gender: As illustrated in *Table 18*, boys in the treatment group scored higher than girls at endline in all four subtasks. Additionally, boys showed higher rates of change on all subtasks, with increased scores of 54% (Letter Sound) to 116% (Invented Word) from baseline to endline, versus 13% (Letter Sound and Invented Word) to 66% (ORF) increased rate of change in scores for girls from baseline to endline.

Table 18: *EGRA Results by Gender and Group*

Subtask	Group Type	Baseline				Endline				Gains				Rate of Change	
		Male	SD	Female	SD	Male	SD	Female	SD	Male	SD	Female	SD	Male	Female
Letter Sound Knowledge	Treatment	6.88	14.113	8.83	14.383	10.57	17.633	9.98	16.067	3.69	16.309	1.15	16.975	54%	13%
	Control	11.18	16.388	15.18	18.703	9.88	16.659	16.64	20.089	-1.31	21.129	1.46	19.774	-12%	10%
Syllable Reading	Treatment	5.33	9.7968	3.22	6.1633	8.35	11.536	4.45	7.8958	3.02	7.5695	1.23	5.7074	57%	38%
	Control	7.10	11.802	8.67	12.018	8.27	12.524	8.63	12.729	1.17	11.697	-0.04	11.478	17%	0%
Invented Word Decoding	Treatment	1.06	2.6862	0.68	1.9559	2.29	4.6050	0.77	2.2707	1.23	3.6324	0.09	2.2445	116%	13%
	Control	2.47	5.8325	2.61	5.6706	2.47	5.4219	2.00	4.1013	0	5.5739	-0.61	4.5175	0%	-23%
Oral Reading Fluency	Treatment	3.77	8.0332	2.97	6.2737	6.51	10.497	4.94	9.1109	2.74	6.4029	1.97	6.2095	73%	66%
	Control	6.07	11.385	6.81	11.266	7.12	12.285	5.74	9.6772	1.05	10.035	-1.07	7.2493	17%	-16%

Though the effect sizes are low for the two subtasks that are statistically significant, they are about 40% to 50% of the estimate for 60 hours of improved classroom instruction. In addition, invented word reading is a measure of decoding skill, and oral reading fluency is a measure of whole-word reading skill.

Results by Age Group: *Table 19* shows the results disaggregated by age groups—children ages 5–7 and ages 8–10. Based on the endline results, older children in treatment showed higher scores on all subtasks compared to younger children in treatment; these results were statistically significant. The highest absolute gains were for the older children in the treatment group, particularly for ORF in which older children read 2.93 more correct words per minute at endline compared to 0.43 more correct words per minute for younger children in treatment. However, the absolute gains for invented word decoding were small (less than one additional word) for both treatment age groups (0.16 additional words for younger children and 0.90 additional words for older children in the treatment group). However, in the treatment group, younger children showed higher rates of change on their scores on Syllable Reading (220%), Invented Word (214%), and Oral Reading Fluency (163%). Improvements in rates of change among the older age group were lower, ranging from 35% with Syllable Reading, to 57% with Oral Reading Fluency, and 68% with Invented Word. It is therefore important to note that even though the absolute gains among the older age group were larger, the rates of change among the younger group were higher. These results are significant, especially when taking the low game dosage into consideration.

It is therefore important to note that even though the absolute gains among the older age group were larger, the rates of change among the younger group were higher. These results are significant, especially when taking the low game dosage into consideration.

Table 19: *EGRA Results by Age Groups*

Subtask	Group Type	5-7 Years							8-10 Years						
		Baseline	SD	Endline	SD	Gains	SD	Rate of Change	Baseline	SD	Endline	SD	Gains	SD	Rate of Change
Letter Sound Knowledge	Treatment	3	8.351219	4.05	12.83179	1.05	12.7368	35%	10.20	16.02069	12.96	18.24855	2.76	18.40795	27%
	Control	5.69	11.12839	5.94	15.10546	0.25	16.93261	4%	18.14	19.37904	18.21	20.58027	0.07	22.07943	0%
Syllable Reading	Treatment	0.51	1.970303	1.63	5.007118	1.12	5.016753	220%	6.51	9.798163	8.80	11.48853	2.29	7.681069	35%
	Control	2.83	7.262558	2.94	9.829943	0.11	9.673625	4%	11.27	13.19264	12.34	13.55774	1.07	12.70321	9%
Invented Word Decoding	Treatment	0.07	0.42507	0.22	1.796375	0.15	1.859015	214%	1.33	2.870998	2.23	4.442125	0.90	3.653231	68%
	Control	1.04	3.672344	0.37	3.20349	-0.67	3.161887	-64%	3.56	6.621279	3.57	5.524054	0.01	6.071313	0%
Oral Reading Fluency	Treatment	0.24	1.034612	0.63	4.148113	0.39	4.292617	163%	5.11	8.589011	8.04	11.24181	2.93	7.096739	57%
	Control	2.08	6.585548	1.49	6.954504	-0.59	6.549409	-28%	9.37	12.81823	9.97	12.67157	0.60	10.11448	6%

5.2.3 Results per subtask

Letter Sound Knowledge: The alphabetic principle is the understanding that words are made up of sounds and that letters are symbols that represent those sounds.[49] The letter sound knowledge subtask assesses children's ability to identify the correct sound associated with each letter of the alphabet. It includes letters presented as stand-alone letters or letters as they appear in the beginning, middle, or end of a word. Children were provided with a sheet of 100 stand-alone letters and asked to name the sound of as many as they could within one minute.

According to the letter sound knowledge subtask results in *Figure 6,* the treatment group scored lower than the control group but showed higher gains. Children in the treatment group read 7.7 correct letter sounds per minute (CLSPM) at baseline and 10.3 CLSPM at endline, resulting in a gain of 2.6 CLSPM. Meanwhile, children in the control group read 13.10 CLSPM at baseline and 13.12 CLSPM at endline, showing almost no gains on this subtask. The rate of change for children in the treatment group is 39%, while those in the control group had a 0% rate of change.

Figure 6: *Letter Sound Knowledge Results by Group Type*

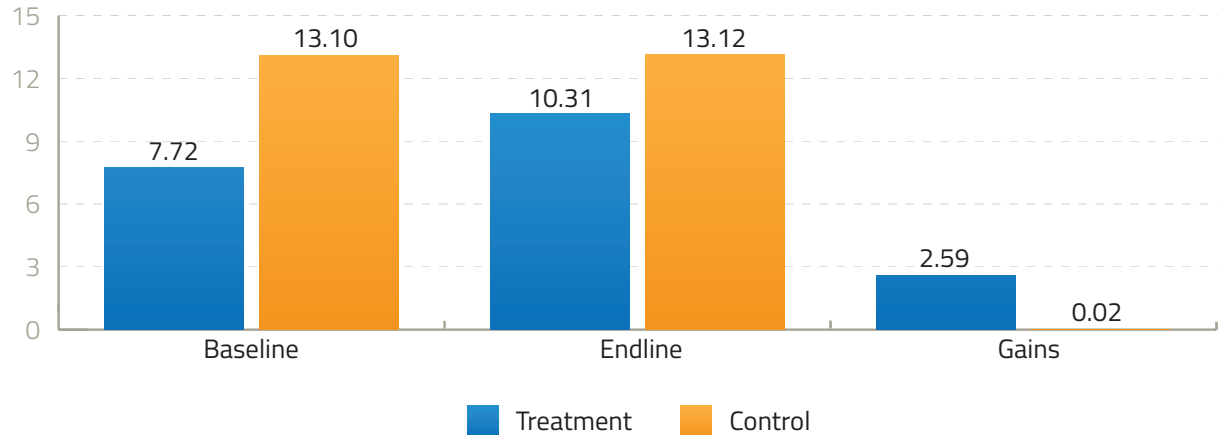


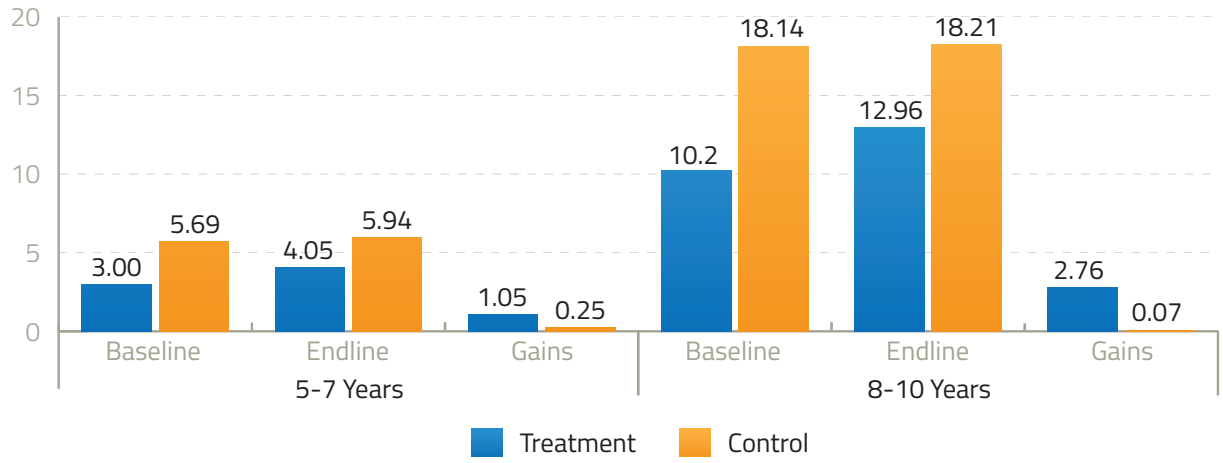
Table 20 shows letter sound results disaggregated by gender. In the treatment group, girls scored higher than boys at baseline, but boys scored slightly higher at endline. Boys demonstrated larger absolute gains: 3.96 for boys compared to 1.15for girls, and a higher rate of change (54% for boys compared to 13% for girls). In the control group, girls demonstrated a gain of 1.46, which was a larger gain than girls in treatment but smaller than boys in treatment. Boys in the control group, however, demonstrated a decrease of 1.3 letters or 12% decline in scores at endline.

Table 20: *Letter Sound Knowledge Results by Gender*

Subtask	Group Type	Baseline		Endline		Gains		Rate of Change	
		Male	Female	Male	Female	Male	Female	Male	Female
Letter Sound Knowledge	Treatment	6.88	8.83	10.57	9.98	3.69	1.15	54%	13%
	Control	11.18	15.18	9.88	16.64	-1.30	1.46	-12%	10%

The letter sound results by age group presented in *Figure 7* indicates that older children had higher literacy scores at baseline and endline, and thus more exposure to letter sounds prior to playing the game. Among treatment children, older children gained 2.76 more CLSPM compared to 1.05 more CLSPM for younger children; however, the rate of change for children in the lower age group was 35%, as compared to 27% in the older age group.

Figure 7: *Letter Sound Knowledge Results by Age Group*

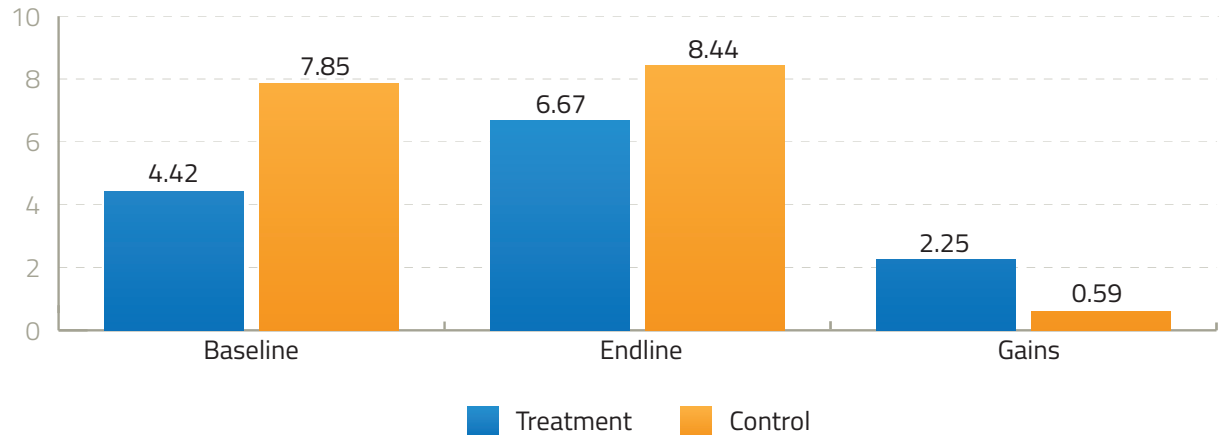


Analysis: Based on the results of the letter sound subtask, it appears the game had a limited effect on improving letter sound knowledge, as compared to Control. One key explanation for these results is the Antura game does not explicitly teach letter sounds; levels one to four of the game focus on letter names, their placement in the word, and diacritics, but not necessarily the sound of the letter. Therefore, treatment children with no prior knowledge, such as those in the 5–7 age group, did not have the opportunity to learn and master letter sounds during the game. For treatment children with previous knowledge, such as those in the 8–10 age group, the introduction of the letter names was likely sufficient to enable children to recall the correct letter sound. Males in the treatment group, although starting from a much lower baseline, made gains that met the male performance in the control group.

Syllable Reading: Another measure of alphabetic knowledge is syllable reading, which assesses children's ability to read syllables with accuracy. Children must apply their knowledge of letter-sound correspondence to various letter combinations, where a letter sound may vary depending on its placement in the syllable or word. The syllable could be composed of one letter and a diacritic or two letters. In this subtask, children were presented with 100 syllables on a sheet and provided one minute to read as many as possible.

As illustrated in *Figure 8,* knowledge of syllables among the treatment and control groups is quite low. Children in the treatment group read an average of 4.42 correct syllables per minute (CSPM) at baseline and 6.67 CSPM at endline, resulting in an absolute gain of 2.25 syllables, but a 51% rate of change. Children in the control group scored higher than the treatment group at baseline reading 7.85 CSPM at baseline and 8.44 CSPM at endline, but demonstrated a rate of change of 8%.

Figure 8: *Syllable Reading Results by Group Type*



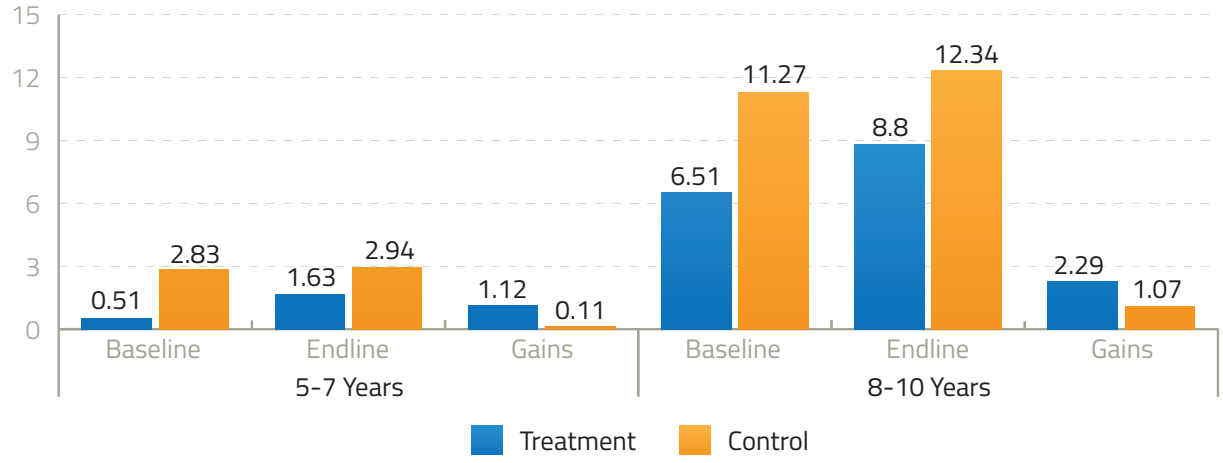
As shown in *Table 21,* boys in the treatment group read more syllables correctly than girls at baseline and endline, and achieved higher gains (3.02 more CSPM compared to 1.23 CSPM for girls), about double that of girls. Similarly, the rate of change was 57% among boys in treatment, compared to 38% among girls in treatment.

Table 21: *Syllable Reading Results by Gender*

Subtask	Group Type	Baseline		Endline		Gains		Rate of Change	
		Male	Female	Male	Female	Male	Female	Male	Female
Syllable Reading	Treatment	5.33	3.22	8.35	4.45	3.02	1.23	57%	38%
	Control	7.10	8.67	8.27	8.63	1.17	-0.04	17%	0%

The syllable reading results by age group (see *Table 19 and Figure 9*) clearly show that in the treatment group, children ages 8–10 had higher syllable reading gains (2.29 more syllables read) than children ages 5–7 (1.12 more syllables read). Comparing older and younger children in the treatment group, older children read 8.8 CSPM at endline, while younger children read only 1.63 CSPM. Children in the control group followed similar patterns, indicating that older children had more exposure to syllables and hence higher performance. However, the rate of change for gains indicates that 1.12 CSPM for younger children represents a 220% rate of change, compared to a 35% rate of change for older children. The results are noteworthy for younger children, especially when taking the dosage into consideration.

Figure 9: *Syllable Reading Results by Age Group*

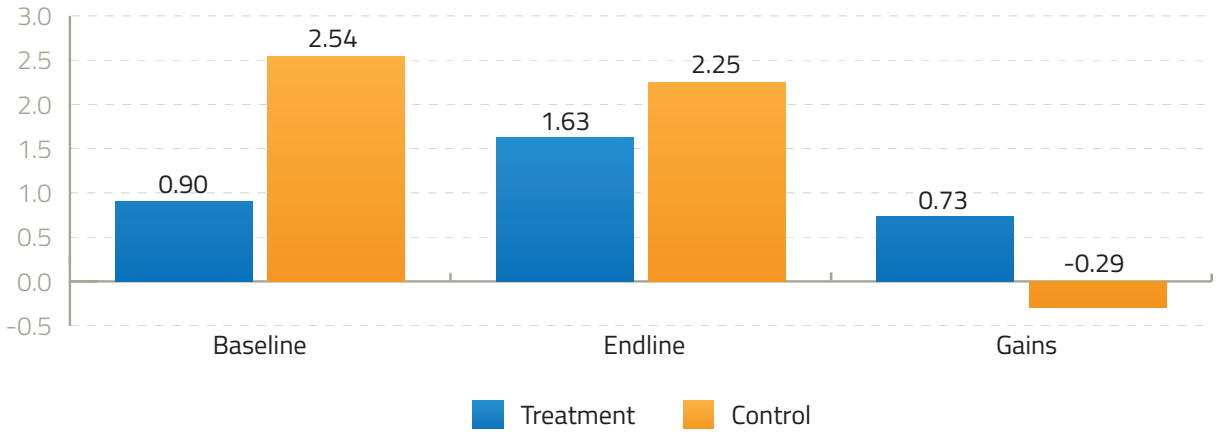


Analysis: Decoding refers to the process of translating a printed word into a sound. For those with prior letter-sound knowledge, these skills taught within the game likely contributed to the gains made in the syllable reading subtask, though the game does not offer sufficient letter-sound game play or instruction for decoding. The syllable reading subtask indicates that 18 more children in the treatment group could identify at least one syllable correctly by endline, which led to a decrease in the number of zero scores. Children in treatment on average identified 7 CSPM, which was 2.4 more CSPM compared to the baseline. Younger treatment children with little or no prior knowledge of syllables learned one more syllable after playing the game, while older treatment children with prior knowledge learned two more syllables. While there were improvements in syllable reading from baseline to endline in the treatment groups, Antura and the Letters did not greatly improve knowledge of syllables for younger children in absolute terms, since it does not teach letter sounds or syllables. The game instead mainly focuses on instruction through letters, letter placement, and building words. Nevertheless, the rate of change among the younger treatment children shows great promise should this trend continue with increased dosage.

Invented Words: To read fluently, children must be able to decode unfamiliar words. The invented word subtask assesses children's ability to apply their knowledge of letter to sound (or phonics) and syllable-to-sound correspondence to decode made-up words. Study participants were given one minute to read from a sheet with 50 invented words that followed the Arabic linguistic structure.

The low mean scores and gain scores in *Figure10* indicate that participants are unable to decode unfamiliar words. The treatment group read 0.9 correct invented words per minute (CIWPM) at baseline and 1.63 CIWPM at endline, demonstrating an increase of 0.73 CIWPM. The control group performed higher at baseline and endline, but showed negative gains of -0.3 CIWPM. Despite these relatively small gains in absolute terms, the rate of change for gains in the treatment group was 81%, while the control group showed a decrease in performance of 11%.

Figure 10: *Invented Word Decoding Results by Group Type*



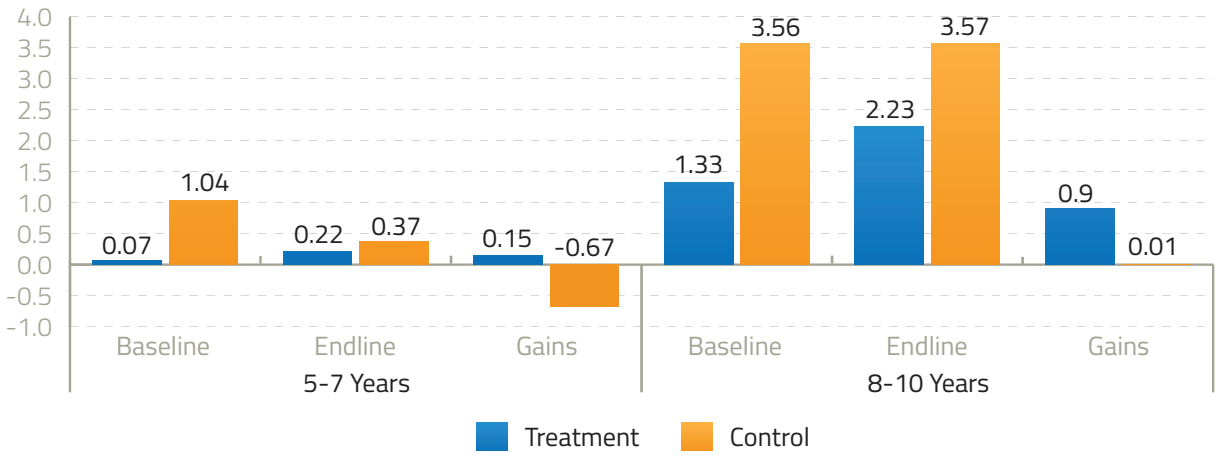
According to the results by gender in *Table 22,* boys in the treatment group scored higher than girls at baseline and endline and had higher gains (1.2 compared to 0.1 for girls). Boys also showed higher rates of change: 116% for boys in treatment, as compared to 13% in for girls in treatment. Boys in control did not demonstrate any improvement, while girls in control demonstrated worsening performance.

Table 22: *Invented Word Decoding Results by Gender*

Subtask	Group Type	Baseline		Endline		Gains		Rate of Change	
		Male	Female	Male	Female	Male	Female	Male	Female
Invented Word Decoding	Treatment	1.06	0.68	2.29	0.77	1.23	0.09	116%	13%
	Control	2.47	2.61	2.47	2.00	0	-0.61	0%	-23%

The invented word results by age group presented in *Figure 11* suggest that treatment children in both age groups showed very minimal gains from baseline to endline. Younger children in the treatment group could decode 0.2 invented words correctly at endline, which was a gain of 0.15, while older children could decode 2.2 invented words correctly at endline, an increase of 0.9 invented words *(see Table 19)*. However, while the absolute gain was small for younger children, it represented a rate of change of 214%, compared to 68% for older children.

Figure 11: *Invented Word Decoding Results by Age Group*



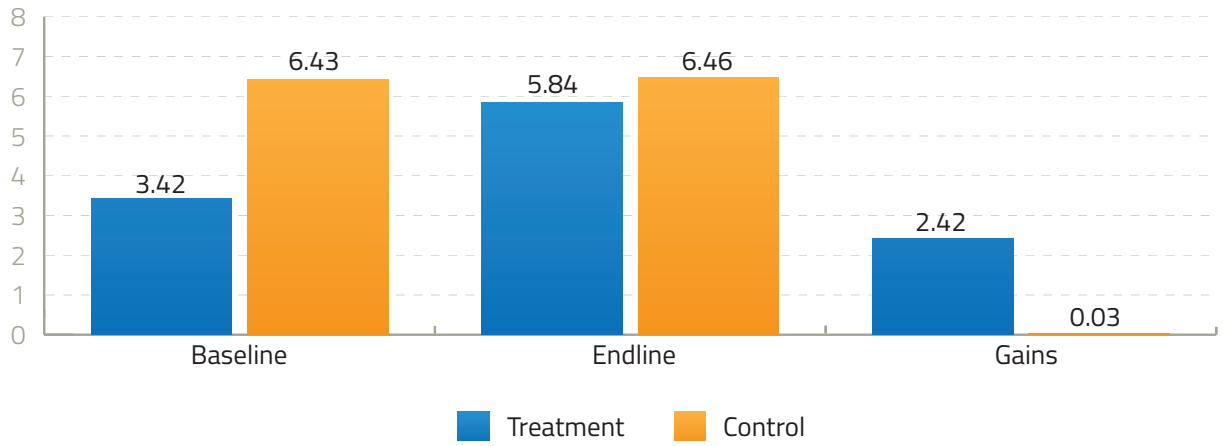
Analysis: The invented word decoding results are directly related to performance on preceding letter sound knowledge and syllable reading subtasks. Without mastery of phonics, children cannot decode unfamiliar words to make larger absolute gains. As the game did not effectively teach letter sound correspondence, children may have faced difficulty in practicing decoding skills—a skill necessary for invented word reading. As a result, there were very limited absolute gains in invented words, but still about one word on average more than the control group. Nevertheless, the rate of change among younger age groups and boys was very high, with rates of change reaching more than 200% among younger children and more than 100% among boys. These improvements show potential for higher absolute gains with perhaps with more time on task.

Oral Reading Fluency (ORF): Fluency is the ability to read text quickly, accurately, and with proper expression to facilitate comprehension. It can be the bridge between word recognition and comprehension. In this subtask, children were asked to read aloud a short story on a printed sheet. The task was timed to one minute.

ORF results in *Figure 12* indicate very low fluency rates among both the treatment and control groups. The treatment group read about half the words than the control group at baseline (3.4 correct words per minute (CWPM), compared to 6.4 CWPM. At endline, the treatment group read 5.8 CWPM, versus 6.5 for the control group. While the treatment group read fewer CWPM than the control group, they had higher gains: children in treatment read 2.4 more CWPM at endline, compared to 0.03 more

CWPM for the control group (for further analysis, see Section 5.2.2 EGRA Summary Results). The rate of change for this gain was 71% for the treatment group, as compared to 0% rate of change for the control.

Figure 12: *ORF Results by Group Type*



In the treatment group, boys performed higher than girls at baseline and endline and showed slightly higher gains (2.74 CWPM compared to 1.97 CWPM for girls). However, the rate of change for both males and females was about the same, as boys showed a 73% rate of change and girls showed a 66% rate of change *(Table 23).*

Table 23: *ORF Results by Gender*

Subtask	Group Type	Baseline		Endline		Gains		Rate of Change	
		Male	Female	Male	Female	Male	Female	Male	Female
Oral Reading Fluency	Treatment	3.77	2.97	6.51	4.94	2.74	1.97	73%	66%
	Control	6.07	6.81	7.12	5.74	1.05	-1.07	17%	-16%

According to results disaggregated by age group in *Table 24,* older children have higher fluency rates than younger children *(Figure 13).* The endline results show older children in the treatment group read 8.04 CWPM, compared to 0.63 CWPM for younger children. Gain scores were also higher for children ages 8–10, in absolute values. Older children in the treatment group read 2.9 more words per minute after playing the game, compared to an increase of 0.39 for younger children. However, the rate of change for younger treatment children was higher than older treatment children (163% compared to 57%).

Table 24: *ORF Results by Age Group*

Subtask	Group Type	5–7 Years				8–10 Years			
		Baseline	Endline	Gains	Rate of Change	Baseline	Endline	Gains	Rate of Change
Oral Reading Fluency	Treatment	0.24	0.63	0.39	163%	5.11	8.04	2.93	57%
	Control	2.08	1.49	-0.59	-28%	9.37	9.97	0.60	6%

Figure 13: *ORF Results by Age Group*

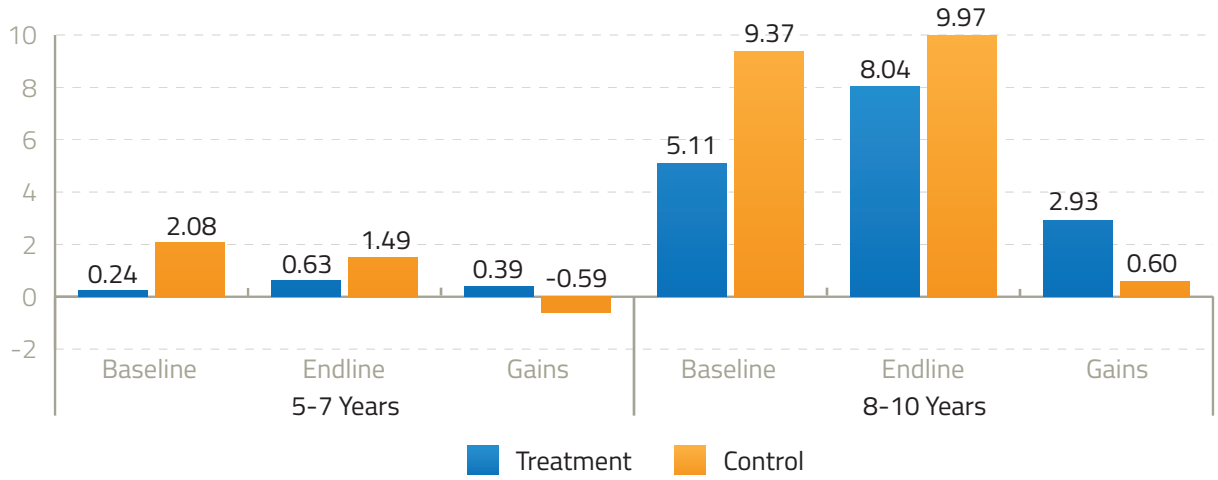


Figure 14: *Boy playing a mini-game in Antura and the Letters.*

Analysis: Beginning readers often rely on sight recognition and memorization to identify words rather than decoding if they have weak decoding abilities. Thus, it is no surprise children had low fluency scores at endline. While the game introduced whole words early and sought to build vocabulary skills, the game did not target the skill level of early children, and did not teach phonics and decoding skills when introducing whole words in early levels. Also, younger treatment children created multiple player profiles to play more at the lower levels. This implies they remained at lower gaming levels and did not progress to higher levels where game play focused more on whole words and fluency. As a result, children with low literacy exposure could not greatly increase their reading fluency in absolute terms. However, the rate of change among younger treatment children was 163% from baseline, indicating great promise. Among older treatment children, ORF gains in absolute terms were larger (almost three more additional words), as compared to about half a word among control children. This indicates children with prior literacy exposure could best benefit from the game to improve oral reading fluency. The rate of change among older treatment children was 57%, the highest rate of change among older children after invented word subtask. The game appeared to reinforce literacy concepts, knowledge of familiar words, word attack and whole word reading for older children, enabling higher gains among older children, and high rates of change among younger children.

5.2.4 Zero scores

There was an overall decrease in the number of children scoring zero on each subtask. The difference was most notable for ORF, where the percentage of zero scores decreased by 11%, from 72% at baseline to 61% at endline, representing about 23 children. However, it is important to note the percentage of zero scores in the control group also decreased by 4% on the ORF subtask, or eight children. It is unknown whether there were conditions in the control group that influenced their improvement or if this was a natural progression.

Moreover, the high percentage of zero scores across all subtasks is of concern *(Table 25).* Over 60% of children in the treatment group and over 50% in the control group could not identify one letter sound,

syllable, or word in a story; and over 75% (75% in control and 80% in treatment) of all children assessed could not decode one invented word. These results indicate that children need explicit instruction in letter sound, decoding, and fluency and more opportunities to practice their reading skills. Given that children in both the treatment and control group samples are out of school children with little or no schooling, these findings may seem natural. Nevertheless, the results identify an important need among this group of refugees.

Table 25: *Number of Children with Zero Scores*

Subtask	Treatment		Control		Difference	
	Baseline	Endline	Baseline	Endline	Treatment	Control
Letter Sound	137	125	107	110	-12	3
Syllable Reading	139	121	117	111	-18	-6
Invented Word	169	162	149	150	-7	1
Oral Reading Fluency	146	123	128	120	-23	-8

Figure 15: *Percentage of Children Scoring Zero on Letter Sound Knowledge Subtask*

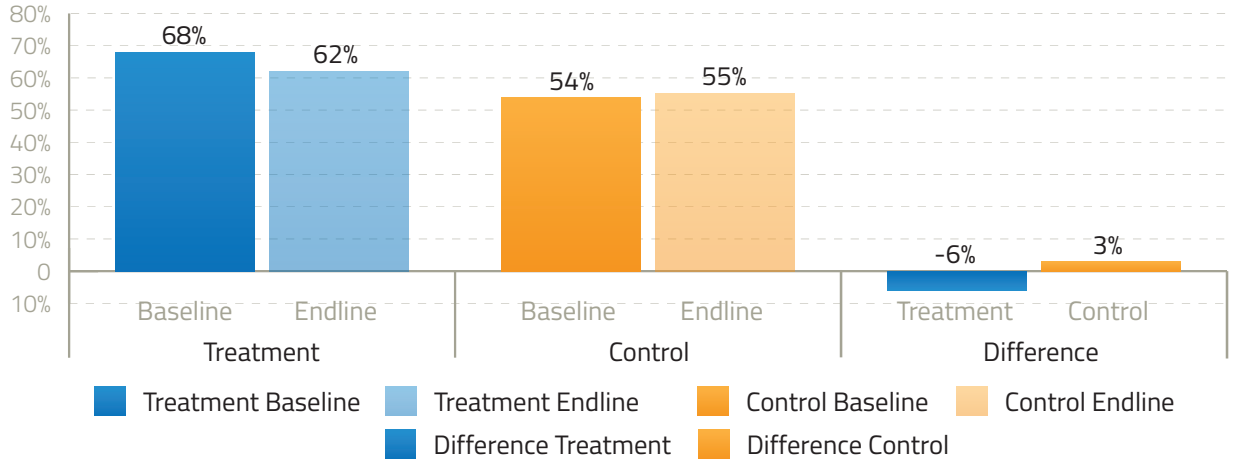
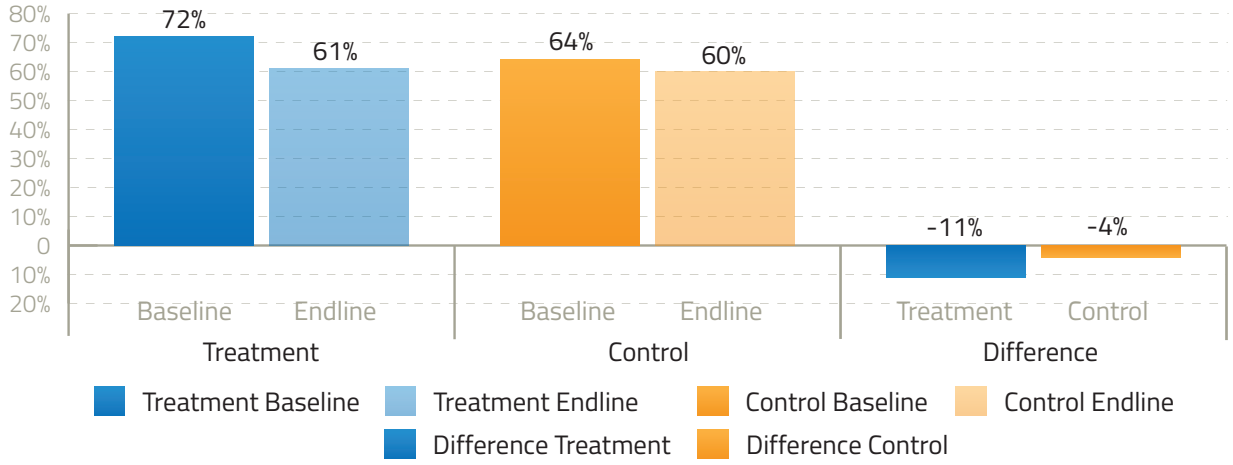


Figure 15 shows the percentage of children who could not identify one letter sound and scored zero on the subtask. Based on the endline results, there were a large percentage of children who scored zero—over 50% for both control and treatment groups. The percentage of children who scored zero in the treatment group decreased by 6% from 68% at baseline to 62% at endline, while zero scores in the control group increased 3% from 54% to 55%.

Figure 16: *Percentage of Children Scoring Zero on ORF Subtask*



On the ORF subtasks, there were a high percentage of zero scores *(Figure 16).* Over 60% of children in the treatment and control group could not read one word correctly and scored zero. The percentage of zero scores decreased by 11% (23 children) for the treatment group (from 72% to 61%) and decreased by 4% (eight children) for the control group (from 64% to 60%).

Reliability

To measure reliability within an EGR Assessment, INTEGRATED ran Cronbach's alpha test after eliminating the zero scores. Cronbach's alpha is an estimate of the inter-subtask correlation of one subtask to another within the same test. In theory, this assesses how consistently each individual subtask measures the common, underlying psychometric construct, which in this case is literacy. Both base and endline EGRA tests had alphas at or lower than 0.8, indicating weak consistency within both tests.

Table 26: *Baseline Reliability*

Item	Obs	Sign	Item-test Correlation	Item-rest Correlation	Average Interitem Correlation	alpha
Letter Sound Knowledge	65	+	0.7979	0.3657	0.6557	0.8510
Syllable Reading	63	+	0.8765	0.5861	0.4189	0.6838
Invented Word Decoding	33	+	0.8501	0.6811	0.4789	0.7338
Oral Reading Fluency	56	+	0.8605	0.6425	0.4129	0.6785
Test Scale					0.5010	0.8006

Table 27: *Endline Reliability*

Item	Obs	Sign	Item-test Correlation	Item-rest Correlation	Average Interitem Correlation	alpha
Letter Sound Knowledge	77	+	0.6880	0.1982	0.6517	0.8488
Syllable Reading	81	+	0.8525	0.6275	0.2565	0.5086
Invented Word Decoding	40	+	0.8561	0.6830	0.3514	0.6190
Oral Reading Fluency	79	+	0.8400	0.6003	0.3048	0.5681
Test Scale					0.4046	0.7311

Correlation

To check for similarity in both the base and endline tests, INTEGRATED conducted rank ordered correlation tests for each of the four subtasks after eliminating zero scores. The correlations below (*Table 28*) show the degree to which endline scores were correlated with baseline scores and vice versa. It is a weak correlation on most subtasks.

Table 28: *Rank-ordered Correlation Coefficients*

EGRA Subtask	Correlation, Baseline to Endline score
Letter Sound Knowledge	0.4318
Syllable Reading	0.5848
Invented Word Decoding	0.4661
Oral Reading Fluency	0.6504

5.2.5 Factors associated with literacy outcomes

Regression Model by Subtask: To provide more reasoning for the quantitative EGRA results, a regression analysis was run to assess the variables that contributed to improved performance for each of the subtasks. The set of variables tested are included in *Table 29.* The evaluation team started by running a regression analysis with all variables, then weeded out non-significant predictors using a stepwise approach.

Table 29: *Variables Associated with Literacy Outcomes*

Variable	Description	Answer Scale
Age	Child's age	■ 5–10 Years
Gender	Child's gender—dummy variable	■ 0 = Males ■ 1= Females
Level	Game level of Antura that child reached	■ 1–6
Last Year of Schooling	Last year child attended formal education	■ 2013 ■ 2014 ■ 2015 ■ 2016 ■ 2017 ■ I didn't attend school
Grade	Last grade level attended by the child	■ 1st Grade ■ 2nd Grade ■ 3rd Grade ■ 4th Grade ■ 5th Grade
Country	The country where the child got his/her most recent education	■ Jordan ■ Syria
Read Alone	Child's self-reported ability to read alone—dummy variable	■ 1=Yes ■ 0=No
Education	Education level of the parents	■ Illiterate ■ Literate but incomplete high school ■ High School ■ Bachelor Degree ■ Vocational Training

For all subtasks, the statistically significant variables associated with improved performance were age and self-reported ability to read alone, and as such the regression analysis shows only those variables with statistical significance. Other variables listed above, such as parental educational attainment and child grade level, did not demonstrate significance in subtasks. The model for each subtask is presented below on *Figures 17–20.*

Letter Sound Knowledge: *Figure 17* below shows that age was the single significant factor associated with higher scores on this subtask, as could also be drawn from the qualitative data. Antura helps older children learn the letter sounds and identify them on EGRA assessments more than it helps younger children.

Figure 17: *Variables Associated with Better Performance on Letter Sound Knowledge*

Source	SS	df	MS
Model	3366.24685	2	1683.12342
Residual	20178.2679	65	310.43489
Total	**23544.5147**	**67**	**351.410667**

Number of obs = 68
F (3, 794) = 5.42
Prob > F = 0.0066
R-squared = 0.1430
Adj R-squared = 0.1166
Root MSE = 17.619

Letter	Coef.	Std. Err.	t	P> I t I	[95% Conf.	Interval]
Age	4.622477	1.57618	2.93	0.005	1.474629	7.770326
Read_Alone	5.031125	4.608651	1.09	0.279	-4.172986	14.23524
_cons	-30.298	13.15708	-2.30	0.025	-56.5745	-4.021506

The coefficient of the age variable suggests that for each year the child is older, an increased knowledge of 4.6 CLSPM is generated. This coefficient is significant where the p-value is 0.05 or less. For the Letter-Sound subtask, age is a significant contributing factor, as its p-value is 0.005.

Syllable Reading: *Figure 18* shows factors that are most significantly associated with better performance on the syllable reading subtask. These were age and self-reported ability to read alone prior to the intervention.

Figure 18: *Variables Associated with Better Performance on Syllable Reading*

Source	SS	df	MS
Model	3649.66331	2	1842.83165
Residual	6188.02787	65	95.2004287
Total	**9837.69118**	**67**	**146.831212**

Number of obs = 68
F (2, 65) = 19.17
Prob > F = 0.0000
R-squared = 0.3710
Adj R-squared = 0.3516
Root MS = 9.7571

Letter	Coef.	Std. Err.	t	P> I t I	[95% Conf.	Interval]
Age	4.451635	.8728507	5.10	0.000	2.708432	6.194839
Read_alone	7.137502	2.552161	2.80	0.007	2.040486	12.23452
_cons	-32.84036	7.286074	-4.51	0.000	-47.39165	-18.28906

The coefficient of the age variable indicates that age is significant (p-value is less than 0.005). The coefficient suggests that for each year older, children can identify 4.5 CSPM more than their younger peers. Also, children who self-reported being able to read alone at baseline tended to identify 7.1 CSPM more than children who reported they could not read alone at baseline, with that variable being significant (p-value is less than 0.05).

Invented Word: *Figure 19* shows factors that are most significantly associated with better performance on the invented word subtask. These were age and self-reported ability to read alone prior to the intervention. The coefficients of these two variables were significant, with both variables' p-values being less than 0.005.

Figure 19: *Variables Associated with Better Performance on Invented Word Decoding*

Source	SS	df	MS
Model	513.28768	2	256.64384
Residual	1132.59467	65	17.4245334
Total	**1645.88235**	**67**	**24.5654083**

Number of obs = 68
F (2, 65) = 14.73
Prob > F = 0.0000
R-squared = 0.3119
Adj R-squared = 0.2907
Root MSE = 4.1743

Letter	Coef.	Std. Err.	t	P> | t |	[95% Conf.	Interval]
Age	1.721642	.373423	4.61	0.000	.975865	2.46742
Read_alone	2.425128	1.091865	2.22	0.030	.2445223	4.605734
_cons	-13.04587	3.117128	-4.19	0.000	-19.2712	-6.82053

The coefficients of the age variable show that with each year older, children can identify 1.7 CWPM. This small change seems to indicate difficulties with this subtask. Also, children who self-reported being able to read alone at baseline could read 2.4 more CWPM than those who reported they were unable to read by themselves at baseline.

Oral Reading Fluency (ORF): Lastly, *Figure 20* shows variables associated with oral reading fluency. Age and self-reported ability to read alone are again significant variables associated with children's achievement on this subtask.

Figure 20: *Variables Associated with Better Achievement on Oral Reading Fluency (ORF)*

Source	SS	df	MS
Model	3897.58987	2	1948.79494
Residual	4913.52777	65	75.592735
Total	**8811.11765**	**67**	**131.509219**

Number of obs = 68
F (2, 65) = 25.78
Prob > F = 0.0000
R-squared = 0.4423
Adj R-squared = 0.4252
Root MSE = 8.6944

Letter	Coef.	Std. Err.	t	P> | t |	[95% Conf.	Interval]
Age	4.962074	.7777868	6.38	0.000	3.408726	6.515421
Read_alone	5.485196	2.2742	2.41	0.019	.9433063	10.02709
_cons	-36.51477	6.492533	-5.62	0.000	-49.48126	-23.54829

The coefficients of the variables associated with achievement on this subtask suggest that children tended to read 5.0 to 5.5 more CWPM if they are older and reported being able to read alone at baseline, respectively. These variables are significant, as their p-values are less than 0.05.

5.3 Qualitative Findings

5.3.1 Antura and the Letters game play

Players entered the game by choosing a human profile. Each player profile started from Level 1, and many children played the same level with multiple profiles. The game asks for the age of the player but does not adjust for skill or literacy level. Players start at the same level and progress based on their achievement. Antura and the Letters divides the game into six levels called Journey Maps, with each level introducing a skill and reinforcing the skill through repetitive but varied game play. There are more than 20 varied mini-games introduced over the six levels. Players begin on Level 1 with an introduction of letter names, shapes, and sounds, then work progressively through letter placement and diacritics in Levels 4–5, and end with spelling out common phrases and vocabulary words in Level 6. *Table 30* shows a more detailed description of each level.

Table 30: *Levels in Antura and the Letters*

Level 1	Introduces the letter names in groupings of two to three, and reinforces their knowledge through the alphabet song, letter name, and shape matching.
Mini-games included	■ Letter name and letter shape match (player matches correct letter tile to requested letter) ■ Sick letters (player removes extra dots and diacritics from letter tile) ■ Rocket letter trace (player traces letters with finger) ■ Letter color trace/tickle (player colors in the letter on the letter tile) ■ Egg hatch (player hatches egg by pressing on the correct letter requested until egg hatches) ■ Hiding letters (player must find the requested letter among those hiding in bushes) ■ Dancing letters (player must place dots in correct place on letter tile) ■ Letter catapult (player must shoot down the requested letter using the catapult) ■ Mouth game (player must place the letter in the mouth from which the letter emerges) ■ Missing letter in word (player must choose/identify the missing letter in the word) ■ Letter in the word (player must select and place the missing letter in the word) ■ Balloon word game (player must eliminate whole words that do not contain the requested letters) ■ Shared letters in paired words (player must identify the shared letter in whole word pairs) ■ Tube game (player is given a whole word and must place that word in the correct letter tube—where there are four tubes and only one tube has a letter that is contained in the whole word)

Level 2	Introduces the letter names and their shapes and how they change shape at the beginning, middle, and end of a word, for the first half of the alphabet. The same games as above are repeated but with some variation.
Game features introduced	■ Missing letter in word (player must select and match one from a multitude of letters) ■ Missing letter in word (player matches the same letter in two words, but letter shape changes due to placement—beginning, middle, or end of word) ■ Picture spelling (player is introduced to a word with its picture and must spell the word letter by letter but does not connect the letters—letters are presented in the alphabetic form— not connected form) ■ Alphabetic order (player must organize groupings of jumbled letters in alphabetical order and ensure they are right-side up) ■ Balloon letter game (player must pop balloons holding up letters not matching the requested letters)
Level 3	Continues from Level 2 to complete the alphabet. The same games as above are used, with some modifications.
Game features introduced	■ Letter name with letter shape match (player matches correct letter tile to requested letter and/or the correct letter with diacritic) ■ Letter-word match (player must select and match the correct letter with the correct letter or correct letter with diacritic in the word) ■ Mouth game (player must place the letter—sometimes with its diacritic—in the place in the mouth from which the letter emerges) ■ Catapult game (player must shoot ball at letters requested—some letters have diacritics) ■ Tube game (player is given a whole word and must place that word in the correct letter tube — where there are four tubes and only one tube has a letter that is contained in the whole word) ■ Word-picture match (player must match a word to its picture) ■ Conveyor belt (player must match a whole word to its picture on a suitcase before the word falls off the conveyor belt; sometimes the word turns around after a few seconds so the word cannot be read, but only when the audio function is activated, which doesn't always work) ■ Balloon image game (player must match image to word and must eliminate word/images not requested)
Level 4	Introduces the main diacritics and reinforces the letters and their placement in whole words, spelling, and diacritics. The same games as above are used, with minor variations.
Game features introduced	■ Missing letter in word with picture (player must place missing letter in word—picture is offered) ■ Word-picture match (player must match word to picture) ■ Catapult picture game (player must shoot the picture corresponding to the word offered) ■ Conveyor belt game (player must match the image to the word on the conveyor belt—word turns around immediately; speed is much faster than in previous levels)
Level 5	Continues to introduce more specialized diacritics, articles, and some vocabulary (adjectives), and reinforces the letters and their placement in the whole word, as well as spelling and diacritics learned in previous levels. The same games as above are used.

Level 6	Moves children into common phrases (greetings), as well as common vocabulary (weather, days of the week, months, years, colors, etc.), and reinforces the letters and their placement in the whole word as well as spelling and diacritics learned in previous levels. The same games as above are used, with the new games below introduced.
Game features introduced	■ Missing whole word match (player must pick the missing word matching the audio sentence and the existing word on the screen) ■ Read the hidden words (player is asked to follow the word reading with the magnifying glass; after reading the sentence, player must choose the correct picture/image corresponding to the sentence read) ■ Dancing letters (player must place diacritics in correct place on letter tile) ■ Question/statement and answer (player must answer the question by choosing a correct answer tile—questions and options have audio options)

Figure 21: *Corrective feedback in the mini-games.*

Each level consists of 45 stages, in which the player plays up to three mini-games per stage, and Antura the dog marks the place the player has reached on the map. Players must complete each stage within the level in its designated order, and must complete the entire level before moving to the next level. Players receive up to three bones upon completion of each game, depending on performance. Player game completion and performance also unlocks accessories for Antura the dog to wear, enabling customization of the dog in the Antura play space. Each level takes at least three hours of continuous and progressive game play to complete. Children could reach level 2.1 on average within the allocated dosage (26.6 hours on average). Twelve children reached Level 5 (six females and six males), but no children completed the game.

Corrective feedback is usually given by a red X appearing for incorrect answers and a green check mark appearing for correct answers (*Figure 21).* Most mini-games give three opportunities to self-correct. In addition, the game begins by demonstrating both game play and correct answers. Some mini-games offer infinite (or only time-bound) opportunities to self-correct, such as the drawing of letters. Other mini-games offer letter matching in which a child can try all letters until the correct one is found. In this sense, guessing can be rewarded. The letter drawing mini-game requires a high level of precision. The same is true of many timed mini-games, where it can be challenging to receive three bones within the time allotted.

Antura and the Letters is intended to provide a staged, step-by-step, and repetitive yet reinforcing approach to language acquisition. In the first three levels, players are introduced to letter names, shapes, and variations in groups of two to three letters. The game does not introduce letter sounds. Mini-games at the letter recognition level sometimes introduced letters not yet covered as well as whole words within complex mini-games. For example, the balloon mini-game *(Figure 22)* introduces

whole words, and the player needs to be able to read the words to eliminate those (up to nine words on the screen) that do not contain the requested letters (up to four letters) within a set period. Other whole word mini-games are introduced in Level 1, where the player has not yet been introduced to how the letter looks in the beginning, middle, or end of the word but is expected to identify the correct letter within that word. This may be too challenging at this stage of language acquisition, when letters are still being learned.

Figure 22: *Balloon Mini-Game*

Nevertheless, some games in Level 2 reinforce letter recognition and spelling. For example, whole words are introduced with pictures, and the player is asked to spell the word letter-by-letter but not yet with its connections. In other games, players are asked to match the same letter within two words where its shape changes due to word placement, reinforcing the aim of Level 2. The alphabetic order game reinforces the alphabet acquired in Level 1.

In Levels 4 and 5, players are introduced to diacritics, articles, and basic vocabulary words. The alphabet-with-diacritics song, first introduced in Level 3 *(Figure 23)* and repeated in subsequent levels, gives a repetitive and clear basis for the introduction of diacritics in Level 4. In Level 5, the player is introduced to some additional diacritics (double diacritics), articles, special letters (such as الألف المقصورة), sun and moon consonants, and some new vocabulary.

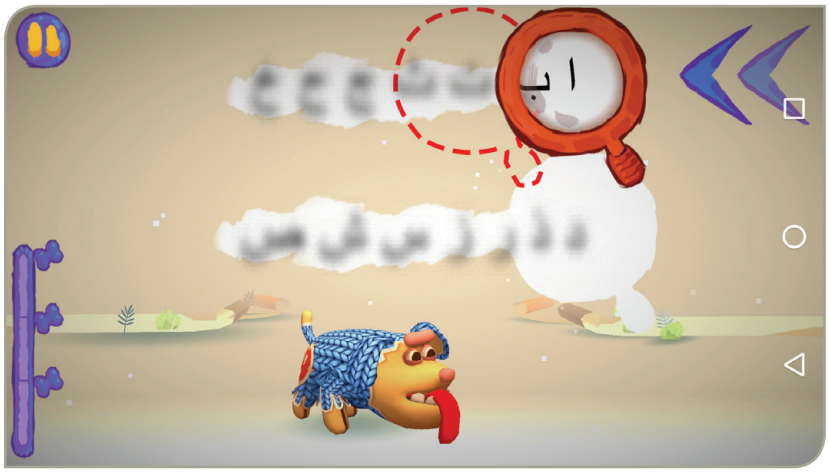


Figure 23: *Alphabet song*

In Level 6, the player is expected to read whole words and sentences and play mini-games that test reading comprehension. This level is quite advanced and can likely only be achieved by those children with previous literacy history. Children testing the game did not reach this level, so child reaction or ability could not be assessed.

The words exposed to the children through the game are inspired from their cultural surroundings and allow the player to expand their vocabulary. Words are accompanied by pictures in some mini-games, which helps reinforce meaning. However, it may be unclear to children at the beginning of the game how many levels there are and when the game will finish.

5.3.2 Children's literacy interface with Antura and the Letters

Children reported in FGDs they learned how to read, write, and pronounce letters as well as read the supplementary diacritics of words. They also stated they could recognize letters and use them to form words. One child stated, "I can cut up the letters, join them together, and I learned how to spell the word rabbit". Children also reported that peers—particularly those who were younger or struggled with reading—are now better able to correctly identify and recognize letters.

This general progress in gaining literacy skills was also reflected in engagement observations and focus groups with parents. In 86% of engagement observations, the child could answer 50% or more of the questions presented in the game correctly. Seventy out of 80 parents confirmed their children were learning to read better through the game. Parents also expressed their children now knew letters and could spell better. Fifty-eight parents said their children wanted to learn more after starting the game. Some parents observed their children reading signs or advertisements or reading on the phone. One mother of a 7-year-old stated, "My husband is away, and my 7-year-old learned to read; her father now sends her messages on my phone and she is able to read them on her own."

Figure 24: *Observing game play.*

Volunteers/teachers also noticed improvement in the literacy skills of children in their classes. One volunteer/teacher noted that children in his class began recognizing letters and spelling after 20 days with the game. Volunteers/teachers also said the children benefited particularly from the alphabet song. Another volunteer/teacher said the children in her class can now distinguish between letter shapes at the beginning, middle and end of the word. All (10/10) volunteers/teachers in focus groups agreed that children in their classes have improved letter recognition and vocabulary. Some volunteers/teachers noted that children in their classes are now reading the names on their tablets and messages on their parent's phones. Words used in the game were unanimously (19/19) seen as relevant to real world word usage by children in play-focused interviews.

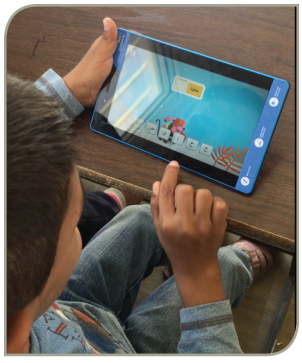


Figure 25: *A child playing an Antura mini-game.*

While general literacy gains were noted by volunteers/teachers, parents and other participants in the study reported some gaps in literacy skills. When asked specifically about what their children learned, parents mentioned letters, letter sounds, syllables, and words, but added that their children did not learn how to combine, link, and create sentences. One parent said that learning through technology made children forget how to write, since they can write on the mobile device but not with pen and paper. Parents in that specific focus group confirmed this message. Only one out of 10 volunteers/teachers noticed improvements in writing or reading fluency among the children in his class.

Children also reported they would like to continue using traditional methods (such as pen and paper) for learning, as writing things out was said to help them learn faster and promote their ability to write. Additionally, some children reported that use of tablets was "distracting and would take away my time from studying." During focus group discussions, children said a combination of both methods (traditional and technology-based) would be best for learning.

Meanwhile, children seemed to agree that the first level of the game matched their level of literacy. Other children reported that their completion of the first stage encouraged them to continue to attempt higher levels. One child stated, "I learned from the first stage, and I got encouraged to keep playing." In gaming observations, children were observed as 34% more engaged with the game's learning content during later game play. The overall level of learning content engagement throughout the playing period was high, at 83%.

However, children were not able to progress through all stages, suggesting the game may not be well matched to the users' skill levels. Some children noted in FGDs that as the stages progressed, so did the level of difficulty. Tablet data collection revealed that children had created multiple profiles and played early levels multiple times with different avatars. On average, children reached level 2.1 out of six, and no child finished the game. Many children encountered difficulties with game mechanics or struggled to understand how the level was meant to be played, even in well-liked levels. As such, while learning outcomes seemed to be unaffected by issues like engagement or frustration, learning outcomes may have been negatively affected by issues of game play. Gaming observations found the difficulties that children encountered were evenly split between the game mechanics (45 occurrences) and learning content (46 occurrences). Within these observed difficulties, there were 29 major difficulties with learning content, where children got more than one question in a row wrong.

Younger children (ages 5–7) reported more challenges with the game in terms of completing stages. One child reported "the dog comes and licks the letters too quickly, I go to swipe and I can't; it's too quick". Younger children also had more issues with the learning content in the Sling/Color Tickle mini-game (four occurrences for young and zero for old). Volunteers/teachers noted that the literacy levels of the children in their classes are mixed, and most had no educational foundation. Volunteers/teachers noted larger improvements for children with no previous literacy than for those who already knew how to read. One volunteer/teacher said she felt the game is best suited for an 8-year-old child. No differences were observed in the ability of female and male children to advance through levels.

Volunteers/teachers suggested the game add numeracy components and stories. There was also a request that the game be translated to support for English literacy.

5.4 Conclusions

Antura and the Letters effectively introduces children to the basics of Arabic literacy, as lessons correspond to the basics of Arabic literacy acquisition. The levels are appropriately staggered such that basics are taught first, with levels building on concepts introduced in previous levels and new concepts taught within clusters. However, many complex mini-games are unnecessarily introduced too early and do not correspond to the level of instruction; this can lead to frustration and guessing when the child has no previous literacy exposure. For example, mini-games in early levels introduce letters not yet covered within whole words and therefore could improve with the inclusion of mini-games more

appropriately matched to early learners. Whole words are regularly longer than three letters, which also may make it challenging for non-readers to play the game without guessing. While the approach is intended to provide a staged, repetitive, and reinforcing approach to language acquisition, many of the mini-games do not effectively teach or reinforce the lesson under review, or introduce whole word reading before the full alphabet, letter variation, and diacritics are introduced. In this sense, the player tasks within mini-games are not unified to the level and lesson. The game could benefit from increased tailoring of mini-games to the level/lesson or letter/diacritic under review through restricting the options from which the child can select a correct answer. While mini-games in higher levels could improve by decreasing difficulty, insufficient testing of game play of Levels 5 and 6 prevented a full analysis of these higher levels.

Nevertheless, game use resulted in positive learning outcomes and a decrease in zero scores across all age groups and genders with a relatively low dosage of 27 hours. Absolute gains in invented words were lowest, confirming that decoding among the entire treatment group remains weak. Gains in letter, syllable, and ORF were about an additional two to three letters/syllables/words in absolute terms, with rates of change being highest in ORF and invented word. Gains among older children were greater than those among younger children across all subtasks. As noted in the qualitative findings, the game seemed to reinforce previous learning. The game also seemed to support literacy improvements best among those children with some prior literacy knowledge, as demonstrated by the statistical significance in variables through regression analysis of age and self-reported ability to read alone.

More than 80% of children who improved from baseline to endline were in the 8-10-year-old age group across all subtasks, demonstrating that for older children, the game is helping their reading fluency and is reinforcing knowledge they already have. Absolute gains were small among younger children, however. The lack of letter sound in the game—a key element of decoding—likely explains why early or younger children were unable to make larger gains in letter sound, syllable reading, invented words, and ORF. Nevertheless, the rates of change among younger children—ranging from 35% (letter) to 220% (syllable)—were larger than those of older children, indicating that the game shows promise for the younger age group. Absolute gains may also increase if Antura and the Letters added decoding and allowed for more time on task.

The DiD results did not demonstrate statistical significance at the 5% level with the ORF gains between the treatment and control groups, because of game play. If the game is revised to address phonics and syllable construction—based on rates of change experienced among younger children—it is likely that literacy outcomes in later testing treatments would demonstrate greater gains among younger children with no previous literacy exposure. Nevertheless, given the dosage of 27 hours, the difference in gains between treatment and control demonstrates that gains can be made in literacy with a relatively low gaming dosage.

The game is very long, as demonstrated by the fact that no children completed the game and only reached level 2.1 on average, despite an average dosage of 26.6 hours. While the game presents many mini-games that are appropriately leveled for literacy, the appropriateness of pace, especially for the younger children, could benefit from revision. This is especially true for mini-games that require fine motor skill development or higher-order literacy skills. The challenges presented to these children may have resulted in the lack of leveling progress among younger children and behind the creation of multiple avatars to play the same level repeatedly.

Finally, it is also important to note the gender differentials observed, whereby boys tended to have much higher gains and rates of change across subtasks than girls. Regression analysis within each subtask did not demonstrate gender being a significant factor in influencing subtask outcomes. Nevertheless, girls tended to cluster slightly more than boys in first grade (54% of girls in the sample were in Grade 1) or second grade (29%), as compared to boys in first grade (50% of boys in the sample were in Grade 1) or second grade (27%). Beyond these gender differentials, qualitative and quantitative findings could not explain gender differentials in performance. This remains an area for further investigation.

The high rate of smartphone penetration among Syrians participating in the control and treatment groups has promising implications for the widespread download and use of literacy games like Antura and the Letters.

5.5 Recommendations

Antura and the Letters could benefit from more focus on teaching letter sounds and building syllables before building words, with a greater focus on phonics and decoding. Specifically, with respect to skills the EGRA subtasks assessed, the game in its current form lacks phonics as a key element for decoding skills, which are essential for improving literacy outcomes.

As the results by age group differed greatly, the game could also benefit from improved pacing for younger children. It may also benefit from the simplification of early levels to focus on shorter words, easier tasks, more time per mini-game, and improved matching of mini-games to levels. This should include using letters/diacritics already introduced, making use of short three-letter words, and matching instruction and game play to the lesson under review.

The game in its current form is too long and could benefit from review of pacing and removal of complex mini-games. This could include removal of randomization of mini-games, at least in early stages, and presentation of more challenges in later stages.

Specific Recommendations for Children with Low Literacy Exposure

- Introduce letter sounds for reinforcement of letter-sound recognition versus letter-name recognition.
- Reduce complex mini-games in early levels, such that the child can feel a sense of achievement earlier in completing a level. Play time can be reduced if complex mini-games are removed from early levels, which may give the child a sense of achievement and desire to continue to the next level.
- Consider making use of three-letter rhyming endings of words or syllable segments when building whole words and changing only the first letter to reinforce blends at an early stage, before moving into introducing longer or more complex words. For example, the game could make use of simple rhyming pairs**.** Use the same concept with diacritics, where the diacritic pattern within three-letter words is the same to reinforce diacritic impact on letters.
- Slow pacing in early stages to allow early learners sufficient time to play with accuracy over speed. Speed of play can be increased as one of the challenges as the child advances.

- Increase letter tracing games in early levels for reinforcement of learning. Ensure that activities within the stages consider fine-motor skill manipulation of initial stages/younger age groups. Increase time allocated for specific activities that involve certain motor abilities to promote successful manipulation by younger children.
- Introduce letter tracing with diacritics matched to sound of letter with diacritic.
- Remove randomization of mini-games in early stages, as the randomization function disrupts game flow when mini-games are too challenging for early learners.
- Reduce the sensitivity of letter tracing to allow for less accuracy.

Specific Recommendations (Regardless of Level)

- Improve matching of mini-games to the level of instruction. The game moves from easy mini-games dealing with single letters to complex tasks dealing with multiple words within the first level. The game should match the more complex mini-games with higher levels, as the child may not have the required skill set to play complex mini-games without guessing.
- Consider only introducing words using letters already introduced. When introducing diacritics, consider only using letters with diacritics already introduced. Ensure that lessons/gaming are staggered to build on previously introduced skills.
- Revise the balloon mini-game to have fewer words, shorter words, and groupings of words that correspond to the letter grouping introduced. Do not introduce words that have letters not yet introduced, such that the mini-game builds on letters acquired.
- Revise the mouth mini-game to be better understood by the child. Consider improving the graphics to better resemble a mouth. Consider demonstrating the objective of the mini-game, should it be kept.
- Review mini-games to better focus on the lesson under review. For example, when introducing diacritics, the letter placement mini-game could focus on the same diacritic with different letters to focus on the effect of the diacritic on the letter. Alternatively, in the letter placement or tube mini-games, the game could better focus on ensuring the child selects letters from a host of options of the same letter with different diacritics, to reinforce the diacritic lesson.
- Consider removing audio cues within some mini-games, such as the image match and balloon games, at later levels to reinforce reading by the child. When the audio is removed, consider allocating more time for game play.
- In the catapult mini-game, consider making all options be the same letter but with different diacritic variations, so the objective is to target the correct diacritic—not the correct letter.
- Introduce options for diagnostics of skills introduced that remove guessing.
- Consider allocating differing play times for differing tasks, as some tasks may need more time than others. For example, consider allocating more play time for letter matching and word reading tasks to enable children to complete full words.

Photo Credit: Norad/Marit Hverven

6. Psychosocial Outcomes

6.1 Measuring Psychosocial Outcomes: Methodological Approach

INTEGRATED employed a dual approach to evaluation: impact evaluation and technical evaluation.

6.1.1 Impact evaluation

The impact evaluation used a longitudinal quasi-experimental design to estimate the impacts of the EduApp4Syria on children's psychosocial outcomes over time, comparing growth in psychosocial outcomes for children using the game to children in matched environments who did not receive the game.

Hypothesis: If the game is applied five to six times per week for up to 30 hours, then psychosocial outcomes of children using the game will improve more than those who did not use the game.

The hypothesis assumes the treatment and control groups displayed comparable traits before the game was implemented, and the conditions existed to confidently attribute observed changes in children's psychosocial well-being outcomes to the use of the game, rather than other factors.

The data trends among the treatment group are potentially indicative of overall wellbeing improvements within three of the five domains assessed.

The primary instrument used to assess psychosocial well-being and outcomes was the Strengths and Difficulties Questionnaire (SDQ), designed to measure children's strengths, difficulties, and pro-social behavior. SDQ was tested and translated into Arabic by Save the Children-Jordan in camp and host community settings among Kindergarten (KG) students in Jordan in 2013. It is a brief behavioral screening questionnaire designed for 3-16-year-olds and measures 25 psychological attributes:

- Emotional Symptoms (5 items)
- Conduct problems (5 items)
- Hyperactivity/inattention (5 items)
- Peer relationship problems (5 items)
- Pro-social behavior (5 items)

The SDQ tool focused on assessing the impact of the game on psychosocial outcomes in specific domains, which have been matched to EduApp4Syria criteria by the research team. *Table 31* draws parallels between each of the five SDQ Scales and the various psychosocial criteria of EduApp4Syria:

Table 31: *SDQ Scale*

EduApp4Syria Criteria/ SDQ Scale	Emotional Symptoms	Conduct Problems	Hyperactive/ Inattention	Peer Relationships	Pro-social Behavior
Happiness	X				
Ability to Play				X	X
Cognition			X		
Sadness	X				
Stress Reactions	X				
Somatic Health	X				
Helplessness, Clinginess, Independence					
Ability to Regulate Emotions		X		X	X

The SDQ, which is available in both the English and Arabic languages, was administered with a subset of treatment children's parents, as follows:

- Parents of children ages 5-10 who are using the games (treatment group)
- Parents of children ages 5-10 who are not using the game (control group)

Table 32: *Participation among Parents*

SDQ (Parents)	Baseline	Endline	Attrition
Control	123	47	76
Antura Treatment	127	53	74

Sampling of parents was conducted through a convenience sample of parents of children selected to participate in the literacy testing of the game, aiming to reach a 95% confidence level and 5% confidence interval of the sample child population. All parents of children were invited to participate. Participation among parents *(see Table 32)* at baseline was less than targeted due to a variety of factors, including lack of adequate outreach to parents, ongoing humanitarian distributions during baseline, and lack of desire to be surveyed.

6.1.2 Technical evaluation

The technical evaluation focused on questions designed to assess strengths and weaknesses of each game in psychosocial outcomes and engagement, as well as provide a qualitative interpretation to the quantitative results. INTEGRATED, in collaboration with CREATE, employed a participatory approach to the technical evaluation using mixed methods that utilized a combination of qualitative and quantitative methods. This composite qualitative and quantitative methodology included FGDs, observation checklists, and KIIs, whereby methods and their composition were selected to enable the collection of primary and secondary data for sufficient depth and breadth. The methods were designed to interlink, allowing the triangulation of data to produce a verifiable body of evidence. The technical evaluation sought to:

- Assess the impact of the game socio-emotional well-being;
- Assess the level of interest and enjoyment among children playing each game (and identify ways in which the games could be improved); and
- Examine how parental engagement may impact the use/engagement in the game.

The evaluation methodology consisted of: a desk review of DGBL on psychosocial outcomes and engagement; semi-structured focus group discussions with child beneficiaries and their parents; in-depth interviews and FGDs with RI staff and volunteers, observation of children using the games for gaming assessment and engagement, and play-intensive interviews with child beneficiaries. Stakeholders groups and the data collection effort are clarified in *Table 33*:

Table 33: *Evaluation Stakeholders and Data Collection to assess the Impact Psychosocial well-being*

Stakeholders	Data Collection Conducted
Donors	■ 3 KIIs
Implementing Partners (RI)	■ 1 KII with Relief International Head of Education ■ 1 FGD volunteers/teachers
Parents of Children in Treatment	■ 10 FGDs Antura
Children in Treatment	■ 10 FGDs Antura ■ 118 engagement observation ■ 43 Gaming observation ■ 22 Play-focused interviews for each game
Game Designers	■ 1 KII

6.2 Limitations

Attrition. Attrition was high among parents, starting with a lower baseline number than planned. This limits the depth of analysis possible as well as the confidence with which analysis can be extrapolated to the group.

Moving. Village 5 residents (Control) were asked to move homes in August, prior to the endline SDQ data collection. As a result, the psychosocial conditions between control and treatment groups cannot be considered the same over the testing period.

6.3 Strengths and Difficulties Questionnaire Results

SDQ Analysis: The SDQ data was analyzed using the SDQ software.[50] Each subscale is scored from Zero to 10. The total difficulty score ranges from 0-40 and is calculated by adding the emotional, conduct, hyperactivity, and peer problems subscales. The internalizing and externalizing subscales range from 0-20. The SDQ is a behavioral screening questionnaire for children and adolescents, consisting of 25 items that measure five subscales: emotional symptoms, conduct problems, inattention/hyperactivity, peer relationship problems, and pro-social behavior. Each of the five measures is classified as "normal, borderline, and abnormal." The "pre" and "post" SDQs can be used to indicate everyday practices by children. Studies using SDQ along with research interviews and clinical ratings have proven the tool's sensitivity to treatment effects. Child and adolescent mental health professionals find the SDQ useful in designing interventions to enhance the psychosocial well-being of these age groups.

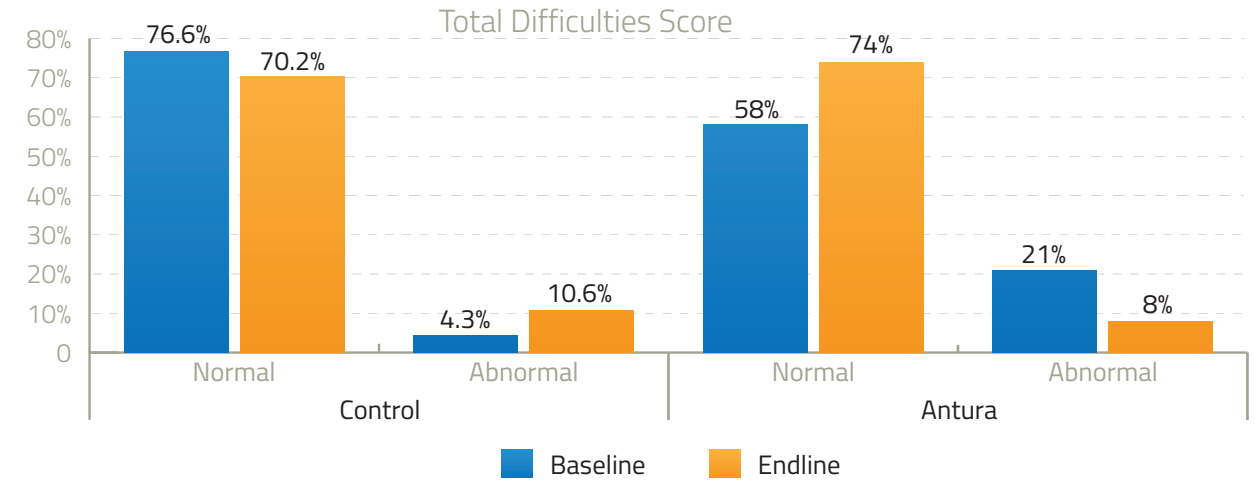
The analysis aims to group students according to the category of psycho-social assessment (e.g. Hyperactivity).[51] *Table 34* shows the results of the SDQ:

Table 34: *SDQ Results*

Overall	Baseline			Endline			Baseline			Endline		
	Antura and the Letters						Control					
	Normal	Borderline	Abnormal	Normal	Borderline	Abnormal	Normal	Borderline	Abnormal	Normal	Borderline	Abnormal
	53						47					
Emotional Symptoms	73.6%	7.5%	18.9%	84.9%	7.5%	7.5%	91.5%	4.3%	4.3%	80.9%	4.3%	14.9%
Conduct Problems	67.9%	17.0%	15.1%	79.2%	11.3%	9.4%	80.9%	8.5%	10.6%	83.0%	6.4%	10.6%
Hyperactivity/ Inattention	60.4%	11.3%	28.3%	60.4%	11.3%	28.3%	63.8%	21.3%	14.9%	68.1%	12.8%	19.1%
Peer Relationship Problems	67.9%	24.5%	7.5%	81.1%	17.0%	1.9%	87.2%	12.8%	0.0%	76.6%	19.1%	4.3%
Pro-social Behavior	88.7%	3.8%	7.5%	90.6%	3.8%	5.7%	97.9%	2.1%	0.0%	91.5%	8.5%	0.0%
Total Difficulties Score	58.5%	20.8%	20.8%	73.6%	18.9%	7.5%	76.6%	19.1%	4.3%	70.2%	19.1%	10.6%

Parents of children engaging the game (as well as the control group) were interviewed at baseline and endline intervals to analyze the difference in overall psychosocial scores within the five subscales. Analysis of results comparing the baseline (pre-test) and endline (post-test) revealed that following the testing period, children's scores within all domains except pro-social behavior increased among treatment children. Overall, children in control and treatment groups displayed normal scores for most of the five categories assessed by SDQ at baseline. However, normal total difficulties scores of the treatment group improved by about 15% between baseline (58.5%) and endline (73.6%), as compared to a decrease in total difficulties scores among the control group between baseline (76.6%) and endline (70.2%), as shown in *Figure 26:*

Figure 26: *Total Difficulties Score*



Within individual domains among treatment children, normal scores within emotional symptoms increased by 11.3%; within conduct problems, normal scores increased by 11.3%; and in both domains, abnormal scores were reduced in the treatment group. By comparison, within the control group, normal scores within emotional symptoms decreased by 10.6%, and normal scores within conduct problems increased by 2.1%, as shown in *Figures 27 and 28:*

Figure 27: *Emotional Symptoms*

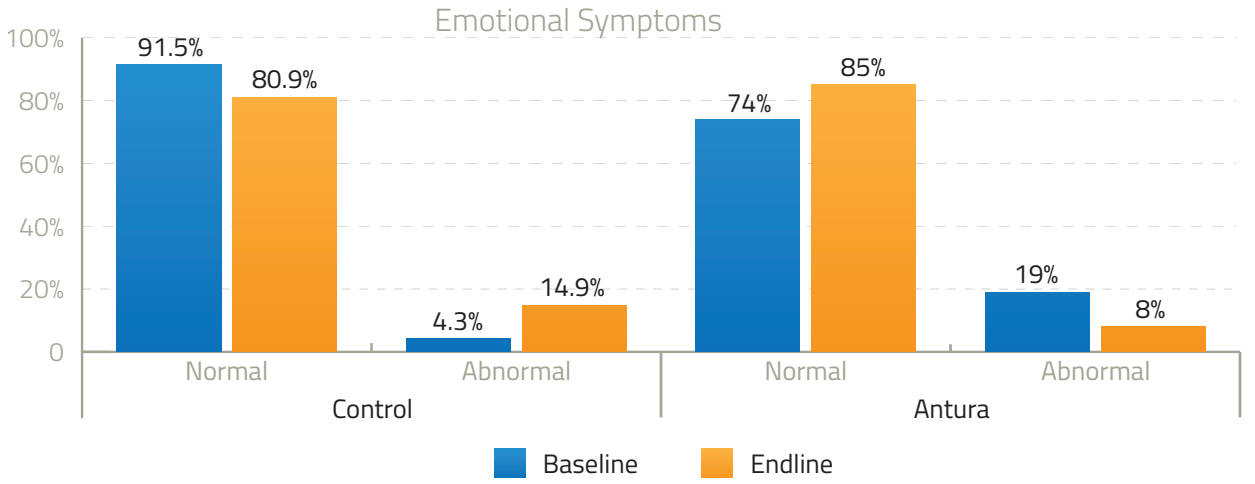
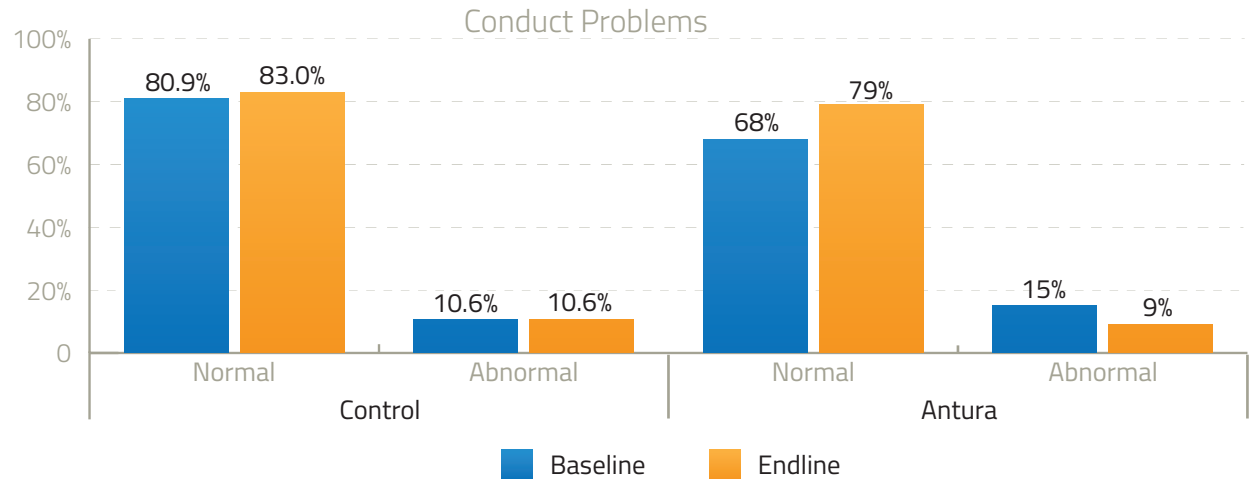


Figure 28: *Conduct Problems*



These positive trends are confirmed by focus group discussions with parents and children. These discussions found that children could correctly identify emotions, while parents indicated their children's behaviors had improved.

Among treatment children, normal scores within the peer relationship problems domain increased by 13% *(Figure 29),* and abnormal scores in this domain decreased by 6%. By contrast, normal scores for peer relationship problems among control children decreased by 10.6%, and abnormal scores for this domain increased by 4.3%. This was confirmed by play observations, in which 85% of children were engaged in play with their peers; parents and children also noted engagement with peers because of game play.

Figure 29: *Peer Relationship Problems*

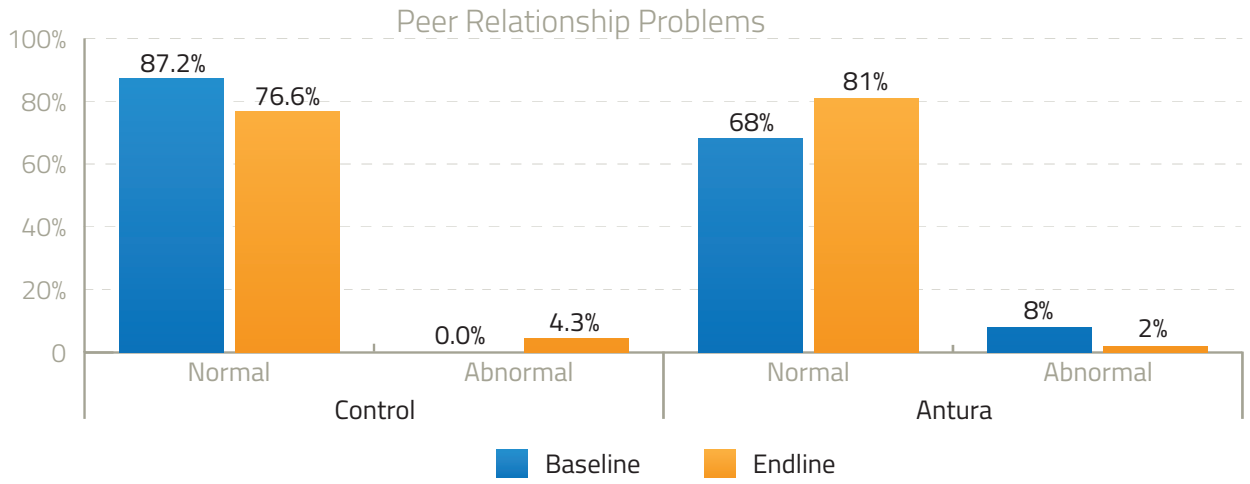


Figure 30: *Moving of home at Azraq Camp.*

On the other hand, scores in hyperactivity and pro-social domains among treatment children did not change during the testing period. Overall, this potentially signifies a positive impact on children's peer relations, conduct, and emotional symptoms, which are supported by qualitative findings where volunteers, teachers and parents reported increased peer interaction because of game play, improved behaviors during the testing period, and increased motivation and happiness associated with the game.

Control group pre- and post-scores do not reveal many improvements, but rather show steady or worsening scores within all domains except pro-social behaviors and an overall decrease within the total difficulties score of 76.6% to 70.2%. While Control children scores stayed steady or worsened and treatment children's scores increased, causation might be difficult to substantiate given the small sample size. Several other underlying factors may also have influenced these outcomes, including control group members' moves to new homes during the testing period *(Figure 30).* Nevertheless, the data trends among the treatment group are potentially indicative of overall wellbeing improvements

within three of the five domains assessed. These improvements are also supported by qualitative findings. However, the small sample size should be considered in qualifying the above.

6.4 Qualitative Findings

6.4.1 Children's engagement with Antura and the Letters

Focus group discussions revealed children reacted positively to and engaged with the game, with a total of 74 out of 86 children saying they enjoyed the game. Children added that the game "taught us the letters" and demonstrated that they were improving in terms of skills, learning, and advancing on the map. One child stated: "I like the game because it's fun and it teaches us and I move up through the stages." Children were excited to complete certain levels and enjoyed interacting with the game's main character Antura ("He made us laugh"). In some FGDs, all children in attendance reported they would like to continue playing the game. Children's enjoyment of the game was also seen in gaming observations, where the average level of overall engagement was ~88%.

In focus group discussions, most parents said their children enjoyed and were engaged in the game, with some (24 out of 80) parents saying the levels and prizes at the end of each level made the children proud of their progress. Some (28 out of 80) parents said that they felt their children were engaged with the game because they were entertained, while other (16 out of 80) parents felt their children were engaged with the game because they wanted to learn. Almost all parents (78 out of 80) reported that their children talk about the game daily, and many said their children regularly sing songs from the game. Some parents stated their children were more engaged in the game than their studies. In some instances, children requested tablets and smartphones for play, but the parents could not provide this. Other parents noted that because of digital games, their children no longer wanted to play outside.

Parents' sense that children would continue playing the game at home was confirmed in play-focused interviews, where almost all children (18 out of 20) stated they would continue playing at home, even though many (15 out of 22) have other mobile gaming options. Whereas more parents thought their children were engaged with the game for entertainment rather than to learn, the reasoning given by children themselves was evenly split between playing for fun and playing for learning purposes.

Bored	Engaged
27	50

Regarding long-term engagement, however, children's responses varied. Initially, children held back from disclosing their level of engagement with the game, but as the FGDs continued, children reported that their level of engagement had dropped as they continued to use the game over time. Upon further probing, about one-third of the participants reported they had experienced some level of boredom (as per the table) and asked for additional activities and mini-games within the game as well as entirely new mini-games. This was the case even though none of the children who played the game completed all the levels. The level of boredom experienced over time also occurred despite higher levels of flow on average (a 23% increase), higher game engagement (a 22% increase), and lower levels of frustration (a 7% decrease) observed during gaming observations in later weeks of the study. A high flow average indicates that the game offers children an engaging level of physical and mental challenge.

Rewards such as receiving bones or leveling up were found to be a motivating and encouraging factor in the children's game play. Children reported feeling happy when they would win or progress through different stages of the game. One child reported: "When the game tells me 'good job,' that makes me happy." Children readily recognized the advancement from stage to stage and noted which levels they had reached, indicating that leveling up contributed to their level of enjoyment and motivated them to continue play. Progress and regression were both clearly displayed to the children. Children also stated they either received stars for progress, the "letters would run away," or they wouldn't receive enough bones to progress to advanced stages. Children understood how the game gave them corrective feedback, with some children noting that if they got the answer wrong twice, they would ask for help from a peer. Sometimes this pertained to game mechanics as opposed to accuracy in answer selection.

These findings were confirmed by parents in focus groups who noted that their children felt a sense of achievement in leveling up. Some parents felt the levels allowed the children to be different and progress, and that children felt happy when they were encouraged to finish levels and move to new ones. Some parents noted that the game is a mix between learning and entertainment, which allowed the children to be engaged with the game. Some parents also said this learning method motivated the children to attend the center, so much that some children asked their parents to bring them an hour before class. Parents noted their children were also motivated using the tablet for learning, with some children reporting that "using a tablet is more fun" than traditional teaching methods.

When parents were asked if they would download the game on their mobile phones and let their children play and/or play with them, most parents (72 out of 80) were willing to let their children use their mobile phones to play the game. When asked how much time they would allow for game play, most parents (72 out of 80) said they were willing to give the child the phone for one hour per day.

6.4.2 Children's emotional and psychosocial outcomes with Antura and the Letters

In focus groups with children, most children reported positive outcomes, with 70 of 86 children stating they felt happy during play, while 12 said they did not enjoy playing the game. The children said they felt comfortable using the game. One child reported, "When I play the game, I feel like I am sitting at home, playing on the phone." This indicates that use of innovative and interactive methods of teaching have a direct impact on children's sense of ease. Children observed in gaming observations generally maintained a decent level of flow, with an average level of ~76% and a mode of 100%. A high level of flow indicates that children could interact easily with the game.

FGDs with parents and volunteers/teachers confirmed that children felt comfortable and happy playing the game, with one parent noting her son "always comes home happy and tells us what he learned." All (10/10) volunteers/teachers noted that children were happily engaged with the game, which motivated them to attend the summer camp. One volunteer/teacher noted the children are traumatized from war, but that entertainment and learning through fun are improving their emotional well-being.

At the same time, children also reported feeling "sad" or "angry" when they would make mistakes or lose. Although they reported these feelings did not make them stop playing, many reported frustrations regarding their inability to finish the game. Some noted that one of the least engaging features of the game was their need to restart letters when they don't get enough bones, saying, "It takes so long for us to win."

In FGDs, 9% of children reported getting frustrated when using the game, with one child stating, "It was hard, I didn't know how to use it at all." However, another child reported that "the more we used it the easier it became." Similarly, half of those interviewed three weeks after beginning play-focused interviews (eight out of 15) believed nothing was confusing about the game. Gaming observations saw low levels of frustration (7%), and the most commonly observed level of frustration was 0%. No frustration was observed over technical issues; children were quick to ask for adult assistance in resolving technical issues (seven observed instances), but as time progressed they learned how to restart the game on their own. Gaming observations revealed that technical issues were rare and consisted of buttons not working (five observed instances), the game freezing (one observed instance), and a glitch in the Paint/Color Tickle level (one observed instance). Errors were usually fixed by restarting the game.

While 79 of 80 parents indicated the game was easy for children to play, one parent said her daughter was frustrated with her lack of progression in the game. The cause for any amount of frustration with the game, although small overall, could be mechanical difficulties. In gaming observations, children were observed 39 times making repeat errors because of difficulty with game mechanics. There was no significant change in the frequency of observed difficulties with game mechanics at different points in the study.

One other positive outcome reported by children was their sense of ownership of the game. Children stated they had created profiles that represented themselves. They also reported enjoying the ability to customize Antura the dog, with one child noting, "I put a hat on mine, and I changed his colors." The ability to resolve technical issues on their own (as mentioned above) also contributed to a sense of ownership, self-confidence, and independence. One child in the focus group stated he was the first to play and navigate Antura and the Letters in his class, and became the "expert." As a result, he began to track everyone's progress in the center, noting which levels children in the center had reached.

Table 35: *Children voting on their favorite game feature*

Feature	# of Children's Votes
Stages/Leveling Up	52/86
Gaming Activities	41/86
Character/Antura	24/86

Figure 31: *The Sling/Catapult Mini-Game.*

Children also indicated a sense of ownership by identifying their favorite features of the game (*Table 35*). There was a large variation in what interviewees considered their favorite part of Antura, showing how the game's variety may enable a variety of users to enjoy the game. This is further evident in that three interviewees (out of 12) stated that expanded offerings (e.g. add more levels or different game modes) would improve the game. This is significant considering no children had finished the game at the time of interview.

Some findings specific to certain Antura mini-games were revealed during gaming observations and play-focused interviews (Table 36). These findings include that only a minority of children understood some of the mini-games and that children experienced confusion and difficulties with the mechanics of some mini-games more than others. Among the total occurrences of difficulties experienced with game mechanics (46 occurrences), 17 were repeated errors on the Sling/Catapult mini-game *(Figure 31)* and 13 were repeated errors on the Paint/Color Tickle mini-game. Still, a large majority of children enjoyed every game they were asked about. During gaming observations, the Sling/Catapult mini-game had over half (six out of 11) of observed occurrences of heightened engagement with game mechanics, when children would stare intensely at the screen or move the screen much closer to their faces.

Table 36: *Mini-games Findings and Observations*

Mini-Game	% of children liking mini-game	Findings of interest from CREATE interviews and gaming observations
Balloon	43.8%	Four out of 19 children stated in play-focused interviews that this mini-game was their least favorite part of the game. In addition, only 16.7% of children interviewed understood how the mini-game works.
Mouth	87.5%	Only 50% understood how to play. Adult observers said they were concerned that the picture on the screen consists of a head. Children were not asked specifically if they knew what was displayed in the picture on this mini-game.
Alphabet	94.4%	In some gaming observations, children were found singing the song together in class. However, in 12 observed instances, children demonstrated confusion with the mechanics of the mini-game and did not understand how to move the magnifying glass along with the song.
Egg	77.8%	Three out of 16 participants identified this mini-game as their least favorite part of the game.
Paint/Color Tickle	88.9%	This mini-game had the most incidents of engagement (12 occurrences), as well as the second-highest rate of observed mechanical difficulties (13 occurrences) and the highest rate of frustration incidents (five occurrences).
Sling/ Catapult	88.9%	31.3% of participants picked this mini-game as their favorite, but it had the highest level of observed mechanical difficulties (17 occurrences).

Figure 32: *Children interacting with the game and each other.*

Storyline: Many children in play-focused interviews could not remember what the story of the game was or even whether the game had a storyline. Over half of interviewees (12 out of 21) stated that Antura does not have a story. While the remaining children stated that the game had a storyline, some could not remember what it was, while others responded by simply describing the game.

Social Skills: Children reported that as they used the game, they would often play with their friends and enjoyed giving and receiving assistance to their classmates. This high level of social interaction was confirmed by parents in FGDs who noted that their children were competing on levels found the friendly competition positive. Most parents (43 out of 49) said their children became more competitive, adding that siblings would come home and discuss the game and progress.

Peer Interaction: Most (52 out of 80) parents in focus groups said their children became more patient, and 58 out of 80 said their children began helping others. All (10/10) volunteers/teachers noted increased competition among their children, and stated that this was a positive influence on the children through play, fun, and increased interaction. Volunteers/teachers reported that children compared performance, adding that game competition was a significant motivating factor. At the same time, some children reported in FGDs that they preferred to play alone to spend more individualized time on the tablet. For other children, playing alone meant they could concentrate and perform better.

Learning Behaviors: Some parents mentioned their children became more motivated to learn and do homework, and the clear majority (71 of 80) of parents said their children had a longer attention span and concentrated more in their studies. One parent said: "The game affected my son's abilities to study; he doesn't get bored while studying." Another parent added that her daughter, who was previously not interested in learning, "became more attentive to class and wanted to learn." However, other parents commented that their children became more distracted and did not want to do their homework. Volunteers/teachers noted that a blended learning approach (traditional plus technology-based) might be motivational. With respect to other attributes, many parents (21 of 28) said the children were more courageous, while other parents (29 of 38) said their children became more interested in learning games. More than half (46 of 80) said their children became more independent.

Age/Gender Differentials: Some findings differed for male and female children and children in different age groups. On average, male children observed in gaming observations maintained a level of flow 7% higher than that of female children. In play-focused interviews, only female children stated that they played Antura competitively (2 of 4 girls, compared to 0 of 5 boys). Younger children had more issues with the Paint/Color Tickle mini-game mechanics than older children (7 occurrences for young versus 3 occurrences for old). Older children were also much more likely to pick the Balloon mini-game as their least favorite level (3 of 7 old versus 0 of 4 young) and as their least favorite part of the entire game (4 of 10 old versus 0 of 8 young), though both age groups seem to equally dislike the Balloon mini-game (5 of 11 old and 2 of 5 young). Older children were also more likely to state wanting to continue playing the game at home for learning purposes rather than for fun (5 of 8 old versus 2 of 7 young). Younger children experienced 5% less overall frustration on average than older children.

6.5 Conclusions

Antura and the Letters engaged the clear majority (85%) of children regardless of literacy level, age or gender, as evidenced by the sharing of experiences with their parents. This was confirmed by parents who said their children felt a sense of achievement in leveling up. It was clear from children and parents that children were most motivated to complete levels. Antura and the Letters was well-designed in its variety of game play and leveling of skill sets, which enabled most children of all ages and gender to play the first two levels. Initial interface with the game was challenging for some children, but with

time children could play with minimal or no adult supervision. By midline, most children could play independently and progress, reflecting a relative ease of use.

Children reacted most positively to gaming activities and character. Antura as a character was positively received, as was the off-task play space, which motivated children to continue playing and sustained interest by unlocking accessories for Antura. The mini-games were many and varied and repeated throughout the levels of play; they also were randomized such that child trajectories were not the same. This likely increased the level of sustained engagement. Some mini-games were preferred over others, but it is important to note that most children did not complete half the game and therefore did not play all the mini-games. As such, their indicated preferences and feedback pertain primarily to the first two levels.

At a psychosocial level, Antura and the Letters appears to have supported the development of positive social outcomes, as demonstrated by both SDQ improvements in peer relationships, conduct, and emotional problems. This was also confirmed by parental feedback noting improvements in patience, independence, and motivation.

Game play was varied, and game mechanics for most mini-games were understood, with some exceptions. The objectives or pathways of the game were not immediately clear to children, which generated some frustration. Some mini-games such as the Mouth mini-game and Conveyor Belt mini-game need revision to be better understood and improved. Among recommended revisions is an easier and clearer start to the game as well as shorter milestones of engagement. The game could also benefit from clearer introduction of objectives, scoring, and travel. While the game clearly communicated progress through rewards and travel on the map, children often felt the game was too long.

Children were effectively rewarded by the game through verbal praise, rewards, and feedback from the game. However, some children were frustrated when they would not earn any bones. Some mini-games are challenging and do not allocate sufficient time; thus, an easier form of game play is needed for younger children who are unable to perform the tasks requested as fast or with as much precision as older children. Younger children tended to be more frustrated with mini-games requesting more advanced fine motor skills, precision, and speed. The game could benefit from a revision of gaming mechanics by requiring less precision or allocating more time to better accommodate younger children. Moreover, some randomized mini-games provided challenges too advanced for early learners and may have disrupted game flow.

At a psychosocial level, Antura and the Letters appears to have supported the development of positive social outcomes, as demonstrated by both SDQ improvements in peer relationships, conduct, and emotional problems. This was also confirmed by parental feedback noting improvements in patience, independence, and motivation. The sense of accomplishment and learning seems to have contributed to children feeling happy. The game enabled them to also feel a sense of ownership and attachment either through avatars or through the customization of Antura. Children wanted to share the game and their learning with siblings and peers, demonstrating positive impact on emotional states and social behaviors. Peer interaction increased during game play, which had positive implications for so-

cial and communication skills. Nevertheless, while a minority, some parental feedback indicated the game was a distraction and reduced attention spans, while SDQ results showed no improvements in hyperactivity/inattention.

While parental engagement with the game was limited due to lack of parental awareness or lack of device/electricity, most parents expressed a willingness to download the game, with one parent noting she used it to increase her own literacy.

6.6 Recommendations

Antura and the Letters could benefit from improved initial interface with the game to ensure pathways are clear, user interface is simple, and game play is clearly understood, to enable play with minimal or no adult supervision. Since Antura and the Letters' objectives or pathways were not immediately clear to children, the game could benefit from clearer introduction of objectives, scoring, and travel. The game could be better streamlined to enable more frequent achievements of milestones and completion of levels through game simplification and improved pacing.

As the Antura off-task play space effectively engaged all children, the game could benefit from increased visibility of the play space. As the game enabled children to also feel a sense of ownership and attachment either through avatars or the customization of Antura, the game could benefit from increased personalization of profiles and off-task spaces.

For younger children, the game could benefit from differentiated pacing and a revision of gaming mechanics to better accommodate younger children by requiring less precision or allocating more time.

While not a requirement of the EduApp4Syria competition, there is room for capitalizing on the positive impact of social interaction in game play to increase pro-social behaviors and psychosocial outcomes. This could be accomplished through the development of sharing mechanisms to support peer interaction or friendly competition.

Specific Psychosocial and Engagement Recommendations

- Improve the storyline and link level and progress mapping to it so children can clearly follow the narrative. Consider showing a destination so children can anticipate the end.
- Showcase the option to dress and accessorize Antura, as this component was a favorite of the children. Consider introducing this option as part of the storyline. Within this space, consider enabling children to personalize accessories for increased attachment, such as creating a dog-collar pendant with the child's name or picture.
- Augment the uses of the children's avatars to include more psychosocial components. These can include aspects related to children's personal identities, likes/dislikes, emotional expression, physical attributes, and abilities. Including this feature promotes self-identity development.
- Enhance the positive reinforcement used throughout the stages to promote children's self-esteem. This could be through varying the terms/visuals used, adding different gaming features, unlocking add-ons children can use to design their avatars, etc.

- Customize mini-game rewards to be activity or skill specific. This would highlight children's strengths, which in turn has a positive impact on psychosocial health.
- Consider reducing the level of performance required to receive three bones to 80–90% accuracy, with a bonus for 100% accuracy to reduce frustration, especially at the early stages.
- Include different emotional expressions for the character's reactions—this would expose children to new emotional vocabulary.

Recommendations for Future Iterations (Currently Outside the EduApp4Syria Specifications)

- Consider better integrating the collection of bones into game play, beyond feeding Antura. What happens if a child does not feed Antura? Feeds him too many bones? What else can bones do?
- Consider introducing mini-games where the child is free to experiment with building their own words (even if words are incorrect or invented).
- Develop a track feature of the game that would remind children to take on activities or stages that focus on their current challenges/stages. These reminders should include a motivational aspect to encourage, or help deal with frustration.
- Add a "sharing" platform that could promote peer social interaction through friendship-building, communication, and social skill development—critical components for the psychosocial development of children.
- Enhance "right brain" features of the game that target children's emotions and spatial skills through imagery, music, and stories. These stories could also be developed based on the variety of different words presented in the game, but could also target a deeper level of psychosocial well-being by including stories that depict feelings as well as problem solving.
- Increase the connection/link between teacher and student if using the game in both formal and informal education settings. For example, this might include a feature in which teachers could allow children to prepare for lessons in advance. This would enable children to feel more confident about participation in class. Additionally, promoting interactions between the student and teacher lessens student pressures to perform and promotes positive social interactions between children and adults.
- Include a creative space within the game for children that could be utilized for writing stories, drawing pictures, creating short clips, etc., based on the words they have recently learned. This would enable use of the right brain hemisphere as well as promote children's imagination and self-expression—important pillars of psychosocial health.

Photo Credit: Norad/Marit Hverven

7. Areas for Further Research

This impact evaluation explores the potential for Digital Game-based Learning (DGBL) on improving literacy and psychosocial outcomes—an area of research that is relatively nascent. The novelty of the EduApp4Syria competition, as well as the ensuing products and its potential to address literacy and psychosocial through DGBL, offers opportunities to further explore methods of education delivery, particularly for out of school children. In a region[52] where armed conflicts over the last 15 years have disrupted education in entire countries—resulting in 13 million Syrian, Iraqi, Yemeni, Palestinian, Libyan, and Sudanese children not attending school[53]—there is a clear need to find innovative methods of education delivery accessible to out of school children.

The high rate of mobile phone subscription and smartphone penetration offers opportunities for remarkable reach of DGBL into the MENA population. According to a study by mobile operators, there were 339 million mobile phone subscribers in the MENA region in 2016, accounting for 60% of the population.[54] Moreover, smartphone penetration among the MENA population has more than doubled over the last three years, currently accounting for 42% of mobile subscribers' phones. This population has one of the fastest growth rates in the world and is expected to reach 65% by 2020, on par with global averages.

The EduApp4Syria competition was an innovative approach taking advantage of the digital revolution while addressing the challenges of education delivery. DGBL's disruptive approach to learning could address some of the challenges faced by educators in the MENA region. DGBL offers the potential to reduce barriers to accessing learning materials, alleviates the need for physical learning environments, and offers the opportunity for iterative and differentiated learning. This provides an opportunity to address quality of education delivery. Serious games or simulations with clear "learning outcomes can create interactive experiences that actively engage the players in the learning process, [whereby] experimentation, graceful failure, and identification of lessons learned can result from game-based learning, where decisions and actions are chosen, consequences experienced, goals are achieved, and feedback furnished."[55]

Although game developers in the region are producing learning games, there are very few such learning games in the region that have undergone rigorous testing. While this report makes one contribution to learning how DGBL can impact literacy and psychosocial outcomes, further research is needed to understand the potential of DGBL for education, psychosocial well-being, and populations in conflict. Some salient issues for further research are outlined below:

Dosage: While education practitioners tend to agree on 60 hours as the minimum dosage to achieve meaningful literacy outcomes, the testing dosage of 26 hours for Antura and the Letters generated some positive results and improvements across all age groups and genders. This leaves much room for discussion on possibilities of greater impact with increased dosage, or alternatively greater impact with gaming improvements. Moreover, games are intended to engage children for prolonged periods and culminate with a sense of accomplishment after completing or winning the game. As such, it may prove challenging to engage in game play for the recommended 60 hours, as Antura is intended to be completed in 15 hours. Moreover, a critical component of gaming is the sense of accomplishment, which may prove challenging to design and realize within a 60-hour time frame. Since DGBL does not

correspond to traditional literacy instructional intervention, perhaps there is room for further research on what type of DGBL dosage would be required to generate impact. Moreover, the evaluation reveals that a standardized pace may not best meet the needs of early or more advanced learners, indicating a need for self-paced game play that may not conform to game play times or dosage requirements. As a result, further research should examine a game-defined approach to dosage, to which this report lends some insight.

Self-Pacing: As noted in the evaluation report, children with different abilities, ages, and exposure to literacy advanced through the game at different rhythms and varied in their perceived challenges. As such, a key recommendation is to modify the pacing of the game to each type of learner. Enabling DGBL to self-adjust according to child performance may lead to improved and differentiated progress through the game. Further research is also needed to examine variation in dosage by age or ability.

Testing Context: While the intended use of the game is for home-based use by out of school children, testing of the game was conducted in a structured manner at remedial education centers to ensure adherence to testing protocols and dosage. As a result, the game was not tested in its intended setting—the home. Since the evaluation focused solely on the testing within the controlled remedial education setting, there is room for further investigation on game use in the home. Further research might also examine appropriate dosage in a home-based, out-of-school setting.

Illiterate Parents: In some instances, parents of children testing Antura and the Letters indicated they had engaged in game play with their children. As some of those parents were illiterate, it may be worthwhile to explore the use of games to improve literacy among an illiterate adult population. While out of the scope of this study, it may be an area for further research once the games are modified.

Psychosocial Outcomes: The testing context and proportion of respondents did not allow for a robust analysis of psychosocial outcomes, as a relatively small proportion of parents (33%) responded and the control group experienced circumstances (moving of home) during the testing period. Such challenges did not allow for equivalence. Nevertheless, the SDQ results show promise for further investigation of psychosocial outcomes within test settings that allow for increased access to parents and equivalence of testing groups.

Pro-Social Behaviors: A key finding in game testing was the level of peer engagement and social interaction generated by game play. The evaluation was designed to track individual children over time, but there may be areas for further research to explore group dynamics in game play over time. Further research could explore ways in which social aspects of gaming—such as peer or class competition or peer sharing/teaching—can impact psychosocial outcomes and learning.

There is a need for better understanding of the impact DGBL solutions can offer, particularly as DGBL content developed in Arabic holds a potential market of more than 350 million native speakers. As such, innovators in the MENA region and beyond are responding to a large and growing market for DGBL in Arabic. In the MENA region, DGBL demand growth rates are projected at 28% in the 2017–2020 period—the fourth highest region out of seven global regions assessed with the 2017–2022 Global Game-based Learning Market.[56] These areas for further research can contribute greatly to DGBL design and development in the MENA region. Research also may advance innovations in education delivery that have the potential to generate combined learning and psychosocial gains among early-grade readers worldwide.

Annex 1: Literature Review

Relevance of DGBL to Syrian Refugee Educational Context

In the Levant region, perhaps the most pressing education challenge is how to address the educational challenges associated with the wave of Syrian school-age refugees, predominantly residing in Jordan and Lebanon. An estimated 2.3 million Syrian children are out of school because of violent conflict in their country.[57] Many of these children must cope with memories of multiple traumas and high levels of stress. The conflict has disrupted their education and the trauma and stress levels often affect their ability to learn. According to the United Nations, it will take about eight to 10 years for displaced refugees to return to Syria,[58] perhaps best encompassed or expressed within the No Lost Generation initiative of the humanitarian response community.

Digital and Game-based Learning (DGBL) is often considered a disruptive approach to learning that holds the possibility to address some challenges faced by educators in the Middle East and North Africa (MENA) region. Specifically, game-based learning offers the potential to reduce barriers to access to learning materials, alleviates the need for physical learning environments, and provides the opportunity for iterative and differentiated learning. Game-based learning also offers the opportunity to address quality of education delivery.

DGBL is gaining momentum in the Middle East. Although research based on these development activities is underway, few studies have been conducted to date on educational games, the markets for DGBL, and the impact of DGBL in the MENA region. Nevertheless, serious games or simulations with clear "learning outcomes can create interactive experiences that actively engage the players in the learning process, [whereby] experimentation, graceful failure, and identification of lessons learned can result from game-based learning, where decisions and actions are chosen, consequences experienced, goals are achieved, and feedback furnished."[59]

A landscape review of 44 projects and studies undertaken of Information and Communications Technology (ICT) for education and specifically for literacy and reading conducted in 2014 by the United States Agency for International Development (USAID), revealed that "the use of mobile ICTs designed to help children learn to read, practice reading (reading to learn), and acquire a broader range of learning skills that support a literate society…provides a new opportunity to re-imagine traditional forms of educational design and delivery."[60] Further, a recent global meta-analysis of 77 randomized experiments evaluated the effects of school-based interventions on learning in developing country primary schools found that the largest mean effect sizes (among school-based interventions) included treatments with computers or instructional technology (0.15)."[61]

Mobile Phone and Internet Access

According to the Groupe Spéciale Mobile Association (GSMA) Mobile Economy report for 2016, there were 339 million mobile phone subscribers in the MENA region[62] in 2016, accounting for 60% of the population.[63] Despite the high rate of subscriber penetration, there is significant variation between the various regions. Nevertheless, smartphone penetration has more than doubled over the last three years—currently accounting for 42% of mobile subscribers' phones—and is expected to reach 65% by

2020, on par with global averages.[64] In some markets, as many at 75–85% of people have smartphones, and mobile penetration in many markets exceeds 100%, though these levels can vary substantially for refugee communities.[65] In Jordan and Lebanon, where most of the Syrian refugees in the MENA region reside, at least 98.5% of refugees live in areas with at least 3G coverage.[66] According to a Regional East African Community Health (REACH) report conducted in Jordan, mobile phone ownership among encamped refugee households was widespread, and smartphone penetration reached 58%.[67] Among those without a smartphone, 47% had access to one through a household member. The study also noted that Syrians use multiple subscriber identification module (SIM) cards due to unreliability of networks in areas where they reside, ease of purchase/pricing of SIM cards, and the spread of highly capable and affordable mobile handsets. Some refugees living near borders use SIM cards from other countries.[68] As a result, contact information/phone line ownership is inconsistent and often out of date.

Examples of DGBL, Similar Solutions, and Their Design

We have found no examples of rigorously tested DGBL solutions for Arabic literacy with applicability to the Syrian refugee context. An example with many relevant features and high complementarity to the EduApp4Syria approach is Qysas (although not a game, it is a rigorously tested Arabic literacy app developed under All Children Reading: A Grand Challenge for Development [ACR GCD]). Another relevant example is Can't Wait to Learn, a game-based Arabic numeracy app developed under E-Learning Sudan ELS). These two examples were chosen as two rigorously-tested digital learning tools with relevance to Arabic language instruction, psychosocial well-being, and education in crisis contexts, and they illustrate some of the challenges and possible solutions to widespread DGBL adoption. The literature indicates there are other DGBL examples in use and/or being developed. Those are touched upon but are not reviewed in depth due to lack of material.

The Qysas early grade digital-leveled library is one example of an entertaining digital platform used for learning purposes. It is designed to help early grade students develop and strengthen foundational Arabic reading skills and reading comprehension skills through access to a library of 125 interactive books leveled from Grades 1–3. Under the ACR GCD grant, Little Thinking Minds (LTM) developed Qysas (stories, in Arabic) content and adapted it for mobile devices, enabling tablet-based or mobile phone-based access by end users to 125 regionally-sourced books through an iterative and practice-based program of gradual literacy strengthening.

Early Grade Reading Assessments (EGRA) conducted in Jordan (2012)[69] and Egypt[70] have shown relatively weak performance in early grade literacy measures. Some areas of the Arab region display high repetition and drop-out rates, especially in poor rural and urban communities.[71] A recent United Nations Children's Fund (UNICEF) International Bureau of Education analysis of the grade repetition phenomenon in the MENA region makes the following three points: 1) there is a significant link between repetition in the first grades of primary education and the learning of reading and writing; 2) there is a need for significant changes in the teaching of reading and writing and for a thorough overhaul of the parameters and traditional practices usually applied to first literacy in formal and non-formal situations; and 3) there is a need for greater awareness of the impact of linguistic factors on school performance in general and literacy in particular.[72] Responding to this gap, Qysas was developed with the aim of increasing access to books through a differentiated and iterative approach to reading, where students in need of supplemental assistance are able to access it wherever they live. The result is a series of interactive and animated books on iOS and Android application that can be downloaded onto

smartphones, tablets, or from pre-loaded tablets. For those students able to access Qysas, the program enables students to access their appropriate reading level, and work up the levels as they pass built-in assessments.

The Qysas app is a leveled digital platform meant to entertain children as they interact with the app's interface. It contains five levels, each made up of three main components: 1) audio reading—a new book is presented to the child through animated voice-over and interactive page turning; 2) practice reading—the child reads the book and turns the pages; if the child cannot read the word, the child can touch the word and the audio function is activated; 3) reading comprehension—the child is asked a series of questions about the story and must answer a minimum number of questions correctly to mark the book as completed. Qysas groups learning into five levels that correspond to the first three grades of elementary school, splitting learning into smaller learning modules, animating texts for extensive reading practice, using the vocabulary drawn from the 40% overlap between fus-ha and spoken Arabic, and using a uniform letter size/font/spacing to advance literacy automaticity. As those stages are completed at the child's own pace, he or she is rewarded and the next level unlocks. Individual student performance is tracked, offering the teacher a student-specific or class-wide dashboard of reading performance.

Although Qysas was tested within literacy clubs in public schools in Jordan, the Qysas learning platform can be used in a variety of learning environments: individually at school, in a group setting or club, and individually or with a peer/parent reader at home. Under the testing environment, school teachers and librarians were trained on how to apply Qysas as a tool for supplementary learning and literacy strengthening. The literacy club approach entailed a one-on-one individualized student interface with Qysas twice a week for the academic year in a classroom setting administered by a teacher for up to 45 minutes each time, for a total of 48 hours. In that time, most children in the testing schools finished all five levels, or 125 books. Where the average for the Arab World is six minutes spent reading annually,[73] this program offered a simple solution to get children reading. An impact evaluation of the program showed that Qysas advances literacy outcomes with statistical significance in syllable fluency, oral reading fluency, and reading comprehension, and significantly reduced zero scores in those three domains, as compared to control group performance (ACR 2017 report forthcoming). Children were motivated to come to school on tablet days as noted by their parents and high attendance rates, and children were also motivated to "level up," comparing reading levels among each other. The practice offered through the Qysas testing environment also resulted in increased reading confidence among students, as noted by teachers and parents. Teachers noted this confidence often transferred into other areas. Due to its universality for Arab children in Grades 1–3, Qysas aims to apply its programming to refugee settings as well.

Can't Wait to Learn-ELS is a game to teach students through play that meets the definition of DGBL more precisely than Qysas. To serve the educational needs of more than 3 million Sudanese children ages 5–13 who are still out of school,[74] a consortium of stakeholders including the Sudanese Ministry of Education, Ahfad University for Women, War Child Holland, and the Netherlands Organization for Applied Scientific Research (TNO), developed a serious game for mathematics based on the Sudanese national out of school math curriculum targeting out of school children. Although Sudan has relatively high primary school enrollment rates, gross enrollment rates range from 85% in Al Gezira to 37% in East Darfur (Ministry of Education Sudan, 2014).

Sudan still faces significant infrastructure challenges around school resourcing, poor learning out-

comes, and low federal spending. The accessibility and quality of education for children in Sudanese states still affected by conflict remains a challenge. The ELS game aims to make "basic [math] education available where formally trained teachers or schools are not present. The initial concept intended to make basic education available without teachers or formal institutions, through the delivery of a serious game on hardware that is placed in a community and supported by local facilitators."[75] The ELS game is also designed to address psychosocial well-being by teaching life skills, consequential decision-making, and solution thinking. It is delivered as "part of an e-learning package that includes access to solar power and community facilitators trained in child-friendly approaches and on how to use the game and tablets."[76]

ELS consists of two game worlds and various mini-games to practice each mathematics concept through autonomous learning.[77] It incorporates two distinct levels: the first level is that of game worlds, which provide the connecting narratives for the second level. The second level is made up of separate mini-games (44 different mini-games, 160 variations of mini-games) designed to address a specific mathematical concept. Game World I asks the player/student to help other children achieve goals in their lives, like becoming a goat herder or doctor, using mainly familiar jobs/roles within the community or jobs less familiar like teacher or engineer. Game World II is a shop where children can buy and sell products or increase the number of products they can sell to enhance their shop. It utilizes an experiential learning approach to engage children, while a management system tracks progress. Each mini-game is designed to addresses a specific mathematical concept. Children work at their own pace, and progress is based on their ability to successfully complete games and tasks.[78]

The children played the mathematics game for a maximum of 45 minutes per day, five times per week in community facilitator-administered learning sessions, for four to six months.[79] Testing among 591 participating children revealed that children playing the game improved their knowledge of mathematics, tested within two different math tests: scores of tested children increased on Test A from a mean of 20 at baseline to a mean of 41 at endline; and increased on Test B from a mean of 32 at baseline to a mean of 41 at endline out of a total score of 60 for Tests A and B. No significant differences for gender were found.[80] Self-esteem, recorded on a four-point Likert scale, was reported to have increased from 1.9 to 2.5 from baseline to endline.[81]

As the humanitarian response community unfolds its piloting of various initiatives in the DGBL space, other examples will continue to emerge. For example, Sahabati ("My Cloud" in Arabic), is a virtual school for children in crisis, primarily those affected by the Syrian crisis. Now in development, Sahabati is designed to provide children and adolescents affected by conflict in the region with the opportunity to continue their education and receive certification for their learning, irrespective of their location and schooling time they have lost. Sahabati offers four core subjects: Arabic, English, Math, and Science, and will be coupled with a system of online assessments and certification.[82] Meanwhile in Lebanon, UNICEF is working with the International Education Association to pilot the Raspberry Pi computer, a hand-held device on which children learn numeracy skills and basic programming.[83]

Contextualization of DGBL and Similar Solution

Within conflict or refugee settings, technological infrastructure presents an even greater challenge, as basic or consistent availability of electricity is often unpredictable, making the simple act of charging a mobile device or tablet problematic. A review of recent projects serving refugees reveals that in Jordan and Sudan, initiatives such as solar-power chargers for refugees, or solar power chargers in community centers in Sudan,[84] have enabled charging of mobile devices in the absence of consistent

electricity. Where DGBL and similar solutions have been applied in these settings, adapting to the context and its limitations have been key, as noted by ELS Sudan: "We believe we have been able to have successful learning outcomes because we have been cognizant of the context and we have adapted to the context."[85] Some adaptations to contexts within the examples of ELS and the Qysas platform cited above were made to enable operation of the app offline.

With Qysas, back-end data collected included the level each student achieved, the score on reading comprehension, the number of times a book was read, and the time spent through individualized tracking with usernames. This data can be linked to a teacher dashboard that aggregates student performance, so the teacher is able to monitor student progress. On the back-end, individual student performance is tracked, offering the teacher a student-specific or class-wide dashboard of reading performance. Back-end data collection serves the learning agenda.

For those DGBL and other digital platforms aiming to be integrated or endorsed within formal education, another potential hurdle is integration of DGBL and similar solutions within curricula, at two levels: integration within national curricula and integration within the classroom. Ministry of Education buy-in for DGBL and/or curriculum reform can become a political challenge of negotiating and approving both content and structure of DGBL delivery within national curricula. As noted by the landscape review of Education under Fire, understanding these political requirements "to help ensure content is appropriately and responsibly delivered, is important."[86] The study notes that content created in the West, not locally, may be laden with "values, beliefs, or incomplete historical representations of both local and global issues."[87] These issues become more salient in areas impacted by conflict. The literature recommends to source content locally where possible. Finally, "working with governments and formal curriculum where possible can support long-term education initiatives."[88] For example, the Ministry of Education was involved in curriculum development of the ELS game in Sudan, while the Qysas platform in Jordan was tested in public schools in Amman with Ministry of Education approval and participation.

Moreover, for integration to occur in the classroom, teachers must be trained and/or experienced in DGBL usage and instruction. They must also be motivated to use app-based resources in the classroom. In the testing of the Qysas platform, teachers/volunteers implementing the testing of the app noted that use of app-based learning platforms in the classroom entails more effort than traditional classroom instruction, and that extra effort requires some recognition to motivate teachers to apply it. This is confirmed by other worldwide studies that document the importance of effective implementation by teachers in realizing the full potential of DGBL.[89]

Finally, as more DGBL programs come online, the literature reveals that a current obstacle to adoption of DGBL by educators is how to differentiate between DGBL applications, and the lack of a rigorous evidence base on the impact of DGBL on learning. In the 2014 landscape review of 44 mobiles for reading projects, "only one Mobiles for Reading (M4R) project contained an adequate randomized control trial (RCT) impact evaluation design."[90] In the MENA region, very few DGBL applications have undergone rigorous testing providing such an evidence base.

DGBL and Learning and Engagement

The literature and empirical evidence reveal that game-based learning environments offer potential for increasing student engagement and motivation, which have natural ties to learning.[91] The literature reveals that the relationship between emotional and cognitive activities is strong: "that positive

affect such as engaged concentration, joy, and excitement can lead to increased learning through better strategy selection, increased persistence, and improved use of mental resources."[92] DGBL also has the potential to mimic the benefit of one-to-one tutoring through interactivity, individualized attention, and feedback. However, learning gains are only possible if the student or player is engaged and motivated by this interaction and use. Studies in computer-based learning environments show how students' emotional states can impact how a student learns—where boredom, confusion, frustration, and anger may lead to decreased motivation and disengagement from the task. To maintain engagement, digital games often offer mechanisms to keep the player engaged such as rewards and adaptive difficulty levels. Disengagement can manifest itself as off-task behavior, although the literature reveals that further investigation is needed to assess whether off-task behavior negatively affects learning or maybe a coping mechanism for negative learning emotions. The literature also notes that certain off-task but in-game behavior may allow the player to remain engaged in the gaming environment, but not with the learning content, which may engender more positive feelings towards the gaming environment. Removal of these off-task, in-game features may decrease positive outcomes such as engagement.[93]

Assessing DGBL Games

For games to be effective learning tools, specific conditions must be met.[94] These conditions include the following, which formed the basis for the assessment of the EduApp4Syria games.

Rationale for the Use of Games: The design and development of games is expensive, and games can result in the need for processing of extraneous information that can inhibit learning.[95] It therefore needs to be carefully considered when the use of games is appropriate. Reasons that have been explored in this context are learners' lack of motivation to study a particular subject, the ability of games to contextualize learning, and the many ways games can engage learners.[96]

Appropriateness of the Learning Mechanic: One of the most important design challenges for games for learning is the learning mechanic, i.e., the essential game play that is repeated throughout the game and is supposed to lead to the desired learning outcomes.[97] When the design of these learning mechanics is not aligned with the learning objectives of the game, then the learning outcomes of the game are in jeopardy.

Appropriate Design for Users and Context: As in any other learning environment, games need to be designed with special consideration for the learners and their unique characteristics, as well as for the context in which they will be used. For example, games designed for use in formal school settings often have different requirements than games designed for use at the learner's home. This has implications for the platform on which games are developed (e.g., mobile, PC, console), the length of play, and many other design parameters.

Balance of Fun and Learning: One of the major arguments to use games for learning is their power to motivate and engage learners. However, designers of games need to balance the desire to make games fun and engaging with the goal to reach the intended learning outcomes. When too many game features are added, the focus on learning content may be lost. For that reason, we have recently begun to focus on activities we call playful learning rather than games.[98]

Annex 2: Methodological Approach

Impact Evaluation

The impact evaluation aimed to establish treatment and control group equivalency to the extent possible in the refugee camp context in Jordan. To address concerns that children within the three assigned groups (two treatment groups and one control group) have underlying differences, children were matched by a range of characteristics. Characteristics of critical importance for establishment of equivalence among groups were:

- Syrian nationality
- Little or no schooling
- Ages 5–10
- Gender balance
- Living in camp setting in Jordan
- A beneficiary of a humanitarian actor
- Access to a mobile phone (if possible)

The study first began by identifying possible test subjects meeting the above criteria. Relief International (RI), a humanitarian relief agency providing remedial education services to Syrian children, offered its education centers as testing centers for the app. Numbers and locations of children served by RI meeting the above criteria in Azraq camp were identified. Azraq camp was chosen as a test site, since Azraq camp hosts refugees who are newer arrivals to Jordan, and thus have lower levels of schooling. To calculate an adequate sample size for the evaluation study, minimum detectable impacts (MDIs) were determined for one key intervention outcome: oral reading fluency. MDIs are the smallest intervention impacts that the evaluation can detect. Based on our assumptions, a planned sample of 225 children per assessment group (225 for Antura and 225 for Control), was determined to be sufficient to lend power to the study, for a total sample of 550 children. To account for an anticipated 25% attrition, a total of 600 students were needed as an initial sample size. A randomized sample of treatment children was selected from Azraq camp where RI operates its remedial education centers, as follows:

Table 37: *Sample group in Treatment and Control groups disaggregated by Age*

Baseline	5 years	6 years	7 years	8 years	9 years	10 years	TOTAL
Control	1.4%	19.8%	18.7%	21.2%	19.8%	16.5%	283
Antura Treatment	1.4%	20.5%	16.7%	20.2%	19.9%	21.3%	366
Endline	**5 years**	**6 years**	**7 years**	**8 years**	**9 years**	**10 years**	**TOTAL**
Control	1%	21%	19%	22%	19%	19%	200
Antura Treatment	2%	18%	15%	19%	24%	22%	202

Attrition

Impact was assessed within the Village 5 and Village 6 cohorts primarily through a measurement of literacy levels using EGRA testing at baseline and endline, and comparing gains through the Difference in Difference approach (DiD). Simultaneously, psychosocial well-being at baseline and endline was measured using the Strengths and Difficulties Questionnaire (SDQ) administered with parents of children in the treatment and control groups.

Participation among parents at baseline was less than targeted at baseline (169 parents per group based on 95% confidence level and 5% confidence interval of the sampled child population), due to a variety of factors, including lack of adequate outreach to parents, ongoing humanitarian distributions at baseline, and lack of desire to be surveyed.

Table 38: *Attrition in the treatment and control groups for SDQ survey and EGRA Assessment*

	Baseline	Endline	Attrition
SDQ (parents)			
Control	123	47	76
Antura Treatment	127	53	74
EGRA (children)			
Control	383	200	183
Antura Treatment	366	202	164

Contamination

Contamination	
Antura Treatment	5 children
Control	12 children

The Azraq camp offers conditions most suitable for controlled conditions, since the Azraq camp has severely restricted Internet access and connectivity, thereby limiting the ability of control subjects to download the apps. In Azraq, control and treatment groups were also separated by administrative villages, with Village 5 remaining physically separated from the other villages for security purposes. To limit cross-contamination between treatment groups, INTEGRATED worked closely with RI to ensure a strong understanding of the importance of non-contamination for the testing period. An exit survey was developed at endline to establish the levels of contamination. Contaminated children EGRAs and their parents SDQs were removed from the study.

Technical Evaluation

Technical evaluation data was collected from key stakeholders as follows:

Focus Group Discussions (FGDs): FGDs were conducted to explore in-depth qualitative issues with parents and app users (children), intended to capture input of beneficiaries (children using the apps) and parents. Where children were involved in discussions, all appropriate protocols and consent measures were taken to create a suitable environment for these groups in addition to consideration of gender differentials, in line with the Inter-Agency Guide to the Evaluation of Psychosocial Programming in Humanitarian Crises and Do No Harm principles. Focus groups with children were led by a child development specialist.

In-depth Interviews: In-depth interviews were conducted with key informants and representatives from partners in implementation, instructors enabling use of the app, app design companies, and the humanitarian actors supporting the testing of the app. The in-depth interviews were guided by semi-structured interview guides. Each interview guide was developed to capture insight into the relevance and effectiveness of the app to provide developers with feedback for improvements. The guides were designed to preserve the potential for a relatively free-flowing conversation, while creating a standardized format to facilitate a reliable, comparative analysis of data for triangulation of information from multiple stakeholder perspectives.

Observation Checklist: An observation checklist was developed to assess various aspects of the apps including ease of use and levels of interest in the app among users (children). One hundred observations were conducted throughout the testing period to assess ease of use and engagement with the game.

In-Depth Gaming Observation Protocol: Observations of game play were conducted for four different personas of app users (see below). An observation checklist was created and conducted by Consortium for Research and Evaluation of Advanced Technologies in Education (CREATE), to collect qualitative insights on research questions related to specific game features and related experiences. We used a purposive sampling based on the personas we defined, which assured that users with specific characteristics were included in the observed sample. Thirty-nine children were observed at least twice for approximately 15 minutes of game play. The first observation took place when children were first exposed to the game. The second observation took place after they had used the game for at least two weeks.

Different Personas

- **Younger Male Player.** This is a 5-7-year-old male user.
- **Younger Female Player.** This is a 5-7-year-old female user.
- **Older Male Player.** This is an 8-10-year-old male user.
- **Older Female Player.** This is an 8-10-year-old female user.

Play-Focused Interviews: Semi-structured interviews were conducted with the players whose game play was observed at least once. The interviews were guided by interview guides, which provided the facilitator with core questions to be asked of each participant, and a question bank from which the facilitator drew based on the play observations. For each question, additional probes were provided in case the question did not elicit sufficient responses. This allowed for the collection of shared data for all interviews, and specific data based on observations of interest for which more detail is required.

Fieldwork Preparation and Data Collection

Assessor Training and Quality Assurance

- *Recruitment of data collection team.* Recruitment of the data collection team was limited to a pool of 70 data collectors who had EGRA experience, were trained and certified and met the 90% inter-rater reliability (IRR) gold standard, and had worked on the ACR GCD, Little Thinking Minds EGRA-based impact evaluation or USAID's Early Grade Reading and Mathematics Project (RAMP) EGRA-based impact evaluation.

- *Enumerator training.* The data collectors participated in an in-person data collection training delivered by evaluation/data collection experts, ensuring strict adherence to data collection protocols, especially those interfacing with children. This training reviewed EGRA protocols and subtasks, and focused on the delivery and testing of the SDQ tool.
- *Pre-Test.* The evaluation team pre-tested the EGRA and SDQ tools among Syrian children and their parents in non-camp settings, to ensure pilot testing for smooth data collection.
- *Emphasis on data quality during the data collection.* During the data collection process all supervisors and enumerators were informed that the quality of their data collection would be monitored by the data collection team in real time, making use of tablet-based data collection and real-time dashboards to monitor progress and enumerator performance. EGRA enumerators met the IRR gold standard (90%), and testing adhered to 8% IRR at baseline and 9% at endline to ensure consistency in data collection.

App Exposure and Dosage

For each app, a minimum dosage was developed in collaboration with the app developers and the humanitarian partner. A review of the literature revealed little consensus on the most effective dosage for DGBL education interventions; most, however, agree that dosage should be linked to the context of the intervention and that factors such as the quality of the intervention and the frequency and intensity of the sessions be considered.[99] Findings generally supported that more intensive interventions produced higher outcomes,[100] but cautioned that more frequent or intense doses may not lead to better outcomes, particularly if the quality of the intervention is subpar.[101]

In the context of this and other DGBL studies, it could be beneficial to further discuss how dosage is defined, and whether dosage should be evaluated by the number of hours spent playing the game, or by completion of the game itself. In considering the latter, dosage may be increased by adding or changing the levels and activities of the game itself, recognizing that dosage beyond completion of the game could have negative effects on engagement; however, adding levels to the game to increase dosage may in turn make the game too large for easy storage or download or negatively impact the game's simplicity. There may also be a place for a follow-up study on ideal dosage in a household setting versus a classroom setting.

A survey of similar interventions shows an average implementation period of two months, with most falling within the range of 20-30 hours.[102] Studies with a lower dose (10 hours or less) over the same period did not have effective results.[103] Integrated Services, Indigenous Solutions (INTEGRATED) proposed a maximum dosage of five hours per week (one hour per weekday) for seven weeks—for up to 30 hours maximum dosage. However, attempting this dosage within the two-month testing time frame posed some risks and challenges—further explored in the risks section.

Annex 3: Baseline Child Demographic Survey Data Analysis Tables

Gender					
	Male	Female	Total	Male%	Female%
Antura	194	172	366	53%	47%
Control	142	141	283	50%	50%
Grand Total	**336**	**313**			

Age													
	5	6	7	8	9	10	Total	5	6	7	8	9	10
Antura	5	65	61	83	70	82	366	1%	18%	17%	23%	19%	22%
Control	4	56	52	61	55	55	283	1%	20%	18%	22%	19%	19%
Grand Total	**9**	**121**	**113**	**144**	**125**	**137**	**649**						

When was the last time you attended school?				
	Antura	Control	Antura %	Control %
2007	0	1		
2013	1	0		
2014	1	1		
2015	5	8	1%	3%
2016	60	27	16%	10%
2017	273	224	75%	79%
I did not go to school	26	22	7%	8%
Total	**366**	**283**		

What was the last grade you attended in school?				
	Antura	Control	Antura %	Control %
1st Grade	192	160	56%	61%
2nd Grade	95	70	28%	27%
3rd Grade	36	30	11%	11%
4th Grade	16	1	5%	
5th Grade	1	0		
Total	**340**	**261**		

In which country did you last attend school?				
	Antura	Control	Antura %	Control %
Jordan	317	219	93%	84%
Syria	23	42	7%	16%
Total	**340**	**261**		

Who are you living with?				
	Antura	Control	Antura %	Control %
Parents	192	102	52%	36%
Father	116	169	32%	60%
Mother	56	11	15%	4%
Brothers	0	0		
Grandmother	2	0	1%	
Grandfather	0	1		
Uncle	0	0		
Total	**366**	**283**		

Do you live with your parents?				
	Antura	Control	Antura %	Control %
Yes	325	277	89%	98%
No	41	6	11%	2%
Total	**366**	**283**		

Do you have a smartphone?				
	Antura	Control	Antura %	Control %
Yes	285	237	78%	84%
No	81	46	22%	16%
Total	**366**	**283**		

Are you able to use the smartphone?				
	Antura	Control	Antura %	Control %
Yes	229	184	80%	78%
No	56	53	20%	22%
Total	**285**	**237**		

Do you know how and where your parents charge their phones?				
	Antura	Control	Antura %	Control %
Yes	272	228	95%	96%
No	13	9	5%	4%
Total	**285**	**237**		

Can your parents charge their smartphone easily?				
	Antura	Control	Antura %	Control %
Yes	232	219	85%	96%
No	40	9	15%	4%
Total	**272**	**228**		

Can they charge their smartphone anytime?				
	Antura	Control	Antura %	Control %
Yes	234	220	86%	96%
No	38	8	14%	4%
Total	**272**	**228**		

Do you have books at home?				
	Antura	Control	Antura %	Control %
Yes	116	156	32%	55%
No	250	127	68%	45%
Total	**366**	**283**		

Do you read at home?				
	Antura	Control	Antura %	Control %
Yes	90	133	78%	85%
No	26	23	22%	15%
Total	**116**	**156**		

How often do you read?				
	Antura	Control	Antura %	Control %
Daily	41	75	45.5%	56%
Once a Week	23	22	25.5%	17%
Twice a Week	26	36	29%	27%
Total	**90**	**133**		

Do your parents or siblings read books with you?				
	Antura	Control	Antura %	Control %
Yes	94	120	81%	77%
No	22	36	19%	23%
Total	**116**	**156**		

Can you read alone?				
	Antura	Control	Antura %	Control %
Yes	78	100	67%	64%
No	38	56	33%	36%
Total	**116**	**156**		

Do you like coming to class?				
	Antura	Control	Antura %	Control %
Yes	363	281	99%	99%
No	3	2	1%	1%
Total	**366**	**283**		

What do you like best about school? (choose all that apply)				
	Antura (366)	Control (283)	Antura %	Control %
I just like it	44	26	12%	9%
Teachers	65	63	18%	22%
Students	64	14	17%	5%
Games	81	35	22%	12%
Learning	324	234	89%	83%

Annex 4: Baseline Parent Demographic Survey Data Analysis Tables

Gender		
	Antura	Control
Female	79%	44%
Male	21%	56%

Are you the head of your household?		
	Antura	Control
No	38%	31%
Yes	62%	69%

What is your educational status?		
	Antura	Control
Bachelor	9%	8%
High School	6%	20%
Higher Education		2%
Illiterate	41%	27%
Literate but incomplete high school	43%	42%
Vocational Training	2%	3%

Do you have a smartphone?		
	Antura	Control
No	23%	14%
Yes	77%	86%

Are you aware of the learning app on your phone?		
	Antura	Control
No	26%	43%
Yes	74%	57%

Are you able to use a smartphone?		
	Antura	Control
No	25%	37%
Yes	75%	63%

Do you let your children use your smartphone?		
	Antura	Control
No	21%	13%
Yes	79%	87%

Do you have access to charge your phone?		
	Antura	Control
No	3%	2%
Yes	97%	98%

Can you charge your phone easily and anytime?		
	Antura	Control
No	34%	20%
Yes	66%	80%

Do you have books at home?		
	Antura	Control
No	80%	62%
Yes	20%	38%

Do you read with your child?		
	Antura	Control
Never	7.7%	2.1%
No	11.5%	57.5%
Yes	80.8%	40.4%

How often do you read with your child?		
	Antura	Control
Daily	48%	42%
Once a Week	24%	11%
Twice a Week	29%	47%

Annex 5: Fidelity of Implementation and Observation Summaries

Fidelity of Implementation Summary

	Attendance	Average # per classroom
1	**Number of students** *Record number: Look at the number of absences in the Attendance Sheet for each session*	13.5
2	**Class Duration** *Record minutes*	45
3	**Over-age students** *Record number: Number of students who are above 10 years*	0.5
4	**Under-age students** *Record number: Number of students who are under 5 years*	0

	Level	Average # per classroom
5	**Number of students who finished the game** *Record number*	0
6	**Number of students who played the game more than two times** *Record number*	0
7	**Number of students who played the game more than five times** *Record number*	0
8	**Number of students who played the game more than six times** *Record number*	0
9	**Number of students who played the game and got bored after finalizing it for the first time** *Record number*	0

	Technology	# of instances overall	
		YES	NO
10	**Tablets availability: Tablets are uploaded with the game (one game only)** *Record number*	10	0
11	**Tablets charged: Tablets are charged before the session** *Record number*	10	0
12	**Tablets broken: Tablets broken and children are not playing** *Record number*	0	10
13	**Electricity availability: Electricity is available to charge the tablets**	10	0

Quality of Delivery		# of instances overall	
		YES	NO
14	**Volunteer trained**	10	0
15	**Assistants trained**	6	4
16	**Volunteer support students: Volunteers are attending to students who need technological help and help in playing the game**	10	0
17	**Volunteer distributing and collecting tablets according to the name on the tablet**	10	0
18	**Students ask for help**	8	2
19	**Volunteer is same as initially trained one**	9	1
20	**Volunteer is adhering to the project as planned**	10	0

Child Participation		Average score per classroom
21	**Child engagement**	4.3
22	**Child interest**	4.3
23	**Request for help (child)**	2.8
24	**Interest in character**	3.9
25	**Interest in music**	4.0
26	**Interest in visuals**	3.9
27	**Interest in the levels**	3.8
28	**Interest in activities**	3.9

Observation Summary

No.	Indicator	# YES	% YES	# NO	% NO	# NA	% NA
1	Child is engaged in app (eye contact is on app, is playing with app, etc.)	114	97%	4	3%	0	0%
2	Child knows how to navigate (moves from stage-to-stage easily, finds different links, etc.)	112	95%	6	5%	0	0%
3	Child loses place on app frequently (more than three times)	4	3%	112	95%	2	2%
4	Child is smiling when playing (observed once)	76	64%	40	34%	2	2%
5	Child gets frustrated when using app (observed once)	11	9%	106	90%	1	1%
6	Child exhibits an understanding of the app	108	92%	8	7%	2	2%
7	Child likes receiving positive feedback from the app (smiles, calls out to volunteer for acknowledgment)	94	80%	19	16%	5	4%
8	Child loses interest after using app for 15 minutes	5	4%	111	94%	2	2%
9	Child asks for help using app	19	16%	96	81%	3	3%
10	Child listens to the story of the app	45	38%	69	58%	4	3%
11	Child laughs using the app	24	20%	90	76%	4	3%
12	Child likes the character of the app	107	91%	9	8%	2	2%
13	Child engages with other children using the app positively	88	75%	28	24%	2	2%
14	Child can follow verbal instruction of app easily	109	92%	7	6%	2	2%
15	Child can follow written words of app easily	85	72%	24	20%	9	8%
16	Child answers 50% or more of questions correctly	101	86%	14	12%	3	3%
17	Child knows when to click on button for next stage of app	113	96%	4	3%	1	1%
18	Child can hold tablet appropriately	114	97%	4	3%	0	0%
19	Child's fine motor skills assist in his/her playing the game	114	97%	4	3%	0	0%
20	Child is gaining a skill (moved levels, started to recognize a letter etc.)	114	97%	4	3%	0	0%

References

1 UNHCR. (2013, November 29). "The Future of Syria: Refugee Children in Crisis." Retrieved from http://www.unhcr.org/media-futureofsyria/

2 UNHCR. (2017, December 7). "Syria Emergency." Retrieved September 28, 2017 from http://www.unhcr.org/syria-emergency.html

3 UNICEF. (2017, March 8). "Hitting Rock Bottom: How 2016 became the worst year for Syria's children." Retrieved from http://www.refworld.org/docid/58c6bdc24.html

4 UNHCR. (2013, November 29). "The Future of Syria: Refugee Children in Crisis." Retrieved from http://www.unhcr.org/media-futureofsyria/

5 No Lost Generation. (2017, April). "Preparing for the Future of Children and Youth in Syria and the Region through Education: London One Year On." Retrieved from http://wos-education.org/uploads/reports/170331_Brussels_paper.pdf

6 *JBS International. (2014, June). *Mobiles for Reading: A Landscape Literature Review.* Retrieved from https://allchildrenreading.org/resources/mobiles-for-reading-a-landscape-research-review/

7 *JBS International. (2014, June). *Mobiles for Reading: A Landscape Literature Review.* Retrieved from https://allchildrenreading.org/resources/mobiles-for-reading-a-landscape-research-review/

8 Ibid., 5.

9 Lim, K. Y. T., Comings, J., Lee, R., Yuen, M. D., Hilmy, A., Chua, D., & Song, B. H. (2018). Guide on Developing Digital Games for Early Grade Literacy for Developing Countries. Quezon City, Philippines: Foundation for Information Technology Education and Development, Inc. and World Vision, Inc., p. 4.

10 Sutton, Michael. (2015, December 20). Game-Based Learning: Disruptive Approaches to Education, Learning, Training, and Development. Retrieved from https://www.linkedin.com/pulse/game-based-learning-disruptive-approaches-education-training-sutton

11 Weigel, Margaret. (2013, May 20). "Outcomes of Game-Based Learning: Research Roundup." Journalists Resource. Retrieved from https://journalistsresource.org/studies/society/education/outcomes-of-game-based-learning-research-roundup

12 Ibid.

13 Bösche, W., & Kattner, F. (2013). "Fear of (Serious) Digital Games and Game-Based Learning?: Causes, Consequences and a Possible Countermeasure." *Developments in Current Game-BasedLearning Design and Deployment,* p. 203-218, accessed 31 September 2017. doi:10.4018/978-1-4666-1864-0.

14 Sabourin, J., & Lester, J.C. (2014). "Affect and Engagement in Game-Based Learning Environments." *IEEE Transactions on Affective Computing, 5*(1).

15 Ibid.

16 Plass, Homer and Kinzer, "Foundations of Game-Based Learning," pp. 260.

17 Off-task behavior is defined as behavior "where a student completely disengages from the learning environment and task to engage in an unrelated behavior." Baker, R.S.J.d. (2007) Modeling and Understanding Students' Off-Task Behavior in Intelligent Tutoring Systems. Proceedings of ACM CHI 2007: Computer-Human Interaction, 1059-1068. Pdf.

18 Sabourin and Lestr, "Affect and Engagement in Game-Based Learning Environments."

19 Hamari, Juho et al. (2016, August). "Challenging games help students learn: An empirical study on engagement, flow, and immersion in game-based learning." Research Gate.

20 Ibid.

21 Lim, K. Y. T., Comings, J., Lee, R., Yuen, M. D., Hilmy, A., Chua, D., & Song, B. H. (2018). Guide on Developing Digital Games for Early Grade Literacy for Developing Countries. Quezon City, Philippines: Foundation for Information Technology Education and Development, Inc. and World Vision, Inc., p. 4.

22 Ibid., p. 12.

23 Engelhart, M.D., Furst, E.J., Hill, W.H. & Krathwohl, D.R. (1956). In Bloom, B.S. (Ed.) *Taxonomy of Educational Objectives Handbook I: The Cognitive Domain.* New York: David McKay Co. Inc.

24 Ibid.

25 Ertmer, P.A. & Newby, T.J. (1993). "Behaviorism, Cognitivism, Constructivism: Comparing Critical Features from an Instructional Design Perspective." *Performance Improvement Quarterly, 6(4),* p. 50–72.

26 Lim, K. Y. T., Comings, J., Lee, R., Yuen, M. D., Hilmy, A., Chua, D., & Song, B. H. (2018). Guide on Developing Digital Games for Early Grade Literacy for Developing Countries. Quezon City, Philippines: Foundation for Information Technology Education and Development, Inc. and World Vision, Inc., p. 17.

27 Ibid., p. 18.

28 Abu-Rabia, S. & Taha, H. (2016). "Reading in Arabic Orthography: Characteristics, Research Findings, and Assessment." In Joshi, R.M. (Ed.), *Handbook of Orthography and Literacy,* pp.324. Mahwah, NJ: Lawrence Erlbaum Associates.

29 Maamouri, M. (1999). "Arabic Literacy." Philadelphia, PA: Linguist Data Consortium-University of Pennsylvania. Retrieved from https://www.ldc.upenn.edu/sites/www.ldc.upenn.edu/files/arabic-literacy.pdf

30 Abadzi, H. (2012). "How to Speed Up Arabic Literacy for Lower-Income Students?" Global Partnership for Education Working Paper Series No. 9. Washington, D.C.: World Bank.

31 Abu-Rabia, S. (1999). "The effect of vowels on the reading comprehension of second and sixth-grade native Arab children." *Journal of Psycholinguistics Research, 28*(1), pp. 93-101.

32 Azzam, Rima. (1990). "The nature of Arabic reading and spelling errors of young children: a descriptive study." Unpublished doctoral dissertation. Teachers College, Columbia University: New York.

33 Maamouri, M. (1999). "Arabic Literacy." pp. 5. Philadelphia, PA: Linguist Data Consortium-University of Pennsylvania. Retrieved from https://www.ldc.upenn.edu/sites/www.ldc.upenn.edu/files/arabic-literacy.pdf

34 Ibid., 6.

35 Ibid., 7.

36 UNHCR. (2017). "Syria Regional Refugee Response, Inter-agency Information Sharing Portal." Retrieved September 31, 2017 from http://data.unhcr.org/syrianrefugees/regional.php

37 UNHCR. (2017, April). "UNHCR Jordan – Azraq Camp Fact Sheet." Retrieved from https://data2.unhcr.org/en/documents/details/59761

38 UNHCR. (2017). "Syria Regional Refugee Response, Inter-agency Information Sharing Portal." Retrieved September 31, 2017 from http://data.unhcr.org/syrianrefugees/regional.php

39 UNHCR. (2017, April). "UNHCR Jordan – Azraq Camp Fact Sheet." Retrieved from https://data2.unhcr.org/en/documents/details/59761

40 Staton, B. (2016, May 27). "Jordan detains Syrian refugees in Village 5 'Jail.'" IRIN News. Retrieved from http://www.irinnews.org/special-report/2016/05/27/jordan-detains-syrian-refugees-village-5-%E2%80%9C-jail%E2%80%9D.

41 UNHCR. (2017). "Jordan Education Sector Working Group. The Meeting Minutes of ESWG Coordination Meeting From ESWG." Retrieved May 14, 2017 from http://www.data.unhcr.org/syrianrefugees/download.php?id=13539

42 UNHCR. (2015). "Comprehensive Child Focused Assessment, Azraq Refugee Camp." pp. 36. Retrieved from http://www.data.unhcr.org/syrianrefugees/download.php?id=9148

43 King, G. & Nielsen, R. (2016). "Why Propensity Scores Should Not Be Used for Matching." This impact evaluation did not use propensity score matching to establish treatment and control group equivalence or match.

44 Figures taken from Relief International staff.

45 As cited by Antura and the Letters app developers.

46 Comings, J. (1995). "Literacy Skill Retention in Adult Students in Developing Countries," *International Journal of Educational Development*, *15*(1).

47 King, G. & Nielsen, R. (2016). "Why Propensity Scores Should Not Be Used for Matching." This impact evaluation did not use propensity score matching to establish treatment and control group equivalence or match.

48 Rate of change = (Gain/ Baseline)X100%

49 RTI International. (2015). "Early Grade Reading Assessment (EGRA) Toolkit, 2nd Edition." Washington, DC: USAID.

50 The Strengths and Difficulties Questionnaire scoring framework and analytics software can be found at http://www.sdqscore.org/

51 Information about the Strengths and Difficulties Questionnaire (SDQ), a brief behavioural screening questionnaire for children 3-16-year old's, can be found at: http://www.sdqinfo.org/d0.html

52 The MENA region, for the purposes of this study, refers primarily to that grouping of Middle Eastern countries united by the Arabic language, where Arabic is the primary language spoken in that country, roughly corresponding to the membership of the Arab League. All together there are 27 countries where Arabic is an official language, but 21 where Arabic is the official language with a majority of Arabic language speakers in: Algeria, Bahrain, Comoros, Egypt, Iraq, Jordan, Kuwait, Lebanon, Libya, Mauritania, Morocco, Oman, Qatar, Palestine, Saudi Arabia, Somalia, Sudan, Syria, Tunisia, United Arab Emirates, Yemen; and six countries where Arabic is an official language, but with a minority of Arabic speakers in: Chad, Djibouti, Eritrea, Israel, Malta, and Tanzania. It is important to note that some references cited in this study use slightly larger mapping of the MENA region to include countries such as Iran and Turkey.

53 UNICEF. (2015, September 3). "Education Under Fire: How conflict in the Middle East is depriving children of their schooling." pp. 3. Amman, Jordan: United Nations Children's Fund Regional Office for the Middle East & North Africa. Retrieved from https://www.unicef.org/mena/Education_Under_Fire.pdf.

54 MENA region is defined by GSMA includes Iran and Turkey.

55 Sutton, Michael. (2015, December 20). Game-Based Learning: Disruptive Approaches to Education, Learning, Training, and Development. Retrieved from https://www.linkedin.com/pulse/game-based-learning-disruptive-approaches-education-training-sutton

56 Adkins, S. (2017, July). "The 2017-2022 Global Game-based Learning Market." Metaari. Presented at the Serious Play Conference 2017. Retrieved from http://seriousplayconf.com/wp-content/uploads/2017/07/Metaari_2017-2022_Global_Game-based_Learning_Market_Executive_Overview.pdf

57 UNICEF. (2017, March 8). "Hitting Rock Bottom: How 2016 became the worst year for Syria's children," pp. 5. Retrieved from http://www.refworld.org/docid/58c6bdc24.html

58 World Bank. (2014). Education in the Middle East and North Africa: Brief. Retrieved from http://www.worldbank.org/en/region/mena/brief/education-in-mena

59 Sutton, Michael. (2015, December 20). Game-Based Learning: Disruptive Approaches to Education, Learning, Training, and Development. Retrieved from https://www.linkedin.com/pulse/game-based-learning-disruptive-approaches-education-training-sutton

60 JBS International. (2014, June). *Mobiles for Reading: A Landscape Literature Review.* pp. 14, 29. Retrieved from https://allchildrenreading.org/resources/mobiles-for-reading-a-landscape-research-review/

61 McEwan, Patrick. (2015). Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments. *Review of Educational Research*, *85*(3), p. 353-394

62 MENA region as defined by GSMA includes Iran and Turkey.

63 GSMA. (2016). The Mobile Economy Middle East and North Africa 2016. pp. 2. Retrieved from https://www.gsma.com/mobileeconomy/mena/

64 Ibid.

65 Kuruvilla, et al. (2017, May). Quoted in SAIS. "A War for Information: Use of Mobile and OTT Services among Syrian Refugees." *SAIS Perspectives*. Retrieved from http://www.saisperspectives.com/2017issue/2017/5/30/rfvuesw9cw5owfgm5st5asa1fh2302.

66 UNHCR. (2016, September). "Connecting Refugees: How Internet and Mobile Connectivity Can Improve Refugee Well-being and Humanitarian Assistance." UNHCR: Geneva, Switzerland.

67 REACH Initiative. (2015, December). "Mass Communications in Azraq Refugee Camp: Jordan Assessment Report." pp. 4.

68 Wall, M., Campbell M.O., and Janbek, D. (2015). "Syrian refugees and information precarity." *New Media and Society, 19*(2), pp. 6-7.

69 RTI International. (2012). "Student Performance in Reading and Mathematics, Pedagogic Practice, and School Management in Jordan." Retrieved from https://jordankmportal.com/resources/student-performance-in-reading-and-mathematics-pedagogic-practice-and-school-management-in-jordan-english

70 RTI International. (2014). "Girls' Improved Learning Outcomes: Final Report." USAID/Egypt. Retrieved from http://pdf.usaid.gov/pdf_docs/PA00JTBC.pdf

71 UNESCO. (2014). *Regional Report on Out-of-School Children*. Retrieved from http://uis.unesco.org/sites/default/files/documents/regional-report-on-out-of-school-children-middle-east-and-north-africa-2015-en.pdf.

72 Maamouri, Mohammed (1999). "Arabic Literacy." pp 4-5. Philadelphia, PA: Linguist Data Consortium -University of Pennsylvania. Retrieved from https://www.ldc.upenn.edu/sites/www.ldc.upenn.edu/files/arabic-literacy.pdf

73 Al-Yacoub, I. (2012, July 14). "Sum of all fears: Arabs read an average of 6 minutes a year, study reveals." Alarabiya. Retrieved from https://www.alarabiya.net/articles/2012/07/14/226290.html

74 UNICEF. "Middle East and North Africa Out of School Initiative: Update: Sudan." Retrieved September 30, 2017 from http://www.oosci-mena.org/sudan

75 War Child Holland. (2015, June). Innovative e-learning: learning mathematics in Sudan with an applied game. Retrieved from https://www.warchildholland.org/sites/default/files/bijlagen/node_14/18-2015/e-learning-sudan-june2015.pdf

76 Dahya, N. (2016). "Landscape Review: Education in Conflict and Crisis - How Can Technology Make a Difference?" Retrieved from http://www.ineesite.org/en/resources/landscape-review-education-in-conflict-and-crisis-how-can-technology-make-a

77 War Child Holland. (2015, June). Innovative e-learning: learning mathematics in Sudan with an applied game. Retrieved from https://www.warchildholland.org/sites/default/files/bijlagen/node_14/18-2015/e-learning-sudan-june2015.pdf

78 Ibid., 26.

79 Ibid., 34.

80 Ibid., 47.

81 Ibid., 54.

82 UNICEF. (2015). "Education Under Fire: How conflict in the Middle East is depriving children of their schooling." Retrieved from https://www.unicef.org/mena/Education_Under_Fire.pdf. p. 12.

83 Ibid.

84 Dahya, N. (2016). "Landscape Review: Education in Conflict and Crisis - How Can Technology Make a Difference?" Retrieved from http://www.ineesite.org/en/resources/landscape-review-education-in-conflict-and-crisis-how-can-technology-make-a

85 Interview with Kate Radford, cited in Dahya, N. (2016). "Landscape Review: Education in Conflict and Crisis - How Can Technology Make a Difference?" Retrieved from http://www.ineesite.org/en/resources/landscape-review-education-in-conflict-and-crisis-how-can-technology-make-a

86 UNICEF. (2015). "Education Under Fire: How conflict in the Middle East is depriving children of their schooling," pp. 26. Retrieved from https://www.unicef.org/mena/Education_Under_Fire.pdf.

87 Ibid., 27.

88 Dahya, N. (2016). "Landscape Review: Education in Conflict and Crisis - How Can Technology Make a Difference?" pp. 30. Retrieved from http://www.ineesite.org/en/resources/landscape-review-education-in-conflict-and-crisis-how-can-technology-make-a

89 Kangas, M, Koskinen, A., & Krokfors, L. (2017). "A Qualitative Literature Review of Educational Games in the Classroom: The Teacher's Pedagogical Activities." *Teachers and Teaching: Theory and Practice, 23*(4), pp. 451-470. Retrieved from https://eric.ed.gov/?id=EJ1131978.

Martinez-Garza, M.M., and Clark, D. (2013). "Teachers and teaching in game-based learning theory and practice." In Khine, M.S. & Saleh, M. (Ed.) *Approaches and strategies in next generation science learning,* pp. 147-163. Hershey, PA: IGI Global.

Martinez-Garza, M.M., Clark, D., & Nelson, B. (2013). "Digital games and the US National Research Council's science proficiency goals." *Studies in Science Education, 49*, pp. 170-208.

Ertmer, P. A., and Ottenbreit-Leftwich, A.T. (2010). "Teacher technology change: How knowledge, confidence, beliefs, and culture intersect." *Journal of Research on Technology in Education, 42*(3), pp. 255-284.

90 JBS International. (2014, June). *Mobiles for Reading: A Landscape Literature Review.* Retrieved from https://allchildrenreading.org/resources/mobiles-for-reading-a-landscape-research-review/

91 Sabourin, J., & Lester, J.C. (2014). "Affect and Engagement in Game-Based Learning Environments." *IEEE Transactions on Affective Computing, 5*(1).

92 Ibid., 45.

93 Ibid.

94 Mayer, R.E. "What Should Be the Role of Computer Games in Education?" 2016. *Policy Insights from the Behavioral and Brain Sciences3*(1), pp. 20-26.; Plass, J.L., Homer, B.D., & Kinzer, C.K. (2015). "Foundations of Game-Based Learning." *Educational Psychologist, 50*(4).

95 Mayer, R.E. "What Should Be the Role of Computer Games in Education?" 2016. *Policy Insights from the Behavioral and Brain Sciences3*(1)

96 Plass, J.L., Homer, B.D., & Kinzer, C.K. (2015). "Foundations of Game-Based Learning." *Educational Psychologist, 50*(4).

97 Plass, J.L., Homer, B.D., Hayward, E.O., Frye, J., Huang, T.T., Biles, M., Stein, M., &Perlin, K. (2012). "The Effect of Learning Mechanics Design on Learning Outcomes in a Computer-Based Geometry Game." *E-Learning and Games for Training, Education, Health, and Sports, 7516*, pp. 65-71.

98 Plass, J.L., Homer, B.D., & Kinzer, C.K. (2015). "Foundations of Game-Based Learning." *Educational Psychologist, 50*(4).

99 Wasik, B. A., Mattera, S. K., Lloyd, C. M., & Boller, K. (2013). "Intervention dosage in early childhood care and education: It's complicated." Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

100 Han, J. (2013). "Dosage effects on language, literacy, and general development for children enrolled in multiple early intervention programs: Head Start, Pre-Kindergarten, and Early Reading First." Dissertation, University of Georgia.

101 Wasik, B. A., Mattera, S. K., Lloyd, C. M., & Boller, K. (2013). "Intervention dosage in early childhood care and education: It's complicated." Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

102 Pitchford, N. (2015). "Development of early mathematical skills with a tablet intervention: a randomized control trial in Malawi." *Frontiers in Psychology, 6.*

Rosas, R., et al. (2003). "Beyond Nintendo: design and assessment of educational video games for first and second grade students." *Computers and Education40* (1). War Child Holland. (2016). "Research report: E-Learning Sudan Phases I and II." *War Child Holland, Ahfad University for Women and TNO.* Retrieved from https://www.warchildholland.org/sites/default/files/bijlagen/node_14/23-2016/research_report_e-learning_sudan_phases_i_and_ii_-_2016_-_online_version_spreads.pdf

103 Ronimus, M., Kujala, J., Tolvanen, A., & Lyytinen, H. (2014). "Children's engagement during digital game-based learning of reading: The effects of time, rewards, and challenge." *Computers & Education, 71.*


ALL
CHILDREN
READING:
A GRAND CHALLENGE
FOR DEVELOPMENT


DL4D
DIGITAL LEARNING
FOR DEVELOPMENT

unicef
INNOVATION

Norad

INTEGRATED
from insight to impact

create

USAID
FROM THE AMERICAN PEOPLE

World Vision

Australian
Aid

IDRC | CRDI
International Development Research Centre
Centre de recherches pour le développement international

UKaid
from the British people

FIT-ED