Feed the Monster

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

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). *Feed the Monster: 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
CIWPM	Correct Invented Words per Minute
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
RCT	Randomized Control Trial
REACH	Regional East African Community Health
RI	Relief International
RTI	Research Triangle Institution
SDQ	Strengths and Difficulties Questionnaire
SEL	Social Emotional Learning
SIM	Subscriber Identification Module
UNHCR	United Nations High Commissioner for Refugees
UNICEF	United Nations Children's Fund
USAID	United States Agency for International Development

Photo Credit: Integrated

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 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 for 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** (FTM).

ACR GCD, Digital Learning for Development (DL4D), and 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 (apps) 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 apps) to a group of children in matched environments who did not have access to the apps. 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 and DL4D to refine key app evaluation questions to 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 foremost experts in their specialization: assessing game use and engagement among children playing digital learning games. The contributions of CREATE ensured depth in qualitative gaming use data used while triangulating conclusions.

Key Findings: Literacy

- Overall, the game resulted in ***positive learning outcomes across all age groups and genders*** with a relatively low dosage of 22 hours. Absolute ***gains were greater in learning of foundational literacy (letters and syllables),*** and less with reading fluency.
- For children with previous exposure to basic literacy, the game, after letter and letter sound introduction, ***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***. Nevertheless, ***younger children showed higher rates of change*** across all subtasks.
- However, the fast pace and lack of differentiation between letters, their voweled variations (letters with vowel diacritics), and syllables may have resulted in ***confusion among children with low levels of literacy***. This was demonstrated by the increase in zero scores in Early Grade Reading Assessment (EGRA) subtasks among a subset of younger and/or less literate children. The game in its current form ***leaves room for child interpretation of decoding,*** a skill needed for increasing literacy outcomes. As a result, there is ***a need to better target phonics and decoding skills***, a key predictor of early reading, and reinforce these early reading skills in the game.
- While the ***Difference in Difference (DiD) assessment results were not statistically significant***, the learning gains and positive trends in ***learning outcomes as a result of game play show promise***, where potentially a larger sample size could detect significance with smaller differentials. These ***upward trends offer promise to better serve literacy outcomes once the game is modified*** to account for recommendations and a higher dosage and/or larger sample is assessed.
- 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 apps*** such as Feed the Monster.

Key Findings: Engagement and Psychosocial Outcomes

- The game is ***well designed in its simplicity and polish***, which enabled children of all ages and genders to ***easily and readily play with minimal or no adult supervision***. Children were easily able to navigate and complete the game, which gave them ***a sense of control and achievement***. Interaction with the monsters—with both correct and incorrect answers—***was entertaining to all children***, and the clear majority of children liked the game. ***Parents and teachers found the game appropriate*** and for the most part viewed its impact with their children as positive.
- The game ***clearly communicated progress both through rewards and incentives,*** such as the growth and collection of monsters, and differentiated performance through the number of stars collected at each stage.

- Game mechanics ***were consistent regardless of task***, which were timed similarly. The rewards and monster reactions had little variation, leading to a sense of repetition that may have reduced the level of sustained engagement, and pacing that may not have been appropriate to each child. As a result, ***engagement was relatively short-lived***, as upon completion of the game, it did not offer any further challenge to the children.
- At a psychosocial level, the game appears to have ***supported the development of positive social outcomes***, as demonstrated by both Strengths and Difficulties Questionnaire (SDQ) results and parental feedback. Children wanted to share the game and their learning with both siblings and peers, ***demonstrating impact on emotional states and social behaviors***. ***Peer interaction increased*** because of game play, while gaming engendered ***high levels of motivation to attend the summer camp***, which had implications for children's ***motivations to learn.***

1.1 General Literacy Recommendations

> Overall, the game resulted in positive learning outcomes across all age groups and genders with a relatively low dosage of 22 hours.

For younger, less literate children, the game could be improved by targeting both decoding and reading fluency skills, two predictors of early reading comprehension. The game should increase oral instructions and corrective feedback, allow for self-pacing or increased game time, pay attention to learning one task or literacy function at a time, and reinforce with repetition. For example, the game could introduce the entire alphabet with letter name *and* letter sound before moving into diacritics and syllables. In addition, the game should provide a better introduction of letters in the beginning, middle, and end position within a word, with clear linkages to their transformation, and better differentiate letter sounds versus letter sounds with diacritics.

For more literate children, the game could be improved by increasing opportunities for reading fluency practice. Game developers should consider building in a non-word/invented word section into the game to teach letter sounds and decoding skills without the ability to rely on sight or memory to guess the answers. For fluency, the game needs more game time for whole-word reading that eliminates guessing; inclusion of short phrases, sentences, and paragraphs that build on previously introduced vocabulary words; and gaming related to reading fluency and comprehension.

Given the promise shown with literacy improvements within the limited dosage, the game designers should consider testing the game again with a larger sample size or longer dosage period after revisions are made to the game to better assess its impact. The current impact evaluation literacy results show significant promise given the relatively low (22 hours) dosage.

1.2 General Engagement and Psychosocial Recommendations

Since the storyline at the beginning is powerful and appropriate, the game could make better links to the storyline throughout the game. As the game introduces new items (such as eggs, gifts, and points), integrating them into the storyline would likely generate increased interest and engagement. A positive conclusion to the storyline with all monsters would also likely generate more attachment to the game.

For all children, the game could benefit from increased personalization—both with respect to improved personal engagement in off-task play areas, as well as avatar personalization. This would likely result in increased engagement and sustained interest in game play.

To sustain engagement for learning, the game could benefit from improved pacing based on individual ability, introducing challenges for more advanced children and more time on task for early learners. This could reduce frustration and boredom, which are negative emotions that could diminish time on task or desire to continue game play.

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 engagement and time on task, as well as increase pro-social behaviors and psychosocial outcomes.

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 engagement and time on task, as well as 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 on the joint goal of game completion.

Photo Credit: Integrated

2. Introduction

2.1 Syrian Refugee Context

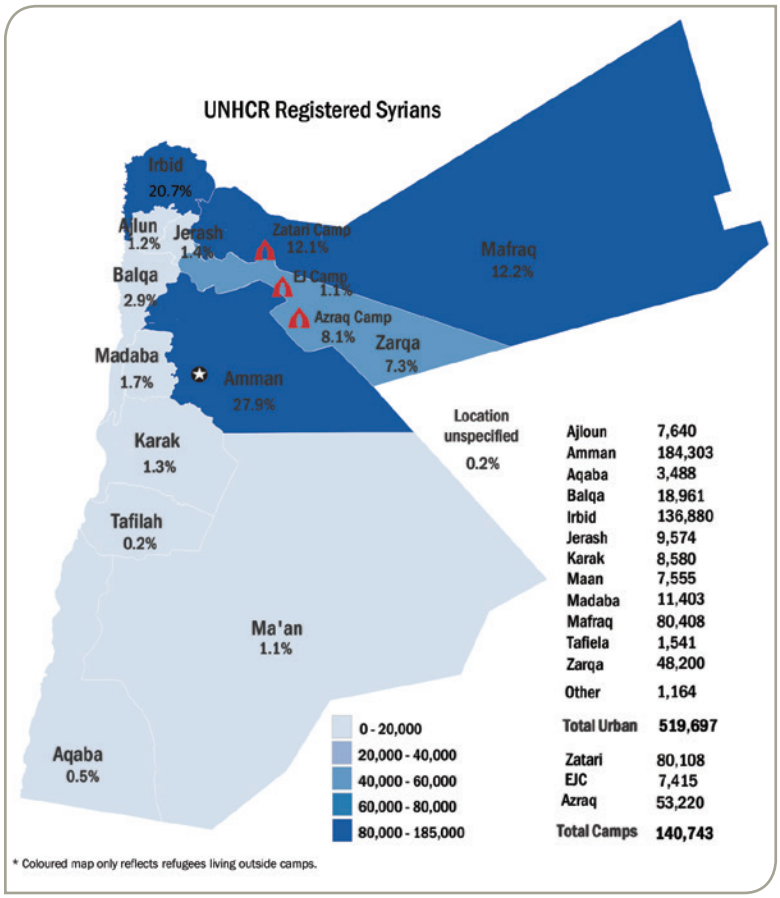


Figure 1: *Map of registered Syrian refugee populations in Jordan, courtesy of UNHCR.*

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. These children are displaced within Syria, refugees predominantly residing in Jordan, Lebanon and Turkey,[1] or in transit camps on their way to Europe. Since the Syrian Civil War broke out in January of 2011, it has displaced more than 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 a 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 the 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-Aged 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.

DGBL, an instructional method by which educational content or learning objectives are incorporated into digital games, is often considered by scholars as a leading method of learning. An innovative approach to learning, DGBL holds the possibility to address some of the challenges faced by educators in the Middle East and North Africa (MENA) region. Specifically, game-based learning offers potential in reducing barriers to access learning materials, alleviates the need for physical learning environments, and offers the opportunity for iterative and differentiated learning. Game-based learning also offers the opportunity 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 to provide such an evidence base. We have found no examples of rigorously-tested DGBL solutions for Arabic literacy with applicability to the Syrian refugee context. However, 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 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, using 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 found to be 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 and leads to the desired learning outcomes. The context in which the game will eventually be used must also be taken into account, as should an appropriate balance between fun and learning. For more literature 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 going to 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.

In seeking to develop a game that would fulfill the EduApp4Syria competition requirements, the team that developed the FTM app focused on three "pillars:"

- an engaging game experience;
- Arabic reading acquisition; and
- improvement in psychosocial well-being.

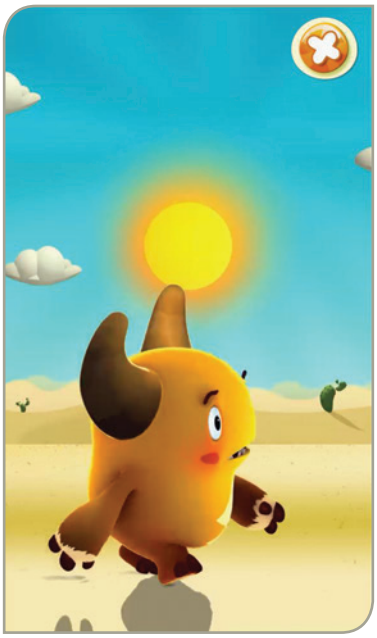


Figure 2: *The monsters being forced to leave their homelands.*

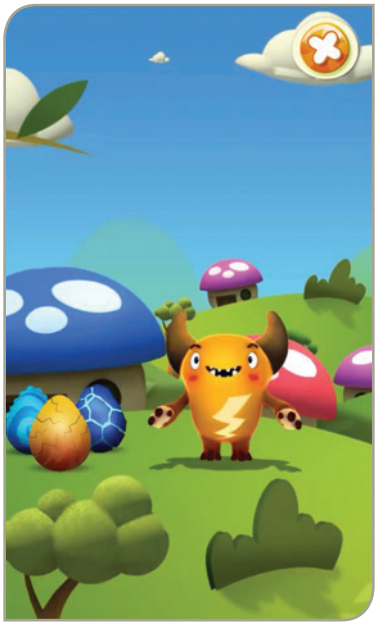


Figure 3: *Image of a grown monster (right) and eggs (left).*

The game's design was intended to engage children by providing a journey of discovery and friendship. Keeping Syrian refugee children in mind as the target audience, the game's storyline describes a fantasy world where friendly monsters were sent into exile by the evil "Harboot" who conquered their land and cast a magic spell to turn them all into eggs (*Figures 2 and 3)*. This storyline is designed to nourish hope within children as they find the eggs and feed them with Arabic letters, syllables, and words to help the monsters evolve, grow, and prosper.

To address the second pillar—Arabic reading acquisition—a team of Arabic literacy experts broke the Arabic Alphabet into small clusters of five to six letters each to match the main game mechanism (solving puzzles). In every cluster, the children must first learn the letters (shape and sounds), then practice the combination of letters and vowels (shape and sounds). At the end of each cluster, they must acquire new words (made of the letters already presented).[8] This process is intended to be simple and engaging and foster a sense among children that learning Arabic is fun.

Finally, to address psychosocial well-being, the game developers embedded the concept of Social Emotional Learning (SEL). Responding to research (that shows that children exposed to crisis-affected settings can experience "toxic stress" that damages executive cognitive functions), the FTM developers sought to provide a caring, nurturing, and predictable learning environment that builds children's social-emotional skills.[9] Within this game environment, children receive positive feedback as they interact with the friendly monsters and comfort and help their monsters when they are tired or upset.

Photo Credit: Integrated

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 randomized control trial design to estimate the impact of the EduApp4Syria games 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 (qualitative) 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 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 Feed the Monster 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 improving children's reading skills and vocabulary? (to be supported by EGRA findings)
Ways in Which the Use of Feed the Monster 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 and 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 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 to the beneficiary group? ■ What do you think of the game in terms of teaching literacy to 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 children's reading skills and vocabulary? In which ways can the game or use of the game be improved to develop children's reading skills and vocabulary? ■ Do Syrian parents and/or instructors consider the game and its content to be 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, communicate, attachment;
- Level of worry while playing the game;
- 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 an SDQ assessment among sampled children's parents, supported by a qualitative evaluation of the games 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 Feed the Monster 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 were the rewards used appropriate for efforts allocated and skills development? ■ How effective was the game in enabling children to achieve intended psychosocial outcomes among early grade learners in the communities targeted? ■ Are there observed gender differentials with respect to use of the games? (supported by SDQ findings)
Ways in Which Use of Feed the Monster 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 well-being, 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 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 skill 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.

> DGBL offers the potential to reduce barriers accessing learning materials, offers the opportunity for iterative and differentiated learning, and alleviates the need for physical learning environments.

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 literacy and psychosocial well-being assessment. Key evaluation questions in this domain were as follows *(Table 4):*

Table 4: *Key Gaming User Experience Evaluation Questions*

Ways in Which Use of Feed the Monster 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.

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 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 students with direct instruction "targeted at their individual learning needs, and adaptive to their learning progress."[10]

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, identification of lessons learned can result from game-based learning, where decisions and actions are chosen, consequences experienced, goals are achieved, and feedback furnished."[11]

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 children 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.[12] For example, some studies show that boys demonstrate "greater involvement with, liking of, and experience in computer gaming," yet learning gains and motivational benefits were found to be roughly equal for boys and girls.[13]

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.[14] 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.[15] 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."[16]

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."[17] 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—including boredom, confusion, frustration, and anger—can lead to decreased motivation and disengagement from the task. Disengagement can manifest itself as off-task behavior,[18] 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.[19]

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."[20] Still, other studies reveal that multi-player DGBL may enhance collaborative learning, group dynamics of positive interdependence, and an intergroup dynamic of competition.[21]

A handbook developed by ACR GCD 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.[22]

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.[23] 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.[24] These three domains of learning ensure the student is engaged on the levels of knowledge, attitudes, and skills, respectively.[25] 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.[26] 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 levels.

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.[27] 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]."[28]

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.[29] The Arabic alphabetic system has 28 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, a large amount which "leads to graphemic difficulty and a significant learning problem."[30] Studies reveal Arabic textbooks overwhelmingly teach reading through whole words, moving away from leveled introduction of letters, partial words, and phonetic learning and towards a desire for direct reading of unvoweled Arabic (consonants written without a symbol for the vowel they are connected to). This places complex demands on young children for language and literacy acquisition. Studies show whole language and whole word approaches are inefficient in lower-income environments.[31] EGRA conducted by Research Triangle Institution (RTI) in Jordan revealed that overall children's knowledge of letter-sound correspondence is low, and as a result, children's 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, and hamza) and can be written 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."[32] Research also found that both skilled and poor readers improved their reading accuracy when they read with vowels."[33]

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 literacy levels and learning achievement in schools.[34] School-aged children generally have about 2,000–5,000 words they comprehend orally and use in communication, which in most other languages forms the foundation on which language acquisition is built.[35] However, in Arabic, fuṣ-ha is the language of instruction, and as such places children at a disadvantage due to diglossia (use of high Arabic—fuṣ-ha—and low Arabic—colloquial). Proficiency in colloquial Arabic sometimes contributes to confusion and difficulty in learning connections, since chil-

dren 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."[36] The literature recommends using vocabulary drawn from the 40% overlap between fuṣ-ha and colloquial Arabic to reduce diglossia challenges.

Table 5: *Challenges to Arabic Literacy Acquisition*

1	The Arabic alphabet has 28 letters, many 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 ultimate 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 most of the new arrivals from Syria, and is the reception point for those Syrians coming in from the border.

Population Profile of Azraq Camp: While the registered population of Azraq Camp is 53,220, the number of residents in the camp is 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. 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).*

Table 6: *Population of Azraq Camp by Age/Gender*[39]

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%

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*

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 Azraq camp. Village 5 has extra security measures in place and a near-complete ban on residents' travel outside of the Village 5 perimeter fence. The Village 5 security measures are in place as a precaution in case 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, Azraq Camp has limited or no Wi-Fi/Internet access apart from weak reception in some areas near the main highway.

Early Grade Education in Azraq Camp: One of the largest challenges in 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 therefore fallen behind. In 2015, over 43% of children ages six to 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 (qualitative) evaluation.

4.1.1 Impact Evaluation

The impact evaluation used a longitudinal randomized control trial 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. The primary instrument used to assess literacy outcomes is the EGRA.

Early Grade Reading Assessment (EGRA): Literacy levels were measured through the EGRA instrument already developed and tested by the Research Triangle Institute (RTI) in Jordan and approved by the Ministry of Education in 2012. Using EGRA subtasks pertaining to basic literacy acquisition for Grades 1–3, the evaluation team conducted a baseline among control and treatment groups. The EGRA focused on the following subtasks corresponding to the FTM game *(Table 8):*

Table 8: *EGRA Subtask Description and Analysis*

EGRA Subtask	Description and Analysis
Letter-Sound Knowledge	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.
Invented Word Decoding	The score for this subtest is a measure of the number of Correct Invented Words Read per Minute (CIWPM). 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] across 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 game. Numbers and locations of children served by RI and 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: Children were selected according to age strata with these results *(Table 9):*

Table 9: *Child Selection Criterion*

Baseline	5 years	6 years	7 years	8 years	9 years	10 years	TOTAL
FTM Treatment	1.4%	20.4%	23.2%	21.1%	17.3%	16.5%	284
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**
FTM Treatment	2%	21%	23%	20%	19%	15%	199
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 within older (ages 8–10) versus younger (ages 5–7) age groupings, gender distribution, and EGRA performance at baseline. This enabled analysis as planned. **There was no statistically 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, the testing of the game was conducted in a structured manner within RI's 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 were instructed to help the children start navigating the game on their own with minimal intervention once game play began.

INTEGRATED conducted the impact evaluation among 284 treatment and 283 control children in the Azraq refugee camp in Jordan, within remedial education centers operated by RI in Village 6 (Feed the Monster) and Village 5 (Control).

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. RI targets children ages five to 18 and aims to provide multiple remedial pathways and learning support for children who have been out of school for several months or years. This program not only enables children to gain the basics of literacy and numeracy but also 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 children, within centers in Za'atari and Azraq Camps. Syrian refugees with teaching qualifications teach courses covering Grades 1 through13 from a central curriculum in Math, Science, Arabic, and English. The educational centers double as locations for child recreation in a safe, supervised space. Syrians are also recruited in the centers to conduct case management for children 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 the Azraq Camp. The Summer Club is not part of the required informal education program, but rather serves to strengthen children's 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, or recreation. For the course of the study, the children registered in the experimental group had "Tablet Class" as one of their two activity classes, whereas the control center in Village 5 continued in 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 among 284 treatment and 283 control children in the Azraq refugee camp in Jordan, within remedial education centers operated by RI in Village 6 (Feed the Monster) and Village 5 (Control). The children were provided with tablets uploaded with the game. The Syrian teachers and volunteers who ran the game testing sessions were 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.

Feed the Monster Testing: Within these classes, Syrian volunteers/teachers supported the roll-out of testing for Feed the Monster, played on tablets provided by ACR GCD and DL4D. Feed the Mon-

ster can be completed in three hours.[45] When children completed the game, they were asked to play multiple rounds.

Treatment and Control Settings: During the summer months in the Azraq Camp, official public schools are not operational. Nevertheless, RI continues to run summer club programming that includes remedial support and recreation classes. These remedial support classes are Arabic literacy, Math, and recreation activities. In this sense the control group was engaged in "business as usual" in remedial centers, and treatment groups used the game during a summer camp activity class.

The treatment and control groups were over-sampled to nearly 300 children per assessment group to account for up to 25% attrition. However, attrition levels experienced were higher than anticipated, reaching 33% among children and above 50% among parents.

Attrition: The treatment and control groups were over-sampled to nearly 300 children per assessment group to account for up to 25% attrition. However, attrition levels experienced were higher than anticipated, reaching 33% among children and above 50% among parents, caused by factors including:

- *General Factors:* The testing conditions in the summer camp in Azraq 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. As such, some children tested at baseline did not join the summer camp program, while others left the camp to spend the summer in other areas of Jordan. Others may have returned when the borders were re-opened in August 2017.
- *Village 6 (Feed the Monster):* Residents of Village 6 can secure permission to leave the camp for five days for every 40 days in residence. When permissions were obtained, families could leave the camp for "vacation leave" to visit relatives in Jordan. Children participating in the study who belonged to families able to leave the camp effectively dropped out of the study. These absences were particularly pronounced at the lead up to the Eid holiday, when the endline was conducted. Moreover, in August 2017, the borders to Syria were re-opened, allowing refugees to return home. Some families who had participated in the study may have chosen to return to Syria.
- *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 refugees who had moved. Where that information was available, the INTEGRATED team 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 parents and children who had participated in the study at baseline. This house-to-house endline data collection was conducted over three weeks from August 28 to September 20, 2017.

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 the criteria (age, years of schooling, gender) for selection of children, as well as household size and educational attainment for parents. Both parent and child surveys aimed to establish home conditions

for use of technology, such as access to electricity and presence of devices including smartphones, tablets, and laptops. This survey provided information 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 not be conclusive, 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. The 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, including focus group discussions (FGDs), observation checklists, and Key Informant Interviews (KII). Methods and their composition were selected to enable the collection of primary and secondary data enough 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 how ease of use could be improved).

The evaluation methodology consisted of: 1) desk review of DGBL on literacy, as well as regionally-relevant games designed to increase literacy; 2) semi-structured focus group discussions with children testing the game and their parents; 3) in-depth interviews and FGDs with RI; 4) observation of children using the games for gaming assessment (CREATE), and 5) play-intensive interviews with child beneficiaries (CREATE) *(Table 10).*

Table 10: *Evaluation Stakeholders and Data Collection to assess the impact on Literacy*

Stakeholders	Data Collection Conducted
Donors	■ 3 KIIs
Implementing Partners (RI)	■ 1 KII with RI Head of Education ■ 1 FGD with volunteers/teachers
Parents of Children in Treatment	■ 8 FGDs
Children in Treatment	■ 10 FGDs ■ 39 gaming observations ■ 17 play-focused interviews for each game
Game Designers	■ 1 KII

4.2 Game Exposure and Dosage

The testing period was scheduled to coincide with summer break when 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. Moreover, given that game play time as cited by the FTM game developers was three hours, the allocated dosage would enable the children to play multiple times.

Attempting this dosage within this two-month testing time frame posed some challenges:

- RI centers rely on generator-powered electricity. In Village 6, the generator broke down for a period of five to six days in August, preventing the charging of tablets and leading to 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. This amounted to two sessions.
- The centers did not always deliver the sessions six times during the week.
- The total dosage for FTM amounted to an average of 22.1 hours against the proposed 30.

Contamination	
FTM Treatment	2 children
Control	12 children

Contamination: The Azraq camp has severely restricted Internet access and connectivity, thereby limiting 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 (fenced in for security purposes) remaining physically separated from the other villages. 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. Nevertheless, it was discovered that some children/parents had downloaded the Antura and the Letters game (the other EduApp4Syria game). The survey team developed an exit survey for the endline assessment to establish the levels of contamination. Contaminated children EGRAs and their parents SDQs were removed from the study.

Software/Hardware: The game was operated on a combination of used and newly purchased tablets distributed and used within RI education centers. The game was uploaded offsite and given to RI for use in the testing period. Tablets provided by ACR GCD and DL4D were used once in the morning shift and once in the afternoon shift. Children were assigned avatars by gender, to ensure unique logins. Since Internet/Wi-Fi is mostly unavailable in Azraq camp, the games operated offline.

4.3 Limitations

Attrition was higher than the planned-for rate of 25%. Despite oversampling at baseline of 600 children (in treatment and control groups), at endline the final sample for treatment and control was about 200 children each, leading to an underpowered sample size.

Dosage was planned to be 31.5 hours, but no child completed the full planned dosage, since sessions were lost to electricity cuts from fuel shortages and generator breakdown. This prevented the charging of tablets. Because the dosage was not achieved, there may be limitations in assessing the full potential impact of the game with additional dosage. The below table shows correlation coefficients between EGRA scores at endline and dosage. It is a weak correlation for all subtasks; higher dosage would likely generate higher results.

EGRA Subtask	Correlation, Endline Scores to Dosage
Letter Sound Knowledge	0.0782
Syllable Reading	0.0919
Invented Word Decoding	0.0388
Oral Reading Fluency	0.1136

Testing conditions were fluid and not easily controlled, resulting in a situation where impact evaluation, design planning, and associated information were not always reflective of realities in the field. This was present during sampling, when the field team relied on 2016–2017 registration information 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 rollout period).

Self-reported data presents its own limitations, in that responses rely on the honesty and introspective ability or bias of respondents. In the context of refugee programming, this places some limitations on the validity of self-reported data and therefore must be considered.

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 regarding smartphone penetration, electricity access, current reading practices, parental education status, and use of home language. This was the demographic survey previously mentioned in section 4.1.1, Impact Evaluation. These contextual factors were collected to better understand the sample population. Some of the findings might help explain the contextual factors for the Feed the Monster game take-up and child performance.

Both parent and child surveys aimed to establish home conditions for use of technology, such as access to electricity and presence of devices including smartphones, tablets, and laptops.

***Educational Status of Parents*:** Of the parents surveyed in Village 6, 27% were illiterate (as compared to 27% in Village 5 Control). About half (52%) were literate but had not completed high school (as compared to 42% in Village 5 Control). One-fifth (20%) had completed high school or had a bachelor's degree (as compared to 28% in Village 5 Control).

***Access to Smartphone and Awareness of Learning Games*:** The majority (72%) of parents surveyed in Village 6 had a smartphone, as compared to 86% of parents surveyed in Village 5 Control. All parents surveyed in Village 6 could charge their phone, while 98% of parents in Village 5 Control could charge their phone. Most parents (59%) in Village 6 were aware of mobile learning games, as compared to 57% in Village 5 Control. Of those who had smartphones, 77% said they let their children use their smartphones, as compared to 87% in Village 5 Control.

Reading: Of the parents surveyed, 64% said they read with their children, as compared to 40% in Village 5 Control. Twenty-three percent of parents surveyed in Village 6 said they have books at home, while 38% of parents surveyed in Village 5 Control said they have books at home. Of the children surveyed in Village 6, 39% 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 element against the EGRA subtask. The table showcases how this game relates to each subtask, with game elements that correspond to reinforcing or learning the subtask tested. The accompanying description offers a breakdown of areas where the game offers learning through game play and potential areas where the child is required to independently interpret links or relations, which may lead to confusion or disorientation.

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

Related EGRA Task	FTM Element	Issues to Consider When Comparing EGRA to Feed the Monster Game Elements
Letter Sound Knowledge	Letter shape and sounds	The game introduces letters with their sounds, but not their names. When the game introduces the letters, some letters may sound the same as others in the monster's voice. Also, when the game introduces letters with diacritics, sometimes the letter and the letter with a diacritic have a similar sound, which can confuse children.
Syllable Reading	Letters with diacritics, syllable segments	The game shows letters as they appear in the beginning, middle, and end of the word but does not clearly link them to their original shape, or clearly show how letters join to form syllable segments, leaving the child to determine those relations by him/herself.
Invented Word Decoding	Letters with diacritics, syllable segments, new words	As above, there is a lack of instruction that enables early learners to clearly distinguish between letters, letters with diacritics, and syllable segments. The game focuses on letter sounds but not on how letters join or how letters transform depending on their placement in the word. While the game uses all forms of the individual letters, it does not make relationships or explain how/why these letters transform, leaving the child to determine that for him/herself.
Oral Reading Fluency	New words	The game introduces many words but does not introduce independent reading. Instead, the focus is on segmented spelling of words. Full words are introduced in the memory game, but the game lends itself to guessing, leaving the children read texts and improve fluency largely on their own.

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[46] on a range of characteristics. Characteristics of critical importance for establishment of equivalence among groups were:

- Syrian nationality
- Little or no schooling
- Ages five to 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. The treatment group included 284 children at baseline, while the control group started at 283 children. However, due to high drop-out rates, the sample was 199 and 200 in the treatment and control groups, respectively, at endline.

In total, there were 83 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 evaluation 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 A - C):*

Table A: *Sample Dropout by Gender and Group*

	Male	Female	Total
FTM	45	38	83
Control	30	41	71
Total	75	79	154

Table B: *Sample Dropout by Age and Group*

	5 Years	6 Years	7 Years	8 Years	9 Years	10 Years	Total
FTM	0	17	19	20	10	17	83
Control	3	11	12	13	18	14	71

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

Subtask	Group Type	Male	SD	Female	SD	Total	SD
Letter Sound Knowledge	Treatment	10.76	13.8862	11.37	16.99828	11.03	15.29502
	Control	15.03	17.9261	13.10	19.37112	13.92	18.66758
Syllable Reading	Treatment	9.49	12.7112	2.92	6.064436	6.48	10.6832
	Control	6.53	8.24928	6.34	11.57067	6.42	10.23253
Invented Word Decoding	Treatment	2.24	4.61793	0.94	2.065775	2.19	3.833288
	Control	2.3	4.07812	2.00	5.074446	2.13	4.650436
ORF	Treatment	7.53	10.8787	2.58	5.626345	5.27	9.162602
	Control	5.1	9.72147	4.22	7.938867	4.59	8.680977

Although the populations experienced somewhat 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 there was no statistical significant difference in the baseline population compared to the endline population.

5.2.2 EGRA Summary Results

Overall Results by Group Type—Treatment Versus Control: As illustrated in *Table 13,* the treatment group showed greater gains in performance from baseline to endline than the control group. The greatest improvement, an increase of 3.78, was on syllable reading subtask followed by oral reading fluency (ORF) and letter sound identification, which showed gains of 2.61 and 2.52, with a rate of change of 41% and 39%, respectively.

Table 13: *Subtask Results for Treatment and Control Groups*

Subtask	Group Type	Baseline		Endline		Total		Rate of Change[47]
		Mean	SD	Mean	SD	Gain	SD	
Letter Sound	**Treatment**	13.21	17.21245	15.73	21.22069	2.52	22.29029	19%
	Control	13.10	17.60739	13.12	18.64871	0.02	20.48555	0%
Syllable Reading	**Treatment**	9.18	12.28838	12.96	15.98812	3.78	13.79012	41%
	Control	7.85	11.90235	8.44	12.59203	0.59	11.57939	8%
Invented Word	**Treatment**	2.83	5.34572	3.52	6.309224	0.69	6.52203	24%
	Control	2.54	5.74136	2.25	4.827161	-0.29	5.090929	-11%
Oral Reading Fluency	**Treatment**	6.71	10.96521	9.32	13.682	2.61	12.68174	39%
	Control	6.43	11.30546	6.46	11.1038	0.03	8.851034	0%

There was virtually no gain for the treatment or control group on the invented word subtask, and both groups scored very low (3.52 correct words per minute [CWPM], 2.25 CWPM, respectively) at endline, indicating a lack of decoding skills. Although the overall gains for the treatment group were modest, the treatment group outperformed the control group on all subtasks showing that the game had some effect on literacy outcomes. Even though these gains are small, the rate of change—in percentage of improvement—was large, especially considering the low dosage of game play.

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 the gains between both groups on each subtask. As noted in *Table 14,* the difference between the gains of the treatment and control groups was significant on the Syllable Reading and ORF subtasks.

Table 14: *Inferential Comparison between the Total Gains of each group, by subtask*

Subtask	Group Type	Total				
		Gain	SD	t-statistic	Pr (ITI > Itl) 0.05	Significance
Letter Sound Knowledge	**Treatment**	2.52	22.290299	1.1677	0.2436	Not significant
	Control	0.02	20.48555			
Syllable Reading	**Treatment**	3.78	13.79012	2.5098	0.0125	Significant
	Control	0.59	11.57939			
Invented Word Decoding	**Treatment**	0.69	6.52203	1.6880	0.0922	Not Significant
	Control	-0.29	5.090929			
Oral Reading Fluency	**Treatment**	2.61	12.68174	2.3510	0.0192	Significant
	Control	0.03	8.851034			

There is currently no literature on dosage for DGBL, nor is there any research that estimates the relationship between impact and dosage for literacy instruction for children in the MENA region. A study by Comings that focused on adult literacy programs in Asia and Africa is commonly cited among education practitioners for early grade literacy, and it estimates that 250 hours of instruction is the minimum needed to produce meaningful reading skills.[48] This level of dosage was not possible within the two-month time frame of evaluation. 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 ORF.

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 ORF.

Effects of the Game on EGRA Scores: To estimate the impact of the game on treatment 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 FTM on children's achievement in this subtask in the treatment and control groups. The analysis helps measure the impact of FTM on EGRA scores after six weeks of implementing the game with the treatment group to test the hypothesis that children's literacy will improve upon introduction to the game. ORF is identified as the ultimate literacy outcome. ORF's importance is

supported by the simple view of comprehension,[49] which states that comprehension is a combination of reading fluency and oral vocabulary. Consequently, when children can fluently read text and have a strong oral vocabulary, they should score high on a comprehension test. Therefore, the mean scores of children on this subtask are being used to make a conclusion on literacy outcomes of children, within text made up of words in their oral vocabulary.

The team used panel data from 199 children in the treatment group as well as 200 children in the control group to examine FTM 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 control group children (effect size); and to attribute the results to FTM and not to chance or measurement error.

To calculate the effect size of the treatment, we used two variables. The first "treatment" is 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 if the scores were taken at endline and zero if it was taken at baseline. Then, a new variable, "DTrX," was generated 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.

As shown in *Figure 4,* the coefficient of the newly generated variable indicates that children in the treatment group scored 2.5 more CWPM than those in the control group when compared over time. However, the *p*-value is 12.5%. The difference between two groups (an experiment versus control group) is judged to be statistically significant when $p = 0.05$ or less. As such, the difference between control and treatment does not show statistical significance.

Figure 4: *Difference in Difference Results for ORF Subtask*

Source	SS	df	MS
Model	1169.19985	3	389.733284
Residual	110842.199	794	139.599746
Total	**112011.398**	**797**	**140.541278**

Number of obs = 798
F (3, 794) = 2.79
Prob > F = 0.0395
R-squared = 0.0104
Adj R-squared = 0.0067
Root MSE = 11.815

ORF	Coef.	Std. Err.	t	P> | t |	[95% Conf.	Interval]
post	.03	1.181523	0.03	0.980	-2.289279	2.349279
treatment	.2885678	1.183007	0.24	0.807	-2.033623	2.610758
DTrX	2.573015	1.673024	1.54	0.124	-.7110582	5.857088
_cons	6.425	.8354632	7.69	0.000	4.785022	8.064978

Effect Size of the Experiment, Comparison at Endline *(Table 15):* The assessment staff calculated an effect size to compare impact against that of school-based interventions. The assessment team 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 the difference between mean gains of the two groups are statistically significant if less than 0.05.

Table 15: *Effect Size of the Experiment and Control Groups When Comparing the Mean 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	2.502613	1.1677	0.2436	21.41445	0.116919	Not Significant
Syllable Reading	3.198945	2.5098	0.0125	12.81451	0.2512908	Significant
Invented Word Decoding	0.9884673	1.6880	0.0922	5.862186	0.1690088	Not Significant
Oral Reading Fluency	2.573015	2.3510	0.0192	10.99267	0.2353	Significant

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

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

The rate of change for girls in treatment was higher than boys in treatment on three of four subtasks: Syllable Reading, Invented Word, and Oral Reading Fluency.

Results by Gender: As evidenced in *Table 16,* boys scored higher than girls on all subtasks in the treatment group. In terms of overall gains among children in treatment, girls read more words per minute on the ORF subtask at endline, while boys achieved higher gains on letter sound. The rate of change for girls in treatment was higher than boys in treatment on three of four subtasks: Syllable Reading, Invented Word, and Oral Reading Fluency.

Table 16: *EGRA Results by Gender*

Subtask		Oral Reading Fluency		Invented Word Decoding		Syllable Reading		Letter Sound Knowledge	
Group Type		Control	Treatment	Control	Treatment	Control	Treatment	Control	Treatment
Baseline	Male	6.07	8.79	2.47	3.52	7.10	11.44	11.18	15.01
	SD	11.385	11.2620	5.8325	5.2124	11.802	13.428	16.388	18.458
	Female	6.81	4.74	2.61	2.18	8.67	7.02	15.18	11.50
	SD	11.266	10.3464	5.6706	5.41434	12.018	10.7233	18.703	15.8400
Endline	Male	7.12	10.90	2.47	4.22	8.27	15.26	9.88	18.54
	SD	12.285	13.1532	5.4219	6.45377	12.524	16.28971	16.659	21.71139
	Female	5.74	7.81	2.00	2.86	8.63	10.78	16.64	13.07
	SD	9.6772	14.15822	4.1013	6.181625	12.729	15.58555	20.089	20.61223
Gains	Male	1.05	2.10	0	0.70	1.17	3.81	-1.31	3.53
	SD	10.035	9.567561	5.5739	4.706347	11.697	12.10231	21.129	21.96872
	Female	-1.07	3.08	-0.61	0.69	-0.04	3.76	1.46	1.57
	SD	7.2493	15.04346	4.5175	7.882171	11.478	15.26948	19.774	22.69007
Rate of Change	Male	17%	24%	0%	20%	17%	33%	-12%	23%
	Female	-16%	65%	-23%	32	0%	54%	10%	14%

Results by Age Groups: *Table 17* shows the results disaggregated by age groups: children ages 5 to7 and ages 8 to10. Based on the endline results, treatment children ages 8 to 10 showed higher scores on all subtasks compared to treatment children ages 5 to 7. The highest gains in absolute values were for the older children in the treatment group, particularly for syllable reading (gains of 3.71 more syllables at endline for older children versus gains of 1.25 more syllables for younger children). However, the rate of change for younger treatment children was 34% on the syllable subtask versus a 27% change rate for the older children in treatment. As for ORF scores, older children in treatment gained 1.96 words compared to 1.30 for younger children in treatment, but the rate of change for younger children was higher than that for the older children: younger children achieved 75% increase from baseline, while older children achieved 18% increase from baseline.

Table 17: *EGRA Results by Age Groups*

Subtask		Letter Sound Knowledge		Syllable Reading		Invented Word Decoding		Oral Reading Fluency	
	Group Type	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control
5-7 Years	Baseline	8.02	5.69	3.65	2.83	0.91	1.04	1.74	2.08
	SD	11.7028	11.1284	7.43172	7.26256	2.78747	3.67234	5.77511	6.58555
	Endline	9.31	5.94	4.9	2.94	1.06	0.37	3.04	1.49
	SD	18.8468	15.1055	14.5199	9.82994	5.14192	3.20349	12.5826	6.95450
	Gains	1.29	0.25	1.25	0.11	0.15	-0.67	1.3	-0.59
	SD	19.3429	16.9326	12.5105	9.67363	4.65078	3.16189	11.6374	6.54941
	Rate of Change	16%	4%	34%	4%	16%	-64%	75%	-28%
8-10 Years	Baseline	17.58	18.14	13.83	11.27	4.4	3.56	10.91	9.37
	SD	19.7698	19.3790	13.6049	13.1926	6.37069	6.62128	12.4747	12.8182
	Endline	19.38	18.21	17.54	12.34	4.92	3.57	12.87	9.97
	SD	22.7108	20.5802	16.4343	13.5577	7.02796	5.52405	14.0748	12.6716
	Gains	1.8	0.07	3.71	1.07	0.52	0.01	1.96	0.60
	SD	24.4954	22.0794	14.7239	12.7032	7.71808	6.07131	13.3497	10.1145
	Rate of Change	10%	0%	27%	9%	12%	0%	18%	6%

5.2.3 Results Per Subtask

The results for each subtask are presented with overall mean scores, results by gender, and results by age groups. The analysis of results draws from the qualitative findings and expert knowledge to explain variations in performance among subgroups.

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.[50] The letter sound knowledge subtask assesses children's ability to identify the correct sound associated with each letter of the alphabet. Children were provided with a sheet of 100 letters that are different according to their placement in the words (ـبـ), (ـفـ), (ـهـ), and some diacritics to differentiate their pronunciation from the actual letter (و), (ـُ). They were then asked to name the sound of as many as they could within one minute.

According to the results in *Figure 5,* treatment children read an average of 13.2 correct letter sounds per minute (CLSPM) at baseline. At endline, the treatment group read 15.7 CLSPM showing a gain of 2.5 CLSPM. Children in the control group read 13.1 CLSPM, resulting in zero gains. The difference in mean gains (2.5 CLSPM) between the treatment and control groups was not significant, as can be seen in Table 15.

Figure 5: *Letter Sound Knowledge Results by Group Type*

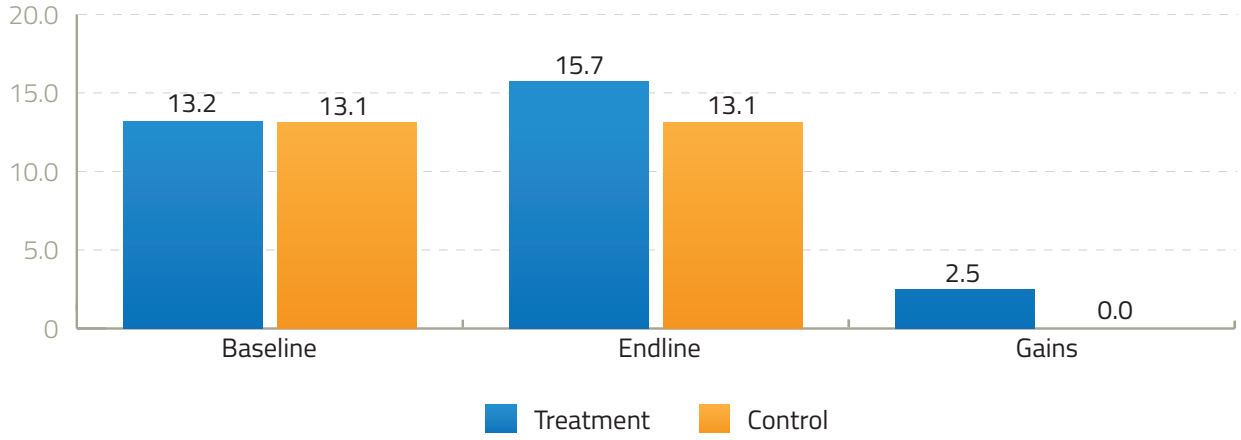


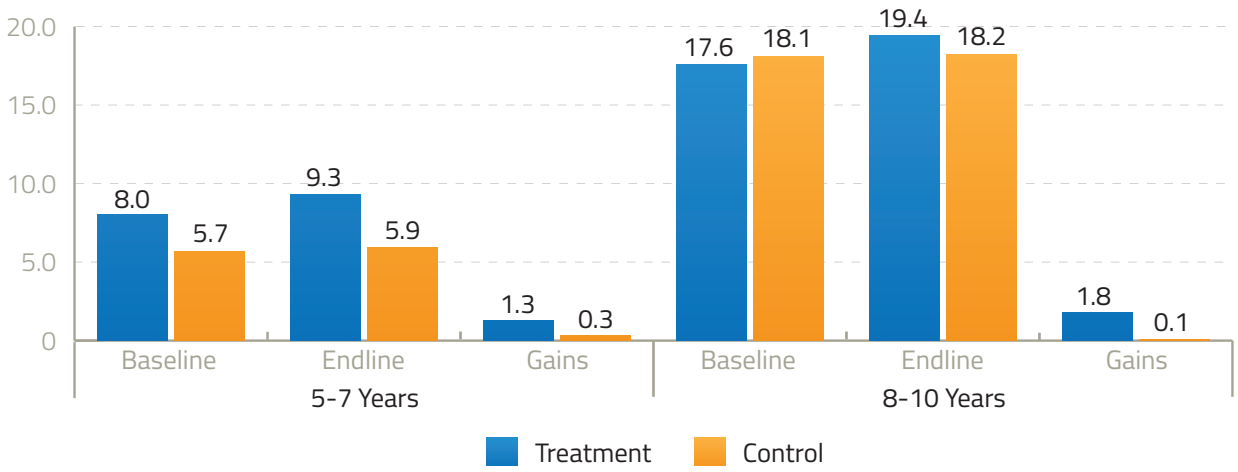
Table 18 shows letter sound results disaggregated by gender. In the treatment group, boys scored higher at baseline and endline than girls and showed higher gain scores and rates of change. Boys read 3.5 more letters at endline, while girls read 1.6 more letters. On the other hand, girls in the control group tended to score higher than boys at both baseline (by four letters) and endline (by six letters). The rate of change shows that boys in the control group had lower performance at baseline by 12%.

Table 18: *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	15.01	11.5	18.54	13.07	3.53	1.57	24%	14%
	Control	11.18	15.18	9.88	16.64	-1.3	1.46	-12%	10%

The letter sound results by age group presented in *Figure 6* indicate similar gains for treatment children in each age group. In the treatment group, children ages 5 to 7 identified 1.3 more letters at endline, while those ages 8 to 10 identified 1.8 more letters. Meanwhile, the control group showed no gains in both groups. In treatment groups, children in the older age group read more letters than children in the younger age group, suggesting older children have had more exposure to letters and thus more familiarity with letter sounds. However, younger children achieved a 16% increase in their scores, while older children achieved only a 10% change in their scores.

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

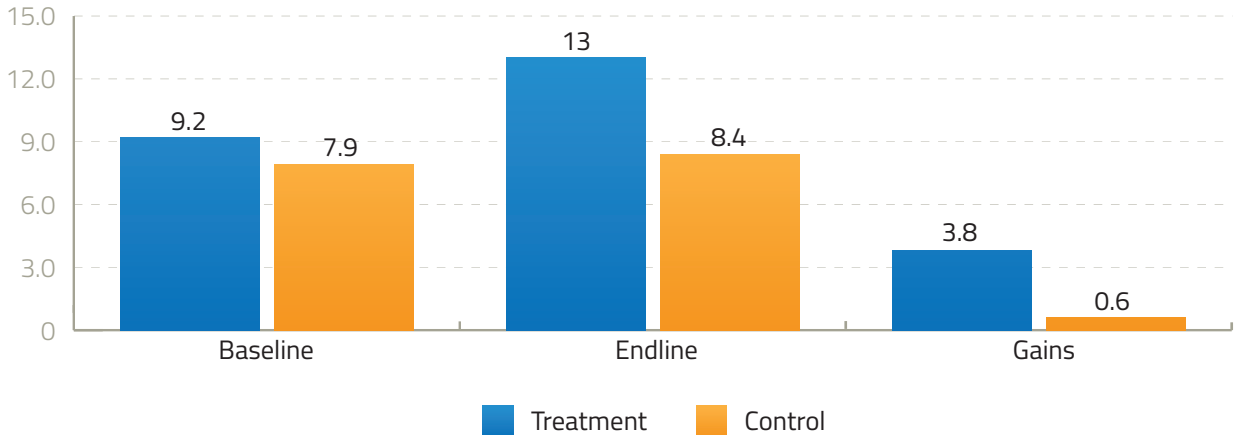


Analysis: Based on the results, the FTM game had only a small effect on improving letter sound knowledge among treatment children. In fact, the number of treatment children who scored zero in this subtask increased, indicating the game may have caused them to doubt their prior knowledge or that it confused them into selecting the wrong answer (See Section 5.2.4, Zero Scores). The qualitative findings showed that the fast-paced nature of the game and lack of instruction might have confounded the EGRA results. For instance, during the game, children are introduced to letter sounds, then letters with the diacritic, and then letters within a word. However, the game offers no explanation of how letters change and insufficient time for children with no or limited prior knowledge of the alphabet to grasp the concept before being asked to complete the task. As a result, children who are new to the alphabet and diacritics may not understand why the correct letter is incorrect without its correct corresponding diacritic (See Section 5.2.4, Zero Scores). Should the game correct for these shortcomings, the potential for increased gains is strong.

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's sound may vary depending on its placement in the syllable or word. 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 7,* children in the treatment group read 9.2 correct syllables per minute (CSPM) at baseline and 13 CSPM at endline, resulting in a gain of 3.8 additional syllables after playing the game. This translates into a 41% gain on the mean scores for this subtask, the largest gain among all subtasks and statistically significant when compared to the gain of the control group, as shown in Table 15. Children in the control group read 7.9 CSPM at baseline and 8.4 CSPM at endline, showing no difference in performance when marginal error is considered.

Figure 7: *Syllable Reading Results by Group Type*



In terms of results by gender in the treatment group *(Table 19),* boys scored higher than girls at baseline and endline. However, the gains were similar for both in terms of absolute values: after playing the game, boys read 3.82 additional syllables, while girls read 3.76 more syllables. However, girls in the treatment group achieved a higher rate of change in their scores, as they showed 54% change from baseline to endline, in comparison to 33% change for boys in treatment.

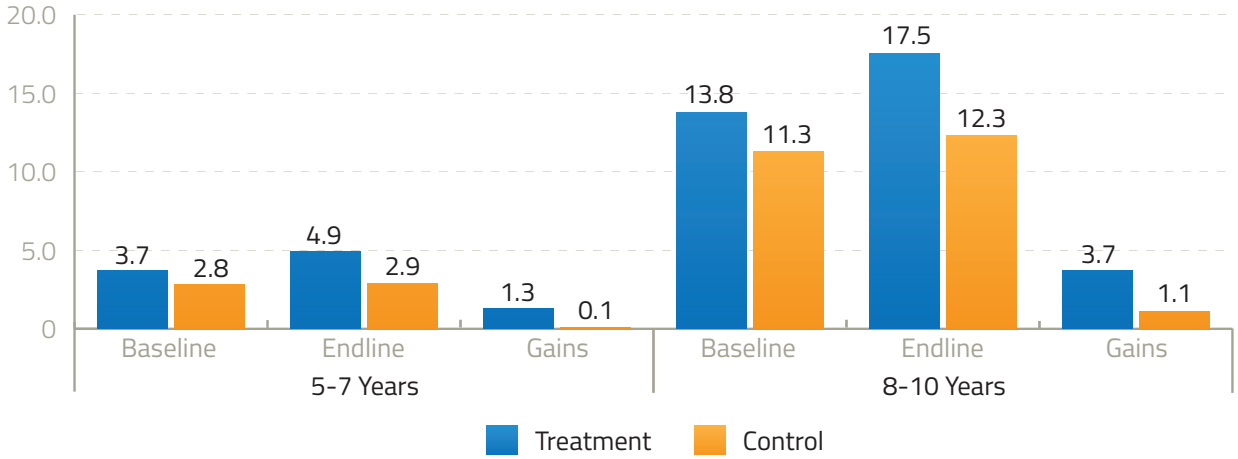
Table 19: *Syllable Reading Results by Gender*

Subtask	Group Type	Baseline		Endline		Gains		Rate of Change	
		Male	Female	Male	Female	Male	Female	Male	Female
Syllable Reading	Treatment	11.44	7.02	15.26	10.78	3.82	3.76	33%	54%
	Control	7.1	8.67	8.27	8.63	1.17	-0.04	17%	0%

The syllable reading results by age group *(Figure 8)* show older children performed better than younger. In the treatment group, participants aged 8 to 10 read 3.7 more syllables at endline with a 27% rate of change, compared to those aged 5 to 7 who read 1.3 more syllables with a 34% rate of change. Meanwhile, older children in the control group gained only one more syllable between baseline and

endline, while children in the younger age group did not show any gains. Like the letter sound subtask, older children had higher absolute gains than younger children, but younger children achieved a slightly higher rate of change on their scores from baseline to endline.

Figure 8: *Syllable Reading Results by Age Group*



Analysis: The syllable reading subtask showed the largest gains (41% rate of change and 3.8 syllables gained) among all subtasks. It also showed the highest rate of increase for boys and children aged 8 to 10, as well as the second highest rate of change among girls and children aged 5 to 7 in the treatment group. Nevertheless, this subtask also showed mixed results. While the treatment group recognized 3.8 more syllables than the control group, there was also an increase in the number of treatment children scoring zero on this subtask.

Qualitative findings revealed that some of the letters sound similar within the game, possibly causing difficultly for some children in distinguishing letter sounds. While the letter sound to letter match game reinforced letter-sound recognition, some letters with similar sounds in the Arabic language were difficult to differentiate in the game, especially during the syllable segment portion of the game. The following syllables were particularly difficult to distinguish: غو and خو or غُ and خُ. The increased number of zero scores among treatment children, therefore, may be a result of the syllables and letters selected in the instrument or due to children's lack of prior knowledge of letter sounds, letters with diacritics, and letter shape variations (See Section 5.2.4, Zero Scores). The game also lacked oral instruction and demonstrative examples of correct answers. Should the game correct for these shortcomings, the potential for increased gains is strong.

Invented Words Decoding: 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 sound correspondence and syllables to decode made-up words. Children 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 *Figure 9* indicate this was a very difficult task for all children. The treatment group read 2.8 correct invented words per minute (CIWPM) at baseline and 3.5 CIWPM at endline, demonstrating an increase of 0.7 CIWPM, or 24%. The performance of the control group decreased from baseline to endline by 0.29 CIWPM showing a -11% rate of change. As shown in Table 15, the increase in mean scores for the treatment group is not statistically significant when compared to the mean change in the scores of the control group.

Figure 9: *Invented Word Decoding Result by Group Type*

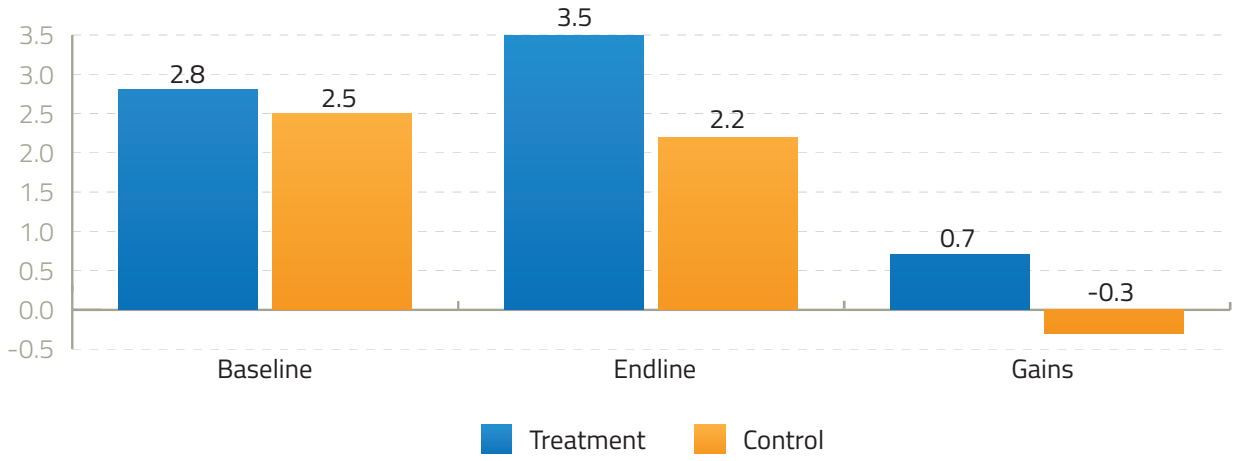


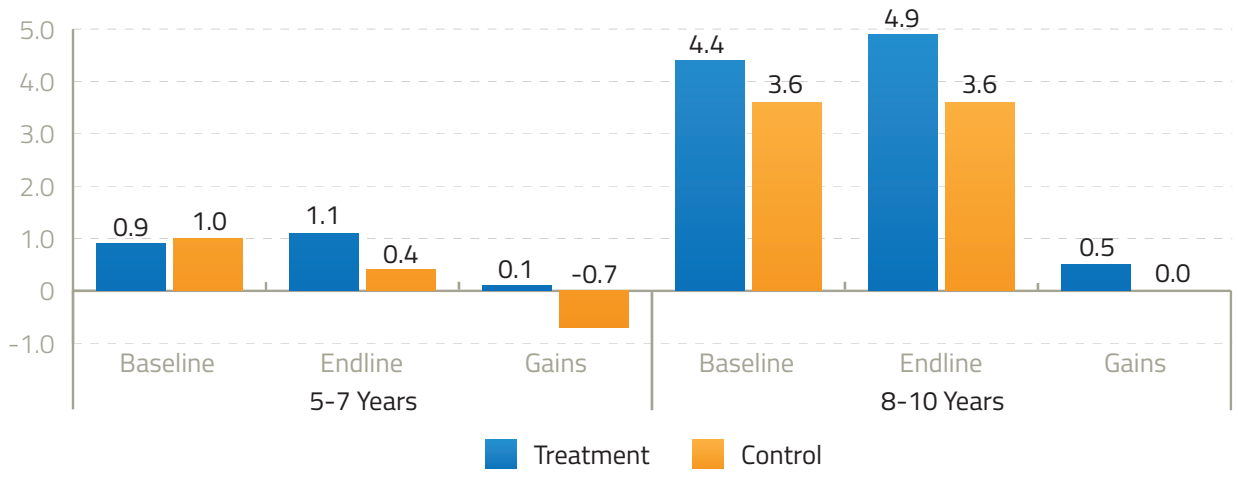
Table 20 shows that boys in the treatment group scored slightly higher than girls at baseline and endline, but the rate of change was higher for females than boys in both treatment and control groups.

Table 20: *Invented Word Results by Gender*

Subtask	Group Type	Baseline		Endline		Gains		Rate of Change	
		Male	Female	Male	Female	Male	Female	Male	Female
Invented Word	Treatment	3.52	2.18	4.22	2.86	0.7	0.68	20%	31%
	Control	2.47	2.61	2.47	2	0	-0.61	0%	-23%

The subtask results divided by age group in *Figure 10* suggest low mean scores and small gains for younger and older children in the treatment group.

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

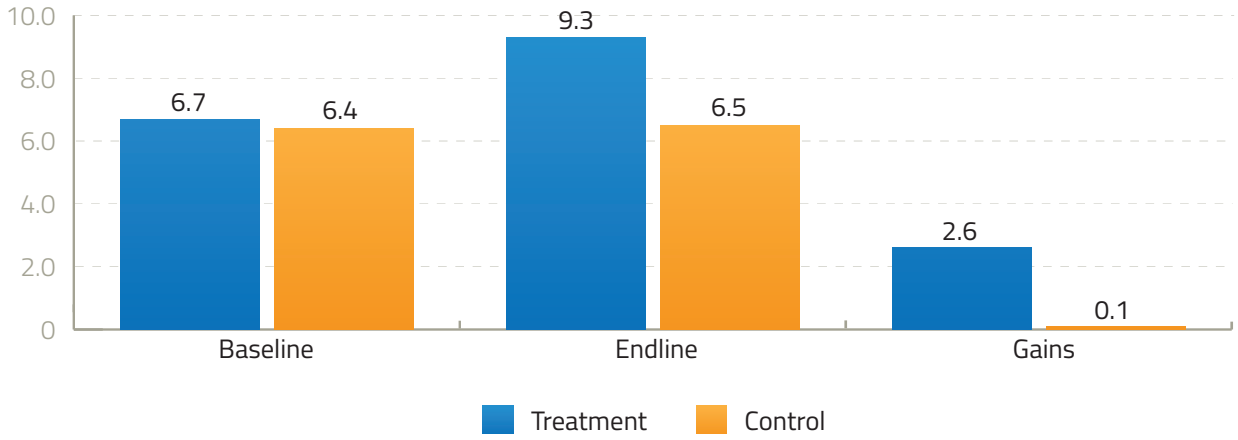


Analysis: Based on the results, the FTM game was mostly ineffective at enabling children to decode invented words accurately. This was likely due to the similarity of letter sounds, lack of instruction, and lack of reinforcement/repetition within the game (as described in letter sound and syllable reading subtasks). Most children in treatment did not master the previous skills and apply them to decode unfamiliar words. Additional support and instruction is needed to help all children improve their literacy skills. In the treatment group, the rates of change in mean scores for this subtask were 16% for the younger children and 12% for the older children. Should the game introduce decoding, there is potential for increased gains.

Oral Reading Fluency: 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 11* indicate the treatment group performed better than the control group. The treatment group read 6.7 correct words per minute (CWPM) at baseline and 9.3 at endline. This marked an increase of 2.6 CWPM compared to the control group, which read 6.4 words at baseline and 6.5 at endline for a gain of 0.1 (for further analysis, see Section 5.2.2, EGRA Summary Results). The increase of mean scores for the treatment group on ORF translate into a 39% increase on the scores from baseline to endline, while the rate of change for the control group was 0%.

Figure 11: *ORF Results by Group Type*



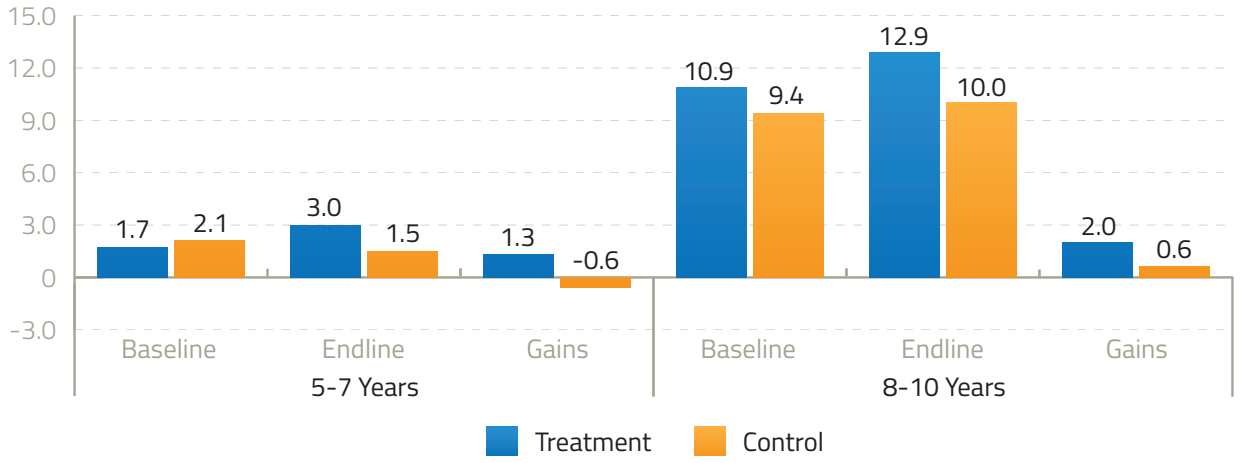
Boys in the treatment group performed better than girls at baseline and endline *(Table 21),* but girls showed higher rates of change than boys (65% compared to 24%).

Table 21: *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	8.79	4.74	10.9	7.81	2.11	3.07	24%	65%
	Control	6.07	6.81	7.12	5.74	1.05	-1.07	17%	-16%

According to the results disaggregated by age group *(Figure 12)* within the treatment group, children in the older age group showed higher absolute gains (2 more CWPM at endline compared to 1.3 more CWPM for the younger group). Older children further expressed higher levels of fluency. Children in the 8 to10 age group read 9–10 CWPM at baseline and 10–13 CWPM at endline, while children in the 5 to 7 age group read 1.7–2.1 CWPM at baseline and 1.5–3.0 CWPM at endline. However, in the treatment group, the rate of change for younger children was 75%, while the rate of change for older children was 18%.

Figure 12: *ORF Results by Age Group*



Analysis: Beginning readers often rely on sight recognition and memorization to identify words, especially if they have weak decoding abilities. Thus, it's no surprise early learners had relatively low fluency scores. Additionally, as the game's focus is on building letter sound knowledge and vocabulary, it does not include short sentences or stories to build fluency skills. While the results validate the qualitative study conclusion that the game reinforced knowledge of familiar words for older children, it did not significantly impact the acquisition of new words. Most notably, 70% of the children whose mean scores improved from baseline to endline were in the 8 to 10 age group, and 72% of children who showed no improvement from baseline to endline were in the 5 to7 age group. Nevertheless, absolute gains made by the older group were less than one word more than the younger age group. Also, the younger age group demonstrated the highest rate of change in this subtask, with a 75% improvement over baseline. This indicates that for older children, the game helped their reading fluency and somewhat reinforced the knowledge they already have, while for younger children it supported their reading fluency. To see a larger increase in absolute gains, more time and literacy exposure is likely required.

This indicates that for older children, the game helped their reading fluency and somewhat reinforced the knowledge they already have, while for younger children it supported their reading fluency.

5.2.4 Zero Scores

Despite higher mean scores at endline over baseline (*Table 13),* there was a higher number of zero scores across all subtasks for the treatment group (*Table 22).* The difference was most notable for letter sound knowledge, where 12%, or 24 more children, scored zero on the endline compared to the baseline (101 at endline compared to 77 at baseline). This could be explained by the games' issues with teaching letter sounds—namely its fast-paced nature and lack of clarity around letters, diacritics, and syllables that sounded similar. These prevented children from mastering skills before advancing to the next level.

Additionally, the high number of zero scores across all subtasks requires further analysis around the pedagogy embedded in the game. *Table 22* shows over half of the children in the treatment and control groups could not identify one letter sound or syllable; three-quarters could not decode one invented word; and over half of children in the treatment group and two-thirds in the control group could not read one word correctly in the ORF subtask. These results indicate that children need explicit instruction in letter sounds, decoding, and fluency and more opportunities to practice their reading skills. For example, the game teaches letter sounds, but does not provide adequate instruction or differentiation between letters in their original form and letters with diacritics, letters within a word, etc., which in turn may hinder a child's ability to correctly learn syllables and whole words. As Dr. Tim Shanahan[51] notes, a child must be offered clear and reinforcing instructions during brain processing to find the correct pathways to appropriately interpret the information. In the interim—until these pathways are clarified—the child may remain confused.[52] One hypothesis for the increase in zero scores could be that access to the previously correct pathways was disrupted by addition of too many new pathways. This might argue for slowing down the introduction of new letters, syllables, and words and increasing the repetition to make stronger pathways and avoid disrupting those that already exist.

Table 22: *Number of Children with Zero Scores*

Subtask	Treatment		Control		Difference	
	Baseline	Endline	Baseline	Endline	Treatment	Control
Letter Sound	77	101	107	110	24	3
Syllable Reading	87	93	117	111	6	-6
Invented Word	125	135	149	150	10	1
Oral Reading Fluency	104	109	128	120	5	-8

Figures 13–16 provide clearer comparison between the treatment and control group in terms of the number of children who scored zero on each subtask from baseline to endline.

Figure 13: Number of Children Scoring Zero on the Letter Subtask, by Group

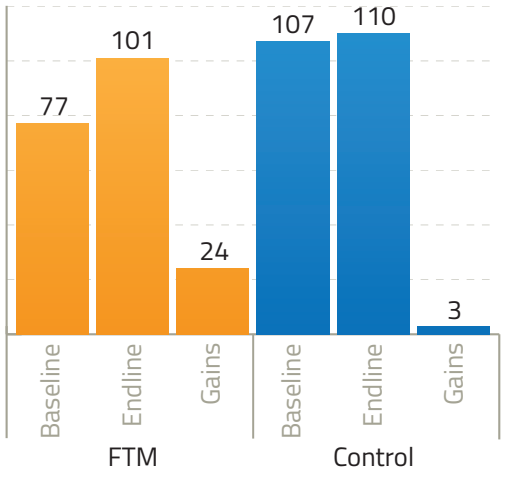


Figure 14: Number of Children Scoring Zero on the Syllable Subtask, by Group

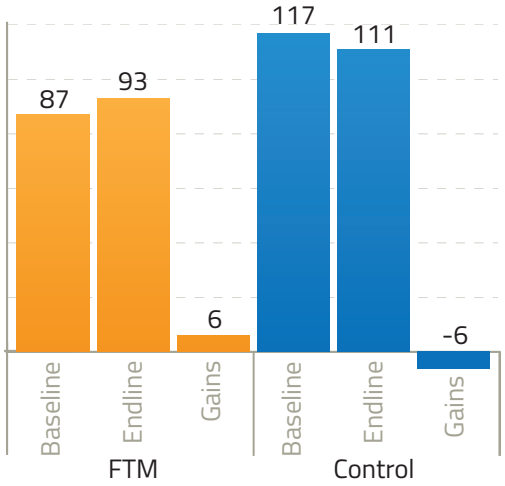


Figure 15: Number of Children Scoring Zero on the Invented Word Subtask, by Group

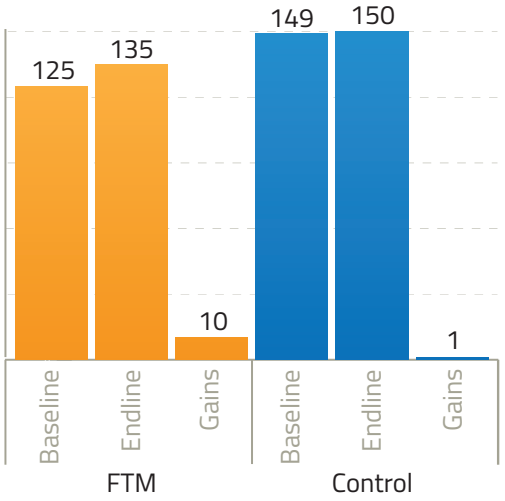
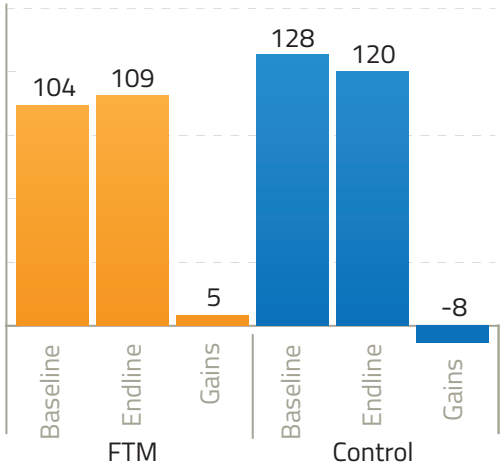


Figure 16: Number of Children Scoring Zero on the ORF Subtask, by Group



To better analyze this phenomenon, the evaluation team disaggregated treatment children according to performance on letter sound and oral fluency reading subtasks from baseline to endline, in three groups *(Table 23):*

- Children who scored zero at baseline and had some positive score at endline—*Children Who Improved*
- Children who scored zero at baseline and at endline—*Non-performers*
- Children who had some positive score at baseline but scored zero at endline—*Confused Children*

Table 23: *Number of Treatment Children with Zero Scores on Letter Sound and ORF by Attribute*

	Letter Sound Knowledge			Oral Fluency Reading		
	Improved	Non-performers	Confused	Improved	Non-performers	Confused
Number of Children	27	50	51	24	80	29
Average Age	7.7	7.54	7.6	7.7	6.88	8.48
Average Score: Baseline	0	0	15.5	0	0	11.59
Average Score: Endline	29.37	0	0	21.46	0	0
Ages 5–7	15	26	27	13	58	6
Ages 8–10	12	24	24	11	22	23
Females	10	26	32	14	49	13
Males	17	24	19	10	31	16
Grade 1 or No Schooling	12	33	33	11	54	15
Grade 2	9	12	11	6	23	8
Grade 3	6	4	7	7	3	6
Grade 4	0	1	0	0	0	0

Analysis of Confused Children: Considering performance on each subtask at endline and disaggregating the results by age group, we observed:

Letter Knowledge Subtask (51 children in total)

Age group 5 to 7 years old

- A total of 27 children scored zero on the letter subtask.
- 17 of the 27 children (63%) were in Grade 1 or with no prior schooling.
- 10 of 27 (37%) children were in Grades 2 or 3.
- 10 of the 27 (37%) children scored positively on at least one other subtask.

Age group 8 to 10 years old

- A total of 24 children scored zero on the letter subtask.
- Only one of the six children had zero scores on all subtasks.
- 18 of 24 children (75%) achieved positive scores on other subtasks.
- 16 of 24 (67%) children were in Grade 1 or with no prior schooling.
- Eight of 24 (33%) children were in Grade 2 or higher.

Conclusion: In the treatment group, children who scored zero at endline but not baseline showed some level of confusion. Those children tended to be younger (53%, ages 5–7) or in Grade 1 or no schooling (69%). Among the older children, most scoring zero at endline were in Grade 1 or with no schooling (67%). Moreover, the majority of older children (75%) scored positively on other subtasks, as compared to the minority (37%) of younger children who scored positively on other subtasks. No gender differentials were observed. We conclude that older children tended to get confused with the letter subtask but could perform on other subtasks. Among younger children, the lack of prior literacy exposure combined with lack of corrective feedback and insufficient letter differentiation in the game may contribute to the confusion, increase in zero scores, and lack of performance on other subtasks.

Oral Reading Fluency (ORF) Subtask (29 children in total)

Age group 5 to 7 years old (six children in total)

- Two of six (33%) children were in Grade 1 or without prior schooling.
- Four of six (67%) children were in Grade 2.
- Four of six (67%) children scored positively on at least one additional subtask.

Age group 8 to 10 years old (23 children in total)

- 10 of 23 (43%) children were able achieve some scores on at least one additional subtask.
- Eight children of 23 (34%) scored zero on all other subtasks.
- 13 of 23 (57%) children were in Grade 1 or without any schooling.
- Four of 23 (17%) children were in Grade 2.
- Six of 23 (26%) children were in Grade 3.

Conclusion: In the treatment group, children who scored zero at endline but not at baseline showed some level of confusion. Those children tended to be older (79%) and in Grade 1 or no schooling (51%). Among older children, less than half (43%) scored positively on other subtasks, as compared to the majority (67%) of younger children who scored positively on other subtasks. However, those four children in the 5 to 7 age group were in a higher grade (Grade 2). Among older children, just over half (57%) are in Grade 1 or have no schooling. No gender differentials were observed. We conclude that zero scores in ORF increase with children in early grades. The confusion with ORF seems to be related to lower literacy exposure regardless of age. Thus, the confusion that generated an increase in ORF zero scores is likely due to lack of literacy exposure/schooling and reading practice. As noted by Dr. Tim Shanahan, this confusion may be explained by the lack of clear and reinforcing instructions to aid the

brain in finding the correct pathways to appropriately interpret information.[53] For younger children and those with less formal instruction in the treatment group, the pathways may not yet be stabilized by instruction (in both game and educational environments), and the child may remain confused.

Analysis of Other Children (Non-performers and Children Who Improved): An analysis of zero scores, after removing confused children, shows an overall decrease in zero scores for letter sound and ORF subtasks. *Table 24* provides the number of children who scored zero on each subtask at baseline, after removing the confused children.

Table 24: *Number of Treatment Children with Zero Scores*

	Baseline	Endline	Difference	Rate of Change
Letter Sound	77	50	-27	-35%
ORF	104	80	-24	-23%

The letter sound reading subtask (after removing confused children from the treatment group) indicates that zero scores decreased by 35% from baseline within the treatment group. Children showing an improvement from baseline, tended to be in Grade 2 or higher (55%), as compared to non-performers, most of whom (66%) were in Grade 1 or had no schooling. With ORF, zero scores (after removing confused children from the treatment group) decreased by 23%. Children who demonstrated an improvement from baseline tended to be in Grade 2 or higher (54%), as compared to non-performers who were in Grade 1 or had no schooling (68%).

***Conclusion*:** As with the analysis of confused children, it appears improvements in literacy outcomes among treatment children are linked to previous literacy exposure. When children are at Grade 1 or less, they likely need more external reinforcement of literacy concepts and increased exposure or practice to make literacy gains. While this analysis (with the removal of confused children) cannot be compared to control group scores, it is nevertheless indicative of the possibilities of a decrease in zero scores as a result of game play, should the issues causing confusion be addressed.

Both baseline and endline EGRA tests had alphas at or close to 0.8, indicating a high consistency within both tests.

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 baseline and endline EGRA tests had alphas at or close to 0.8, indicating a high consistency within both tests.

Table 25: *Baseline Reliability*

Item	Obs	Sign	Item-test Correlation	Item-rest Correlation	Average Interitem Correlation	alpha
Letter Sound	122	+	0.8172	0.5021	0.7562	0.9030
Syllable Reading	112	+	0.9059	0.7595	0.5279	0.7703
Invented Word	74	+	0.9232	0.8427	0.5311	0.7727
Oral Reading Fluency	95	+	0.8615	0.6908	0.5976	0.8167
Test Scale					0.6062	0.8603

Table 26: *Endline Reliability*

Item	Obs	Sign	Item-test Correlation	Item-rest Correlation	Average Interitem Correlation	alpha
Letter Sound Knowledge	98	+	0.8641	0.6397	0.7520	0.9010
Syllable Reading	106	+	0.8989	0.7422	0.6183	0.8294
Invented Word Decoding	64	+	0.9286	0.8562	0.6014	0.8190
Oral Reading Fluency	90	+	0.9028	0.7701	0.6390	0.8415
Test Scale					0.6564	0.8843

Correlation

To check for similarity in the baseline and endline tests, INTEGRATED conducted rank ordered correlation tests for each of the four subtasks after eliminating zero scores. The correlations below (*Table* 27) show the degree to which endline scores were correlated with baseline scores and vice versa. It is a weak correlation on most subtasks.

Table 27: *Rank-ordered Correlation Coefficients*

EGRA Subtask	Correlation, Baseline to Endline score
Letter Sound Knowledge	0.3181
Syllable Reading	0.5840
Invented Word Decoding	0.6296
Oral Reading Fluency	0.7388

5.2.5 Factors Associated with Literacy Outcomes

Regression Model by Subtask: To provide more insight into the quantitative EGRA results, a regression analysis was conducted to assess the variables that contributed to improved performance for each of the subtasks among treatment children. The set of variables tested are in *Table 28*. The evaluation team began by running a regression analysis with all variables in *Table 28*, then weeded out non-significant predictors using a stepwise approach.

Table 28 : *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	Level of FTM that the child reached	Level 0–Level 86
Last Year Schooling	Last year the child attended formal education	■ 2015 ■ 2016 ■ 2017
Grade	Last grade level attended by the child	■ Grade 1 ■ Grade 2 ■ Grade 3 ■ Grade 4
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

Regression analysis was conducted for each subtask and with all above variables. With all subtasks, the statistically significant variables associated with improved performance were Age and Self-Reported Ability to Read Alone. As such, the regression analysis shows only those variables with statistical significance. *Figures 17–18* show the model for Letter Sound and ORF.

Letter Sound Knowledge: *Figure 17* shows the age and self-reported ability to read alone were the most significant factors associated with higher scores on this subtask, as was evident in the qualitative data.

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

Source	SS	df	MS
Model	5232.32403	2	2616.16202
Residual	34153.626	77	443.553584
Total	**39385.95**	**79**	**498.556329**

Number of obs = 80
F (, 77) = 5.90
Prob > F = 0.0041
R-squared = 0.1328
Adj R-squared = 0.1103
Root MSE = 21.061

Letter	Coef.	Std. Err.	t	P> | t |	[95% Conf.	Interval]
Age	5.183702	1.811634	2.86	0.005	1.576278	8.791125
Read_Alone	8.272376	4.875505	1.70	0.094	-1.435996	17.98075
_cons	-28.40566	14.70283	-1.93	0.057	-57.68274	.8714177

The coefficient of the Age variable suggests that for each year the child is older, an increased knowledge of 5.1 CWPM is generated. Similarly, a child who reported being able to read alone prior to playing the game can identify 8.2 more CWPM than a child who reported not being able to read alone.

ORF: *Figure 18* shows variables associated with oral reading fluency among treatment children. Age and Self-reported ability to read alone are again statistically significant variables associated with children's achievement on this subtask. For ORF, the level reached by children on FTM was a significant variable, although it affected their achievement on this subtask only slightly.

Figure 18: *Variables Associated with Better Achievement on Oral Reading Fluency*

Source	SS	df	MS
Model	4615.12562	3	1538.37521
Residual	9391.75749	73	128.654212
Total	**14006.8831**	**76**	**184.301094**

Number of obs = 77
F (3, 73) = 11.96
Prob > F = 0.0000
R-squared = 0.3295
Adj R-squared = 0.3019
Root MSE = 11.343

ORF	Coef.	Std. Err.	t	P> | t |	[95% Conf.	Interval]
Age	5.075942	.9806102	5.18	0.000	3.121589	7.030295
Read_Alone	5.903389	2.696899	2.19	0.032	.5284766	11.2783
Level	.0818473	.0449858	1.82	0.073	-.0078092	.1715038
_cons	-39.59108	8.645579	-4.58	0.000	-56.8217	-22.36047

The coefficients of the variables associated with achievement on this subtask suggest children tend to be able to read 5.1–5.9 more CWPM if they are older and reported they could read alone at baseline. Although a significant variable, the level tended to have less effect on child achievement as it helps them read only 0.9 CWPM, showing little effect of the game on mean scores for this subtask.

5.3 Qualitative Findings

5.3.1 Feed the Monster Game Play

Figure 19: *Image of the eggs that children can hatch to introduce new monsters.*

FTM was introduced to 283 children in Azraq camp during the summer camp program operated by RI in 2017. At endline, 201 children's results were captured, with 82 children lost to attrition. Children played Feed the Monster for 45 minutes per day, five times a week for seven weeks, for an average dosage of 22.1 hours, with girls playing 22.3 hours on average and boys playing 21.8 hours on average. Children playing FTM had an average attendance rate of 71.2%.

Children enter the game by choosing an animal avatar. The game does not assess or ask the level of literacy of the child, and does not adjust for skill or literacy level. Children start at the same level and progress based on their achievement. FTM divides the Arabic Alphabet into small clusters of five to six letters each. In each cluster, the child:

- is introduced to the letter and its sound;
- must feed the monster the correct letter based on matching the letter to a copy of the letter first, then matching the letter to its sound;
- is introduced to the letter's voweled variations (diacritical markings) and their sound;
- must feed the monster the correct letter and diacritic combination;
- is introduced to the letter in a syllable segment in its written form and sound;
- must feed the monster the correct syllable;
- is introduced to the single letter or voweled variation of that letter; or a syllable segment or letter sequence within a word and its sound;
- must feed the monster the correct letter, voweled variation, syllable segment, or letter sequence;
- is introduced to a word using the letters in the cluster and its sound; and
- must feed the monster the letters in the correct spelling order.

When these tasks are completed, the child moves on to the next letter grouping. Each letter grouping period lasts 15 to 20 minutes. Correct answers make the monster happy and earn the child one to three stars, depending on performance. The child earns points throughout the game and periodically earns new monster eggs (*Figure 19*) and special items to feed the monster. If the child gets all answers incorrect, the monster is sad, but the child can make the monster happy by petting it. A monster also becomes sad when it is neglected, and each new monster displays a different negative emotion to be addressed through play.

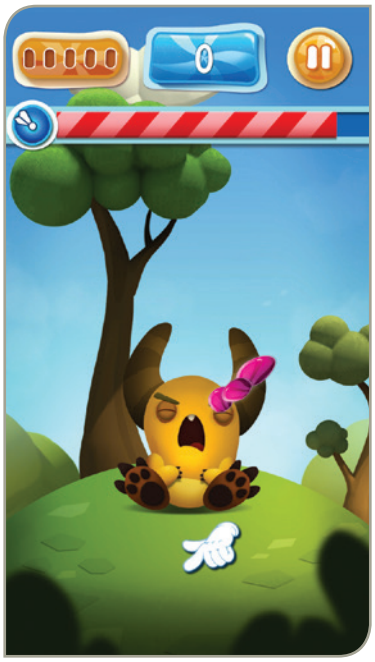


Figure 20: *A monster spitting out an incorrect answer.*

Corrective feedback is given by the monster spitting out the incorrect answer *(Figure 20)*. There is no opportunity to self-correct, nor does the game demonstrate the correct answer when an incorrect one is given. There is no instruction or demonstration on what the correct answer is, or why the incorrect choice made by the child was incorrect. Guessing can be rewarded, as in some game play the choices offered are all correct. To reinforce learning, there is a need to provide options to self-correct, or provide options for the game to demonstrate the correct answer. In all above tasks, the child can progress through the game, even if they feed the monster incorrect letters.

The above approach offers the child an opportunity to see and hear the letter sound and match the letter once to its shape and again to its sound. The game does not name the letter, so the child is not introduced to the letter name. When the child is introduced to the letter's voweled variation, the reason for the introduction of the diacritic is not clear, and this stage comes shortly after the letter is introduced. All letters in that grouping are introduced with voweled variations (with diacritics). The child must feed the monster the correct letter and vowel combination (letter and diacritic combination), and incorrect options of just the letter (without the correct diacritic) are offered as incorrect options to feed the monster (*Figure 21*). If the correct letter with the incorrect diacritic is fed to the monster, the monster spits it out as incorrect. Since this stage comes immediately after the introduction of the letter, it may confuse children who are new to the alphabet and diacritics. For children with little or no literacy exposure, there is a need to better introduce and distinguish letters and their voweled variations (diacritics). However, for children with previous exposure to diacritics, this can be reinforcing because it allows children to recall the correct voweled variation of the letter.

Figure 21: *Game featuring letters with and without diacritics.*

Shortly thereafter, the letter is introduced within a syllable segment, but the letter may have changed shape within the syllable segment. The game does not explain how it has changed shape or how it is the same letter. There is also no explanation offered as to why some letters connect and some do not. A mechanism is needed to explain how letters connect/don't connect for children with no prior literacy knowledge. Moreover, the sounds of the letter, voweled variation, and syllable can sometimes sound similar, thus potentially confusing children between the three. For children with previous exposure to the alphabet, diacritics, and letter shapes based on placement within the word, these exercises can be reinforcing, since it enables recall of the correct letter shapes based on word placement. For children with no prior literacy exposure, the game must better distinguish between letter sound, voweled variation, and syllable.

While the letter sound to letter match segment reinforced letter-sound recognition, some letters sound similar in the Arabic language and may be difficult to distinguish in the audio of the game, such as (ظ and ث), (س and ص), (غ and خ), (ط and ت). This becomes particularly challeng-

ing during the syllable segment portion of the game when syllable segment sounds are difficult to distinguish, such as غو and خو. This potentially poses challenges for the child, particularly if he or she has little to no prior knowledge of the alphabet, diacritics, and letter shape variations. The game may need to better distinguish between letters with similar sounds.

Figure 22: *Game featuring letters with voweled variations.*

When letters are introduced within the word, the whole word is not spoken—only the desired letter sound. The same is applied to voweled variations and syllable segments. When the whole word is introduced, this is the first time the child hears the whole word —whereas until that point the child only sees the whole word, and is orally asked for one part of it (letter, voweled variation, or syllable segment). The child must feed the monster the correct shape of the letter based on its placement in the word, or the correct letter-syllable combination based on its placement in the word (*Figure 22*). This may offer some confusion to the child who cannot read, as there is no explanation why the desired letter is now surrounded by two to three other letters, or why the correct letter is incorrect if it does not assume the correct shape based on placement in the word. The same confusion can apply to the syllable segment. For children with no prior exposure to early literacy, the game should offer better pacing and breakdown and/or build-up of the word through gradual introduction, connection, and sounding out of letters coming together to build a word. However, for those with previous literacy exposure, the current approach allows the child to read the syllable segment as part of the word, or spell the word letter by letter or by syllable segment, which can be reinforcing.

The words the children were exposed to in the game were inspired by their cultural surroundings, and they enable the child to expand vocabulary. Play-focused interviews revealed that 15 of 16 child participants agreed that words used in the game were generally relevant to real-world usage, indicating that FTM successfully addressed the discrepancy between fuṣ-ha and colloquial dialects.

The current version also has basic letter puzzles, letter-in-a-word puzzles, and full-word puzzles. Extra mini-games are unlocked when certain monsters are hatched and begin to grow. These mini-games include: Petting Game, Letter Tracing, Memory Game, and a Word-to-Picture Matching Game. However, if a child did not unlock each monster, the child could not play the corresponding mini-games. The game also does not indicate that such mini-games are available to be unlocked. In the Beta (final) version, the developers intend to expand to sentences and short reading paragraphs.

5.3.2 Children's Literacy Interface with Feed the Monster

Focus group discussions with children targeted two different age groups, ages 5 to7 and ages 8 to 10. Children reported in FGDs that the game taught them letters and different words using all the letters learned. Most children reported they could correctly spell words. This finding is confirmed in the engagement observations conducted by INTEGRATED, in which 93% of children could correctly answer 50% or more of the questions posed by the game.

Figure 23: *Game featuring longer words.*

While children generally succeeded in the learning challenges presented in the game, it became more difficult as the game progressed. Children reported in FGDs that as words got harder, or when they were unable to correctly identify a letter in the middle of a word, the game became challenging. Children also stated, "When you know the word you are able to pick out the letters," reflecting that prior literacy is helpful to improve or advance game play. Almost all children noted that they could spell words with up to five letters but faced more difficulty when words were longer (five to seven letters) (*Figure 23*). These experiences were also reflected in the gaming observations conducted by CREATE, where children were seen having more difficulty with learning content in later weeks (*Table 29*). While Arabic reading acquisition was challenging for some, however, engagement with the game's content rose over time, and most children could recognize letters through in the game.

Table 29: *Instances of Difficulty with Learning Content*

	Earlier Weeks	Later Weeks
Overall learning content	Nine instances across 11 observation periods	64 instances across 11 observation periods
Letter-to-letter matching levels	Four instances across 11 observation periods	20 instances across 27 observation periods

Although FTM included immediate feedback for players, children generally did not understand from this feedback why their answers were wrong or how to improve. Reactions to feedback were rare, with children more often demonstrating disengagement with feedback. More than half (three of five) of the play-intensive interviewees stated they did not know why their answers were wrong in the game.

Although most parents in focus groups were aware their children were playing a literacy game, very few parents had played the game with their child. In focus groups with parents, 28 of 32 parents felt their children were indeed learning from the game. Although some parents mentioned the game was difficult for their children at the beginning, most indicated the game was easy to play for children. One parent mentioned the game is harder for her younger child. Some teachers and volunteers noted that children who were academically challenged were more attached to the game but took longer to find solutions, whereas stronger children would move quickly through the game but get bored. When asked if the game takes into consideration the different levels of literacy among children, some volunteers/teachers felt it did, since it starts from a very basic level and becomes increasingly difficult. Therefore, the children who are more advanced made progress quickly through the phases, whereas the less advanced started learning from the beginning and made progress at their own pace. One volunteer suggested having two different games, one for early grades and one for Grades 3 and 4:

"Children in the first and second grade faced problems in using the tablet and playing the game more than children in the third and fourth grade." Volunteers and teachers said the game was an effective tool that improved the children's spelling and their ability to combine letters and syllables into words. All volunteers/teachers agreed there was an improvement in letter recognition, with most (seven of eight) indicating there has been an improvement in vocabulary; however, none (zero of eight) felt the game improved reading fluency.

When asked what parents felt their children were learning, parents said their children were learning how to identify and read letters and words, and how to combine letters, distinction of letters, and other auxiliaries of reading (*harakat, hamza, tanween*). Some parents mentioned improvement in spelling, reading, writing, and pronunciation, with one parent stating that her child started reading letters and words on television. Still others noted their children were now tracing letters and spelling. On the other hand, some parents indicated their children learned how to use the tablets and mobile phones.

When asked if parents noted any changes in literacy performance over the last six weeks, most parents (29 of 35) in FGDs stated their children's literacy skills had improved. Of those who noted improvement, they cited improvements in reading, writing, pronunciation, and spelling, in addition to increased motivation to learn and do homework and increased independence in reading and writing.

Of 98 children playing FTM, 100% had finished the game at least once, 31% had finished the game at least twice, and 3% had finished the game more than five times (this self-reported data was collected the fifth week of testing). Many said they improved from two-star completion to three-star completion, showcasing the ability of children to improve in the tasks of the game. On the last play, boys on average reached level 62 of 86, while girls on average reached level 67 of 86. Although children were progressing, many FGD participants did not want to play again after collecting and feeding all the monsters; instead, they wanted new levels, new challenges, or some sort of variation. Four of 11 children in play-focused interviews expressed interest in more game content, with some asking specifically for more advanced levels.

Parents and teachers recommended the game extend to reading full sentences and short paragraphs. The KII with game developers noted the need for challenge and variation. Planned future versions will add new levels using a level editor to add or delete items without code change. This should enable the developers to focus on new types of game content to keep children engaged.

Although many children participating in FGDs expressed a desire for more variation, 14 of 17 children interviewed one-on-one still said they would continue to play FTM at home, even though over 10 of those 17 children have other mobile gaming options in their homes. Half of those who said they would continue playing at home named learning as their motivation.

Further, almost two-thirds of parents (24 of 38) in FGDs said they were willing to let their children use their mobile phones to play the game. When asked how much time they would allocate to game play, most parents said they were willing to give the child the phone for one hour per day.

5.4 Conclusions

Overall, the game resulted in positive learning outcomes across all age groups and genders with a relatively low dosage of 22 hours. More specifically, the game supports the qualitative findings of im-

provement in literacy outcomes, in that knowledge of letter sound, syllable reading, and oral reading fluency improved among the sampled population over control, with absolute gains greater in learning of foundational literacy (letters and syllables) and less with reading fluency. Since the game's focus is on building letter-sound knowledge and vocabulary and does not include short sentences or stories to build fluency skills, these EGRA trends match the game.

Figure 24: *Children playing FTM in Azraq Camp.*

FTM effectively introduces letters and letter sounds in small groupings for all children regardless of literacy level. For children with previous exposure to basic literacy, the game, after letter and letter sound introduction, effectively reinforces previously skills acquired. With respect to syllable reading, analysis showed that 67.9% of children who improved from baseline to endline were in aged 8 to 10 years, while a large proportion of younger children showed either no improvement or decline in the mean scores. Most notably, with respect to ORF, 70% of children whose mean scores improved from baseline to endline were aged 8 to 10, while 72% of children who showed no improvement from baseline to endline were ages 5 to 7. Regression analysis across subtasks confirms this, as variables influencing positive performance showed statistical significance with Age and Self-Reported Ability to Read Alone. Nevertheless, while older children with more exposure to literacy showed the highest absolute gains, younger children showed higher rates of change across all subtasks. This indicates that for older children, the game is helping reading fluency and reinforcing knowledge they already have. For younger children, the game modestly improved literacy levels, but more time or literacy exposure is required to make larger absolute gains.

These upward trends offer promise for improved literacy outcomes once the game is modified to consider recommendations and a longer dosage and testing period can be instituted.

FTM, however, does not offer sufficient corrective feedback for learning and can generate confusion. The game also enables progress even if incorrect answers are given or if the child guesses his/her way through. The game could improve with well-placed diagnostics that send children back to earlier levels to reinforce foundational skills. The fast pace and lack of clearly explained differentiation between letters, voweled variations, and syllables may have resulted in confusion among children with low levels of literacy, as demonstrated by the increase in zero scores in EGRA subtasks among a subset of younger and/or less-literate children. An analysis of children whose zero scores increased disaggregated by age group demonstrated the younger age group scored a higher percentage of zeros compared to those in the older age group with letter sound; in ORF, those children whose zero scores increased tended to be in Grade 1 or with no schooling (albeit older). As a result, the game should target phonics and decoding skills—a key predictor of early reading—and reinforce early reading skills in the game.

While FTM was reinforcing for children with prior literacy knowledge, it could improve by focusing on individual EGRA subtasks; breaking the literacy acquisition down with more instruction; and increased gaming spent on differences between letters, letters with diacritics, letters within syllables and their relationships. Specifically, with respect to skills the EGRA subtasks assessed, the game in its current form leaves room for child interpretation of decoding skills that are needed for improving literacy outcomes. Where children had previous literacy learning, that interpretation was known in advance. Children without that prior knowledge were left to interpret linkages and decoding, due to possibilities for guessing as well as the lack of corrective feedback and instruction on letter relationships.

While the DiD results were not statistically significant, the learning gains and positive trends in learning outcomes because of game play showed promise, where potentially a larger sample size could detect significance with smaller differentials. Alternatively, potentially a larger dosage may also result in gains in DiD results. These upward trends offer promise for improved literacy outcomes once the game is modified to consider recommendations and a longer dosage and testing period can be instituted.

Finally, the rate of smartphone penetration among Syrians participating in the study is very high, with most allowing their children access to their smartphone. This has promising implications for the widespread download and use of literacy games such as Feed the Monster. While over half the surveyed population were aware of learning games, increased awareness of games for learning among this population is necessary.

5.5 Recommendations

For younger, less- literate children, the game could be improved by offering features known to boost decoding and reading fluency skills, two predictors of early reading comprehension. Other recommendations include: increasing oral instructions and corrective feedback; allowing for self-pacing or increased game time; targeting one learning task or literacy function at a time; and reinforcing with repetition. For example, the entire alphabet could be introduced with letter name and letter sound before moving into diacritics and syllables. Also, better introduction of letters in the beginning, middle, and end of word could be made with clear linkages to their transformation. The game could also better differentiate letter sounds versus letter sounds with diacritics.

For more literate children, the game could be improved by increasing opportunities for reading fluency. Consider building a non-word/invented word section into the game to teach letter sounds and decoding skills without the ability to rely on sight or memory to guess the answers. For fluency, consider increasing more game time to whole word reading to eliminate guessing; or include short phrases, sentences, and paragraphs that build on previously introduced vocabulary words and provide gaming related to reading fluency and comprehension.

Given the promise shown with literacy improvements within the limited dosage, consider testing the game again with a larger sample size or longer dosage period after revisions are made to the game to better assess its impact. The current impact evaluation literacy results show significant promise given the relatively low (22 hours) dosage.

Specific Recommendations for Children with Low Literacy Exposure

- Introduce letter names with letter sounds to better differentiate between letters, their sounds, and letters with diacritics.
- Introduce the whole alphabet before moving into letters with diacritics to establish a solid understanding of all the letters and their sounds.
- Introduce letters with their diacritics for the whole alphabet before moving into syllables to establish a solid understanding of letters with their diacritics. This will help reinforce decoding skills.
- Introduce letters shapes in beginning, middle, and end of word before moving into syllables. Provide oral instructions that clearly demonstrate how letters change and connect based on their word placement before moving into syllables. This will help support decoding.
- Build syllables letter-by-letter to reinforce letter connections, diacritics, and phonics before asking for the child to identify letters within syllables.
- Enable children to trace letters, then voweled variations and syllables early on to reinforce learning of letter shape recognition and letter connections. This could also build fine motor skills.
- Extend early game play through increased repetition of the above to establish a solid foundation of letter and letter sound recognition, letter shape and letter connections, and letters with diacritics, before moving into word building.
- Consider making use of three-letter rhyming endings of words or syllable segments when building whole words and changing only the first letter. This will reinforce blends at an early stage before moving into longer or more complex words. For example, the game could make use of simple rhyming pairs. The same concept can apply to diacritics, where the diacritic pattern within three-letter words is the same to reinforce diacritic impact on letters.
- Slow down the pace in early stages to enable early learners sufficient time to play with accuracy over speed. Speed of play can be challenged as the child advances.
- Increase oral instructions on the correct answers; allocate more game time to reinforce previously taught skills; provide corrective feedback through opportunities to self-correct; and provide oral instructions to reinforce learning and eliminate guessing from children. Demonstrate correct answers to reinforce oral instructions for visual learners.
- Introduce options for diagnostics of skills introduced that remove guessing.

Specific Recommendations for Improvement of Literacy (regardless of level)

- Consider allocating varying play-times for different tasks, since some tasks may need more time than others. For example, consider slowing down the pace of spelling stages, especially when words are longer, to enable enough time for children to complete full words.
- Consider introducing games in which the child is free to experiment with building his or her own words (even if words are incorrect or invented).
- Consider instituting an option to make words reappear during word/syllable spelling sections.
- Introduce options for self-correction.
- Introduce options for diagnostics of skills introduced that remove guessing.
- When new words are introduced, consider embedding their meaning (through images recalled in the memory of the monster, for example) in case words unfamiliar to the child, or where the colloquial word differs from the fuṣ-ha word.
- Showcase the mini-games early on to entice the child to unlock and play them regularly.
- Slow down the pacing of the game with respect to learning new skills. Consider allocating more and repetitive gaming to individual letters and completing the entire alphabet before moving into syllables.
- In line with literacy best practices, consider finding ways to engage parents in encouraging children to discuss their learning after gaming. Consider engaging parents in encouraging their children to continue practicing their reading skills. This can be accomplished through use of other digital learning platforms such as Qysas, or by revisiting Feed the Monster for reinforcement and practice.

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 experimental design (randomized controlled trial) to estimate the impacts of the EduApp4Syria games on children's psychosocial outcomes over time, comparing growth in psychosocial 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 psychosocial outcomes of children using the games will improve more than those who did not use the games.

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 psychosocial well-being outcomes to the use of the game(s), rather than other factors. The primary instrument used to assess psychosocial outcomes is the Strengths and Difficulties Questionnaire (SDQ), administered with parents of a subset of children in control and treatment groups.

SDQ: Psychosocial well-being was measured through the SDQ instrument, designed to measure children's strengths, difficulties, and pro-social behavior. It was tested and translated into Arabic by Save the Children-Jordan in camp and host community settings among their Kindergarten (KG) children in Jordan in 2013. It is a brief behavioral screening questionnaire targeted for children ages 3 to 16 and asks about 25 psychological attributes along five topics:

- 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 competition criteria by the research team. *Table 30* draws parallels between each of the five SDQ Scales and the various psychosocial criteria of the EduApp4Syria competition:

Table 30: *SDQ Scale*

EduApp4Syria Criteria	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:

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

Table 31: *Participation among parents*

SDQ (Parents)	Baseline	Endline	Attrition
Control	123	47	76
FTM Treatment	125	65	60

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's child population. All parents of children were invited to participate. Participation among parents (see *Table 31*) at baseline fell below expectations 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 (Qualitative) 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 on 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 use of and engagement in the game.

The evaluation methodology consisted of: 1) desk review of DGBL on psychosocial outcomes and engagement; 2) semi-structured focus group discussions with child beneficiaries and their parents; 3) in-depth interviews and FGDs with RI staff and volunteers; 4) observation of children using the games for gaming assessment and engagement; and 5) play-intensive interviews with child beneficiaries. Stakeholders groups and data collection is clarified in *Table 32:*

Table 32: *Evaluation Stakeholders and Data Collection to assess the impact on Psychosocial well-being*

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	■ 8 FGDs
Children in Treatment	■ 10 FGDs ■ 102 Engagement observations ■ 39 Gaming observation ■ 17 Play focused interviews
Game Designers	■ 1 KII

6.2 Limitations

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

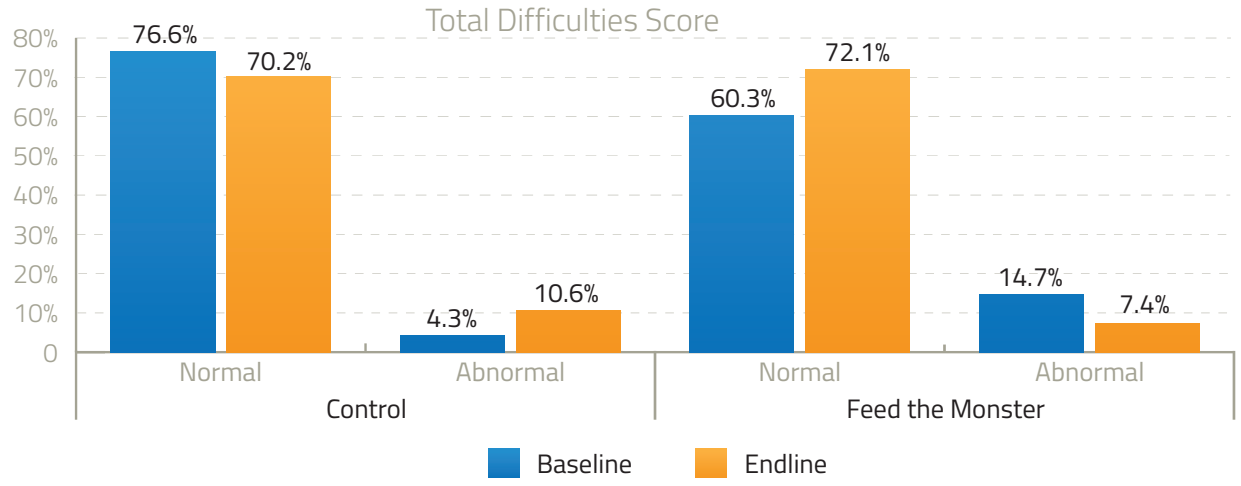
SDQ Analysis: The SDQ data was analyzed using the SDQ software.[54] Each subscale is scored from zero to 10. Each of the five subscales—emotional, conduct, hyperactivity, and peer problems—were classified as "normal, borderline, and abnormal" by way of the questionnaire analysis. The total difficulty score ranges from zero to 40 and is calculated by adding the subscale scores. The internalizing and externalizing subscales range from zero to 20. The "pre" and "post" SDQs can be used to indicate everyday practices by children. Studies use SDQ along with research interviews, and clinical ratings have shown this tool is sensitive to treatment effects. Child and adolescent mental health services as well as other services with emotional and behavioral difficulties find the SDQ useful in designing interventions to enhance the psychosocial well-being of these age groups. The analysis groups students according to the category of psycho-social assessment (e.g. Hyperactivity).[55] *Table 33* shows the results of the SDQ:

Table 33: *SDQ Results*

Overall	Baseline			Endline			Baseline			Endline		
	Control						Feed the Monster					
	Normal	Borderline	Abnormal	Normal	Borderline	Abnormal	Normal	Borderline	Abnormal	Normal	Borderline	Abnormal
	50						68					
Emotional Symptoms	91.5%	4.3%	4.3%	80.9%	4.3%	14.9%	73.5%	17.6%	8.8%	79.4%	5.9%	14.7%
Conduct Problems	80.9%	8.5%	10.6%	83.0%	6.4%	10.6%	64.7%	14.7%	20.6%	82.4%	5.9%	11.8%
Hyperactivity/ Inattention	63.8%	21.3%	14.9%	68.1%	12.8%	19.1%	55.9%	16.2%	27.9%	73.5%	5.9%	20.6%
Peer Relationship Problems	87.2%	12.8%	0%	76.6%	19.1%	4.3%	72.1%	17.6%	10.3%	88.2%	11.8%	0%
Pro-social Behavior	97.9%	2.1%	0%	91.5%	8.5%	0%	92.6%	2.9%	4.4%	94.1%	4.4%	1.5%
Total Difficulties Score	76.6%	19.1%	4.3%	70.2%	19.1%	10.6%	60.3%	25.0%	14.7%	72.1%	20.6%	7.4%

Parents of children playing 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, most children in the 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 20% between baseline (60.3%) and endline (72.1%), as compared to a decrease in normal total difficulties scores among the control group between baseline (76.6%) and endline (70.2%), as shown in *Figure 25:*

Figure 25: *Total Difficulties Score*



Within individual domains among treatment children, normal scores increased by 5.9% within emotional symptoms; 17.7% within conduct problems; 17.6% within hyperactivity; and 16.1% within peer relationship problems. Within the control group, normal scores decreased by 10.6% within emotional symptoms and peer relationship problems, while normal scores increased by 2.1% within conduct problems, and 4.3% within hyperactivity, as shown in *Figures 26 and 27:*

Figure 26: *Conduct Problems*

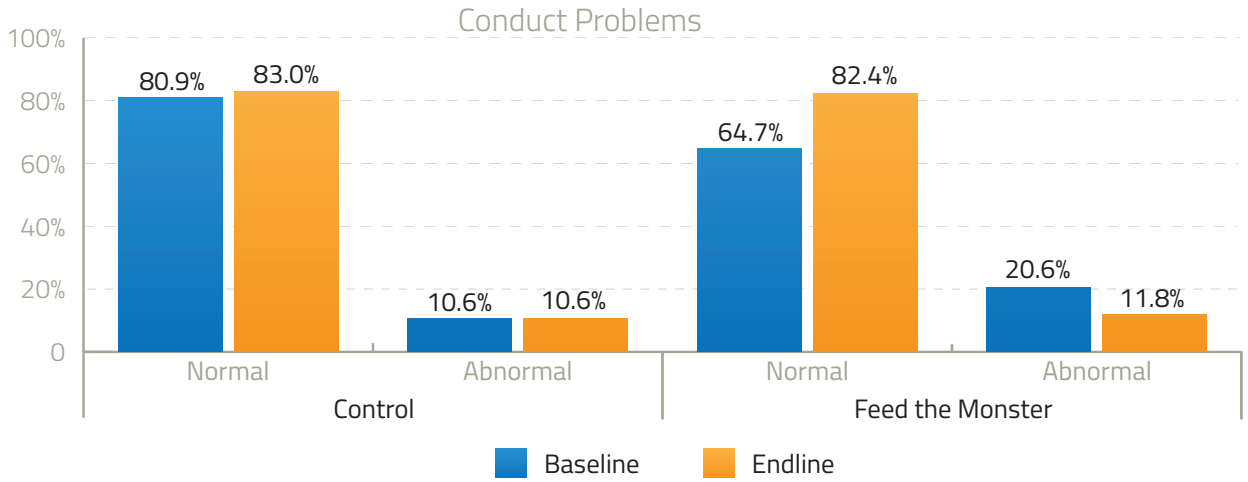
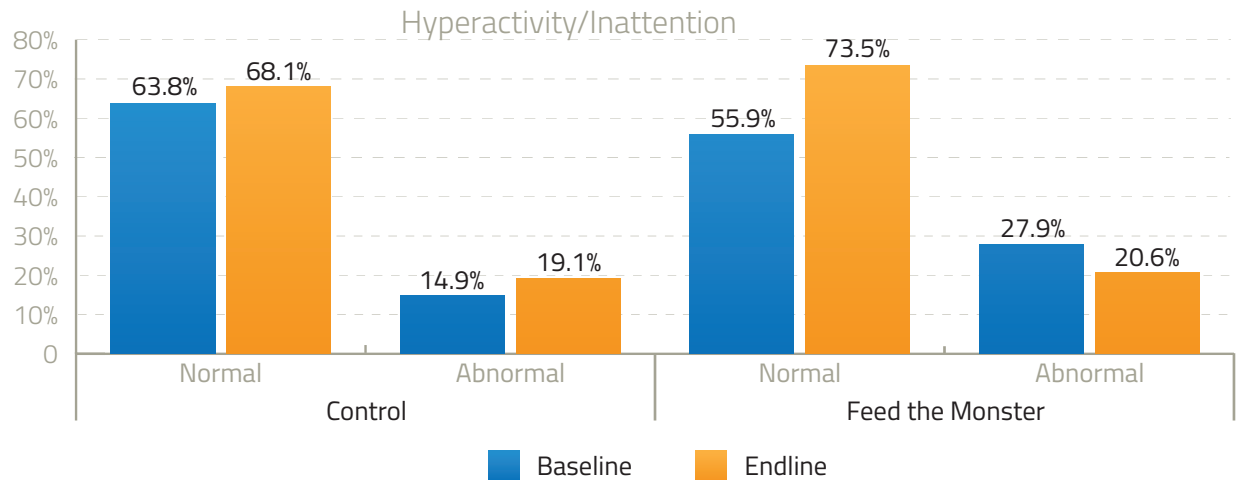


Figure 27: *Hyperactivity/Inattention*



Responses provided by treatment children and their parents in FGDs revealed children could correctly identify emotions and engaged with their peers because game play. *(Figures 28 and 29).*

Figure 28: *Peer Relationship Problems*

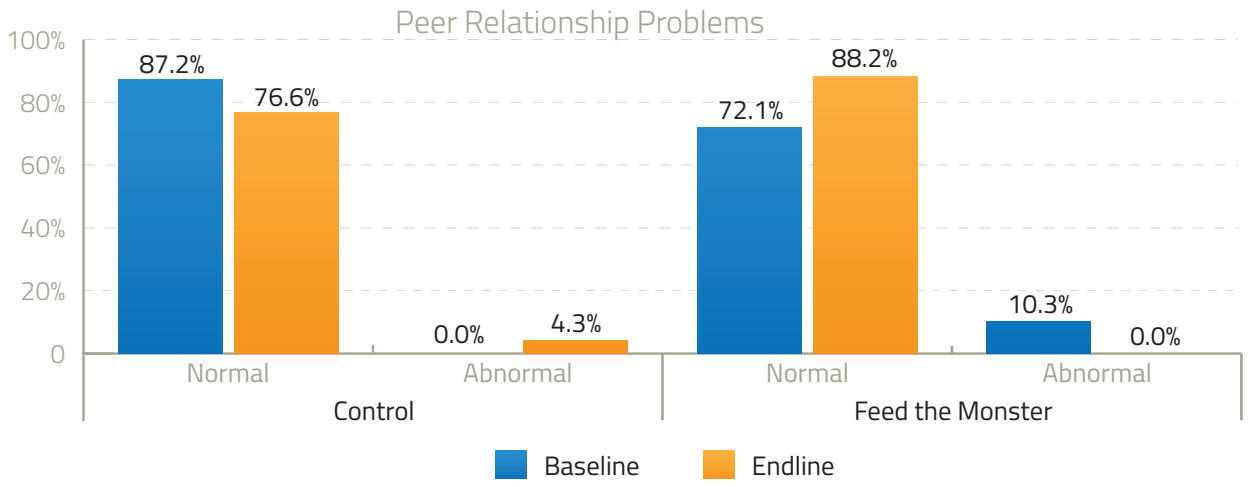


Figure 29: *Emotional Symptoms*

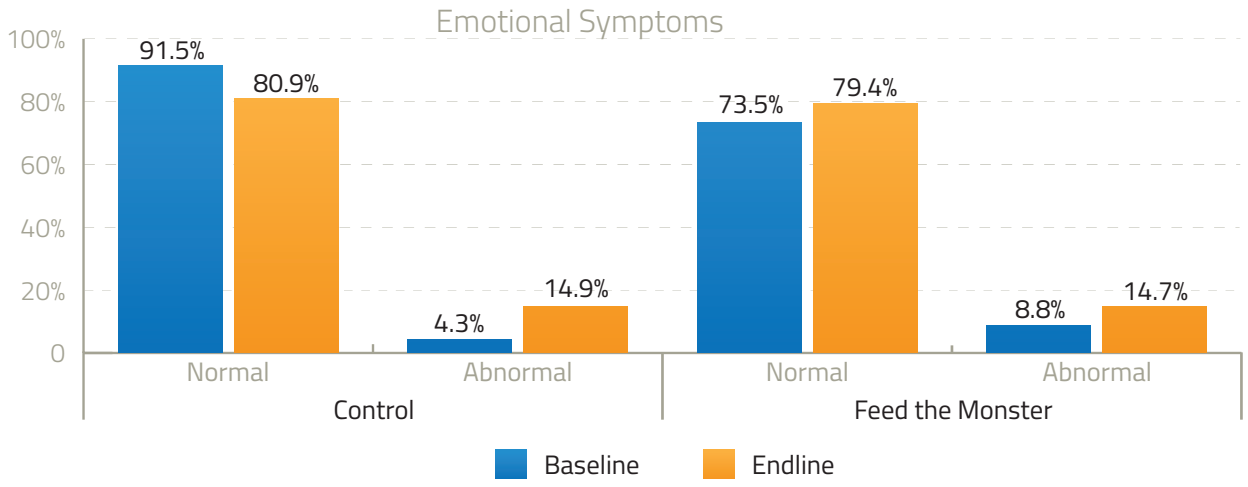


Figure 30: *A child showing her progress in FTM.*

Among treatment children, pro-social behaviors exhibited a 1.5% improvement in normal scores between baseline and endline, whereas the control group exhibited a decrease in normal scores of 6.4%. Treatment children also showed a decrease in abnormal scores, most notably in peer relationship problems (10.3% reduction), conduct problems (8.8% reduction), and hyperactivity (7.3% reduction). By contrast, the control group displayed no reduction in abnormal scores in any domain except pro-social behavior, where abnormal scores were reduced by 2.9%. This potentially signifies a positive impact on children's peer relations, conduct, and hyperactivity levels, supported by qualitative findings where volunteers and parents reported that children's motivation to learn and their focus increased during testing period.

The control group shows steady or worsening scores within all domains except pro-social behaviors and an overall decrease of normal scoring from 76.6% at baseline to 70.2% at endline. The treatment group had an overall increase in normal scoring from baseline to endline. However, underlying factors (i.e. Village 5 residents being asked to move homes during the testing period) may have influenced these outcomes. As such, these external factors did not enable equivalence for comparison of treatment to control groups. Nevertheless, the data sets among the treatment group are indicative of overall well-being improvements within four of the five domains assessed, showing promise for a positive gaming effect on psychosocial outcomes.

6.4 Qualitative Findings

6.4.1 Children's Engagement with Feed the Monster

The FGDs with children revealed that children enjoyed playing the game. All children interviewed readily recognized the character and name of the game with enthusiasm, with most children in FGDs (41 of 56) liking the game. This was confirmed by play-intensive interviews, where seven of 14 children responded to a question about their least favorite part of the game by saying there was nothing, while three more children said they enjoyed everything. Some parents indicated their children talk about the game at home in a positive manner, adding that their children are enjoying the game and progressing from one phase to another. Thirty of 34 parents in focus groups confirmed their children were connected to the game and were still engaged in playing it after six weeks of playing.

Further, some parents indicated their children were much more motivated to attend the center due to game play. A few parents mentioned their children used to skip school or lacked enthusiasm about going to school, but this changed since they started playing the game: "They love the game and they can't wait to leave the house to go to school and play it." Parents noted their children were competing with their friends to score points, and the competition among peers contributed to their sense of achievement. Some other parents noted their children had increased motivation to learn and do homework. Others

noted the game motivated children to learn their lessons, which in turn had an impact on their academic performance. A few parents noted that using games for learning improved results.

Table 34: *Children's Engagement with Game*

Experienced boredom	30/56
Completed game	30/56
Completed game more than three times	8/56

Engagement proved to be short term, however— three of 34 parents mentioned their children's enthusiasm was decreasing, and 30 of 56 children in focus groups reported boredom from the continued exposure over time (note: at about 15–17 hours of dosage). One child reported "It's boring. It makes me think, and we don't want to think. We want to play—we want something new." Another child reported, "I don't like it anymore because it's not fun, I finished it six or seven times." Multiple rounds of completion were observed for FTM, and it was evident the children played the game multiple times due to a lack of alternative activities/subsequent stages within the game or an entirely different game or activity. In engagement observations, 13% of children lost interest in playing the game after 15 minutes of play. *Table 34* above reflects quantitative information regarding sustained interest.

When asked about the children's level of engagement, teachers or volunteers stated that the level of enthusiasm dropped in the last 10 days because children were getting bored with the game. One volunteer explained that "dedicating 45 minutes each day for two months for playing the game is too much time, and the children want to do something else now." Volunteers and teachers felt playing the game once provided the needed challenge for the child; this challenge wasn't presented in the following repetitive rounds.

Despite an observed decrease in enthusiasm for some children, 14 of 17 child interviewees said they would play FTM at home, with half of these children explaining they would play because the game is fun. In keeping with this enthusiasm, 10 of 56 children reported in FGDs that they had already downloaded the game at home and were playing the game with their siblings and peers. This was corroborated in focus groups with parents. This engagement at home highlights children's connection to the game in that they wanted not only to play at school, but also engaged with the game willingly beyond the classroom.

Children reported that the positive reinforcement they received from the game encouraged them to keep playing. They also liked receiving verbal praise, such as "good job." Additionally, the children expressed their desire to maximize rewards, saying: "I don't like it when I get two stars, I want three" *(Figure 31).* This indicates positive reinforcement encouraged additional trials and sustained engagement/motivation. Another child reported that receiving the rewards encouraged him to keep playing: "The fatter the monster got, the more I was encouraged to play." The reinforcements encouraged children to play, as they kept attempting to win to unlock additional games.

However, the volunteers/teachers felt that points and rewards presented inside the game were unclear and insufficient and should be better defined. One volunteer mentioned that children didn't notice they had rewards and points that were increasing as they progressed through the game. This volunteer's observation was confirmed by play-intensive interviews, where interviewees demonstrated

Figure 31: *Three out of three stars.*

inconsistent ideas of how to earn rewards like new monsters. Among the responses children gave to explain how new monsters were acquired were the following: children must win the game, children must press letters, children do not have to do anything, or children acquire new monsters when picking which monster to use. Two children thought they would receive monsters simply by playing the game, and others did not know at all how to get them. Beyond misunderstanding how to get new monsters, only half understood how to make monsters grow.

There were also concerns that children may have difficulty playing the game on their own; however, FGDs and observations revealed that children require minimal adult supervision to play FTM. At the beginning of the playing period, children reported that they needed adults to explain game functions. As time passed, however, the need for supervision decreased. Children stated that over the course of the period, adult supervision was; they grasped the mechanisms of how to play after the first time, for some in as little as 30 minutes. In only 7% of engagement observations did the child get frustrated when using the game. Later weeks of game play had a higher flow average (a 16% increase from the first week) as recorded in gaming observations. A high flow average indicates that the game offers children an engaging level of physical and mental challenge.

6.4.2 Children's Emotional and Psychosocial Outcomes with Feed the Monster

FGDs with volunteers noted that the storyline at the beginning is appropriate and describes symbolically what happened to the Syrian people and children. They also noted Syrian children should aim for reaching the end of the game and celebrating the success. Some volunteers thought the name wasn't culturally appropriate and could be perceived as scary for children. Children themselves did not seem to identify with the story, with only one of 16 interviewees correctly describing the story narrative in an interview.

Children in focus groups reported they felt happy when they played, happy when they were praised by the game, and happy when they leveled up. At the same time, they reported also feeling frustrated and/or angry when they lost or were unable to figure out how to spell letters. They also expressed frustration when the "monster spit up the letters." On the other hand, one boy reported intentionally feeding the monster the incorrect letters because the way in which the monster spits out the letters was entertaining to him. The fact that children remembered specific elements of the game soliciting an emotional response highlights the game's ability to encourage children to understand their feelings.

Table 35: *Children's favorite FTM Features*

Feature	Number of Votes (from Child FGDs)
Stages	16
Activities	28
Character	20
Music	13
Visuals	14

Children were eager to share about their favorite aspects of the game. In play-intensive interviews, five of 17 children expressed that feeding their monsters was their favorite part of the game. In focus groups, children also repeatedly mentioned enjoying the graphics, music, and different activities, citing the "dragon game," the "worm," and the memory game as favorites (children referred to the games according to the monsters that unlocked them). *Table 35* reflects different elements of the game favored by the children.

Children also found ways to personally connect to the game: "I would like to take pictures of myself and put them on the game." The game also prompted children to develop empathy, as one child reported, "I would like to download the game at home, so my sister can play it because she's sick," reflecting that this child believes the game will make his sister feel better since it had a similar impact on his own emotional state.

Figure 32: *Children interacting with each other while playing FTM.*

When asked about perceived changes in their children's emotional and psychological well-being, parents' responses varied. Almost half of parents (20 of 38) did not perceive changes in the mood of their children. Among those who perceived a difference, the changes noted included that children became calmer and tidier and returned from school (camp) feeling happy. Some parents noticed an increased tendency to help others, a higher level of independence and interaction with others, and less conflict with peers. "My son is feeling better, he is more active. He used to stay on his own and not deal with others; now he is interacting more." Still others noted higher levels of concentration and obedience, higher inclination to play and experiment, and decreased levels of fear. One parent mentioned her child's personality grew stronger and became more expressive. "My child used to be reclusive and liked to be on her own. Now her personality grew stronger and she mingles with the neighbors and goes to the market to shop."

Regarding social skill development, responses varied. Some children reported they preferred to play on their own because they didn't want to share tablet time and wanted to learn more independently; others reported they liked playing with peers and friends, stating "I don't get bored when I play with my friends." However, in 84% of engagement observations, the child observed was engaged positively with other children while playing. When they played together, children stated they assisted one another with reading, writing, and technological errors.

Volunteers reported that children compared performance (stars acquired and levels attained), and that game competition was a significant motivating factor. In gaming observations, children were frequently observed sharing progress among peers and with adults in the classroom, as well as helping each other. In interviews, seven of nine children expressed their desire to share with their friends and compete with peers motivated them to share their progress with other children. Children mentioned to their parents that the game is about combining letters into words and competing with their friends to score points, which contributes to their sense of achievement. These findings have important implications for social, communication, and problem-solving skills. Parents noted that children came home to "play teacher" with younger siblings, and that they had improved relations with neighborhood children.

Some technical difficulties occurred during game play, but overall, no frustration was observed. There were 30 total observed occurrences of mechanical difficulties, and 15 of these occurrences were glitches while dragging letters to feed the monster. While difficulties with dragging did not interrupt the children's play extensively during observations, more play testing is needed to better determine the effects of the drag mechanic on frustration.

Among those who perceived a difference, the changes noted included that children became calmer and tidier and returned from school (camp) feeling happy.

The children reported in FGDs that tablets would shut off sporadically and that in those moments they did not know what to do. Fourteen of 56 children reported glitches, mostly related to the game freezing. Eight such glitches were observed during gaming observations; five times when the screen was loading between game rounds; and three times during other periods in the game. Children were most often observed asking for help from an adult in the classroom when this occurred. The issue may be due to overheating, since devices with glitches tended to be running "hot." Errors were usually fixed

by restarting the game. Overall, technical glitches did not appear to trigger frustration considering the very low average level of 6% for child frustration from gaming observations.

Some differences were observed between male and female children. More female children had technical issues in the waiting period between levels (six of 23 females versus three of 16 males) that could detract from their enjoyment of the game. Females were more engaged in choosing between their monsters than male children, with nine of 23 female children observed as engaged in this process and only three of 16 males demonstrating the same engagement. Females also understood better how to make the monsters grow, with zero of seven female children misunderstanding that changing between monsters is what makes monsters grow, as opposed to three of eight males. There were no significant differences for gender regarding flow, game engagement, or frustration.

Differences were observed between younger and older children, with older children demonstrating a better understanding of game rules and mechanics. Many older children (five of eight) understood that feeding the monsters is what makes them grow, whereas none of the six younger children interviewed could identify what makes monsters grow. During the 39 gaming observations conducted for FTM, older children experienced 9% more flow and 5.6% more mechanics engagement than younger children. This suggests older children experienced slightly more flow and mechanics engagement than younger children.

6.5 Conclusions

Figure 33: *In one game, children could pet sad monsters to make them happy.*

The game is well designed in its simplicity and polish, which enabled children of all ages and genders to easily and readily play with minimal or no adult supervision or interface. Children were easily able to navigate and complete the game, which gave them a sense of both control and achievement. They were also easily able to collect, hatch, and grow monsters, which encouraged them to continue playing. Interaction with the monsters—both with correct and incorrect answers—was entertaining to all children, and most children liked the game. Children reacted very positively to the graphics, music, and activities of the game, and most engaged with the monsters and gaming activities (such as the dragon, worm, and memory games). Parents and teachers found the game appropriate, and generally viewed its impact on their children as positive.

The game clearly communicated progress both through rewards and incentives such as the growth and collection of monsters, and differentiated performance through the number of stars collected at each stage. These incentives encouraged game play and offered positive reinforcement, which in turn motivated their continued engagement, up to a certain point. However, some elements of the game were unclear to the child. Game mechanics were consistent regardless of task: tasks were timed similarly and the rewards and monster reactions were mostly unvaried, leading to a sense of repetition that may have reduced the level of sustained engagement.

Also, pacing did not cater to the needs of each child. As a result, engagement was relatively short-lived; once monsters were collected and three stars on each level were achieved, the game did not offer any further challenge to the children, nor did it evolve with them. There is a need to address sustained engagement, as boredom, confusion, frustration, and anger can lead to decreased motivation and disengagement from the task.

At a psychosocial level, the game appears to have supported the development of positive social outcomes, as demonstrated by both SDQ results and parental feedback. Children wanted to share the game and their learning with both siblings and peers, demonstrating impact on emotional states and social behaviors. Peer interaction increased because of game play, which had positive implications for social and communication skills. Moreover, gaming engendered high levels of motivation to attend the summer camp, engage with peers in competition and assistance, and complete the game. This had impact on children's motivation to learn.

6.6 Recommendations

While the game's opening storyline is powerful and age-appropriate, developers could make better links to the storyline throughout the game. As the game introduces new items (such as eggs, gifts, and points), integrating them into the storyline would likely generate increased interest and engagement. A positive conclusion to the storyline with all monsters may also generate more attachment to the game.

For all children, the game could benefit from increased personalization—including improved personal engagement in off-task play areas as well as avatar personalization. This would likely result in increased engagement and challenge, resulting in sustained interest in game play.

To sustain engagement for learning, the game could benefit from improved pacing based on individual ability, introducing challenges for more advanced children, and more time on task for early learners. This could reduce frustration and boredom–negative emotions that could diminish time on task or desire to continue game play.

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 engagement and time on task, as well as increase pro-social behaviors and psychosocial outcomes. This could be accomplished through the development of sharing mechanisms to support peer interaction or friendly competition. Sharing gaming through group game play—as practiced in classrooms in the testing of this game—might also maximize social interaction on the joint goal of game completion.

Specific Psychosocial and Engagement Recommendations

- Consider creating an appropriate ending for the grown monsters that ties to the original storyline for more impact and closure. Consider better integrating the discovery of new eggs to the storyline or to literacy achievements so the child is actively seeking to unlock new eggs and collect monsters. Consider better integrating the scoring to objectives in game play.

- Consider developing more options for child engagement with the game. For example, consider creating a space for naming the monsters or engaging with the monsters in non-learning play to increase levels of engagement. Explore ways in which this personal space can further engage the child in character interaction. This might compel the child to come back to play more.
- Consider enabling the personalization of the monsters by enabling the child to name the monster. This could reinforce letter-sound recognition and spelling. Consider developing a name plate for each monster for the child to personalize.
- Consider augmenting 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 and can increase engagement.
- Consider evolving the game with the monsters. For example, consider accompanying each newly-acquired monster with a more difficult or challenging level. In this way, collecting monsters is not simply an iterative practice of the same game, but increases the challenge in the game.
- Consider including different emotional expressions for the character's reactions. This would expose children to new emotional vocabulary.

Recommendations for Future Iterations (currently outside EduApp4Syria specifications)

- Consider adding a "sharing" platform to promote peer social interaction through friendship building, communication, and social skill development among children, a critical component of psychosocial development. Since peer interaction was a significant success in testing, it could effectively generate increased engagement. In lieu of an online space, this could also be explored in group settings.
- Consider developing a track feature that would remind children to take on activities or stages that focus on their current challenges/stage. These reminders should include a motivational aspect to help encourage trials.
- Consider enhancement of "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. Stories could also target a deeper level of psychosocial well-being by including those that depict feelings as well as problem solving.
- While intended for independent use by out-of-school children, the game–if used in both formal and informal education settings–could increase the connection/link between teacher and student. The inclusion of a feature that could be manipulated by teachers and allow children to prepare for lessons in advance would give children more confidence to participate in class. Such student/teacher interaction may also promote positive social interaction between children and adults.

Photo Credit: Norad/Knut Hverven

7. Areas for Further Research

The high rate of mobile phone subscription and smartphone penetration offers opportunity for remarkable reach of DGBL into the MENA population.

This impact evaluation explores the potential for DGBL to improve literacy and psychosocial outcomes: an area of research that is relatively nascent. The novelty of the EduApp4Syria competition as a process—as well as the ensuing products and their potential to address literacy and psychosocial outcomes through DGBL—offer opportunities to further explore methods of education delivery, particularly for out-of-school children. In a region 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[56]–there is a clear need to find innovative methods of education delivery that out-of-school children can access. The high rate of mobile phone subscription and smartphone penetration offers opportunity 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.[57] 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.[58]

The EduApp4Syria competition was an innovative approach taking advantage of the digital revolution while addressing the challenge of education delivery. DGBL's disruptive approach to learning could address some of the challenges faced by educators across the MENA region. Beyond the MENA region, digital game-based learning offers the potential to reduce barriers in accessing learning materials, alleviate the need for physical learning environments, offer the opportunity for iterative and differentiated learning, and address quality of education delivery. Serious games or simulations with clear "learning outcomes can create interactive experiences that actively engage the children 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]

Although game developers in the region are producing Arabic learning games, there are very few such learning games in Arabic that have undergone rigorous testing to offer discussion on the role of DGBL in education. While this report makes one contribution to DGBL's impact on literacy and psychosocial outcomes, there is a need for further research to better understand the potential of DGBL on improving literacy outcomes and psychosocial well-being, particularly among populations in conflict. Some of the salient issues raised within this report 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 22 hours for Feed the Monster generated some positive results and improvements across all age groups and genders. This leaves much room for discussion on possibilities of increased impact with increased dosage, or alternatively increased impact with gaming improvements. Moreover, games are intended to engage children for a prolonged

period and leave them with a sense of accomplishment of completing or winning the game. As such, it may prove challenging to engage in game play for the recommended 60 hours, as witnessed with the three-hour game play time in this current version of FTM. Since DGBL does not correspond to traditional literacy instructional intervention, further research is needed to discover how much DGBL dosage would be required to generate impact. Dosage over the course of a series of games or digital platforms that build on each other is another area requiring further research. Qysas and other early grade digital reading platforms, for example, could be used as follow-on material after children finish simpler literacy games.

Differentiated or 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. In this way, 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, though this may not conform to game play times or dosage requirements. Enabling DGBL to self-adjust according to child performance may lead to improved and differentiated progress through the game. This may allow more time to explore differentiated or self-paced DGBL designs and impact. 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. As such, some parents in focus groups indicated that different behaviors emerged in terms of parental interaction with the child and child interaction with siblings in the test setting. 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.

Blended Learning: While not in the scope of this study or within the aims of the EduApp4Syria competition, approaches to blended learning or integrating DGBL into the classroom or curricula could be an area for further research. Feed the Monster provided proof of concept that children with previous literacy exposure tended to perform better, pointing to the possibility that DGBL—when combined with more formal instruction—could lead to greater literacy gains. Research could also examine the role of DGBL within national curricula and the classroom. As noted by the landscape review of Education in Crisis (2017), understanding these political requirements "to help ensure content is appropriately and responsibly delivered, is important." The literature recommends, where possible, to source content locally, and notes "working with governments and formal curriculum where possible can support long-term education initiatives".[60] On another level, integration within the class-

In many of the cited regional examples of DGBL use, teacher training on the use of and motivation with DGBL becomes a critical component of its delivery, application, and use. Worldwide studies confirm the importance of effective implementation by teachers in realizing the full potential of DGBL.

room requires that teachers be trained and/or experienced in DGBL game play and instruction. In many of the cited regional examples of DGBL use[61], teacher training on the use of and motivation with DGBL becomes a critical component of its delivery, application, and use. Worldwide studies confirm the importance of effective implementation by teachers in realizing the full potential of DGBL. The EduApp4Syria Impact Evaluation was delivered in a remedial classroom setting, an area ripe for further learning and testing of new methods.

Illiterate Parents: In some instances, parents of children testing Feed the Monster indicated that they had engaged in game play with their children. Some of these parents were illiterate; therefore, 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, since 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. Further research could explore ways in which social aspects of gaming—such as group dynamics, peer/class competition, or peer sharing/teaching—can impact psychosocial outcomes and learning. There is a need for better understanding of the impact that DGBL solutions can offer, particularly as DGBL content developed in Arabic holds a potential market of over 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% growth in the 2017–2020 period—the fourth highest region out of seven global regions assessed with the 2017–2022 Global Game-based Learning Market.[62] 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.

Further research could explore ways in which social aspects of gaming—such as group dynamics, peer/class competition, or peer sharing/teaching—can impact psychosocial outcomes and learning.

Annex 1: Literature Review

Relevance of DGBL to Syrian Refugee Educational Context

In the Levant region, perhaps the most pressing education challenge has been 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.[63] 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,[64] 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."[65]

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."[66] Further, a recent global meta-analysis of 77 randomized experiments evaluated the effects of school-based interventions on learning in developing country primary schools and found the largest mean effect sizes (among school-based interventions) included treatments with computers or instructional technology (0.15)."[67]

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[68] in 2016, accounting for 60% of the population.[69] 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.[70] In some markets, as many at 75–85% of people have smart-

phones, and mobile penetration in many markets exceeds 100%, though these levels can vary substantially for refugee communities.[71] 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.[72] According to a REACH report conducted in Jordan, mobile phone ownership among encamped refugee households was widespread, and smartphone penetration reached 58%.[73] 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.[74] 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, however, 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 that 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 in-depth materials to review.

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, reading comprehension skills, and simultaneously provide 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)[75] and Egypt[76] (Research Triangle Institution [RTI]) 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.[77] 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.[78] 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 applications 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, he or she can touch the word to activate the audio function; 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.

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 user names. This data can be linked to a teacher dashboard that aggregates student performance, so the teacher is able to monitor student progress. Individual student performance is tracked, offering the teacher a student-specific or class-wide dashboard of reading performance. Back-end data collection served the learning agenda.

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,[79] 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 school children ages 5–13 still out of school,[80] a consortium of stakeholders including the Sudanese Ministry of Education, Ahfad University for Women, War Child Holland, and Netherlands Organisation 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 high primary school enrollment rates, gross enrollment rates range from 85% in Al Gezira to 37% in East Darfur.[81]

Sudan still faces significant infrastructure challenges around school resourcing, poor learning outcomes, 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."[82] 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."[83]

ELS consists of two game worlds and various mini-games to practice each mathematics concept through autonomous learning.[84] It incorporates two distinct levels: the first level is that of the 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, for example, 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. Children work at their own pace, and progress is based on their ability to successfully complete games and tasks.[85]

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.[86] 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., No significant differences for gender were found.[87] Self-esteem, recorded on a four-point Likert scale, was reported to have increased from 1.9 to 2.5 from baseline to endline.[88]

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 Syria 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 the 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.[89] Meanwhile in Lebanon, UNICEF has been 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.[90]

Contextualization of DGBL and Similar Solutions

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,[91] 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: "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."[92] Some adaptations to contexts within the examples of ELS and the Qysas platform cited above were made to enable operation of the app offline.

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."[93] 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."[94] 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."[95] 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 form of recognition to motivate teachers to apply it. This is confirmed by other world-wide studies that document the importance of effective implementation by teachers in realizing the full potential of DGBL.[96]

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."[97] 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.[98] 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."[99] 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 may be 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 learning content, which may engender more positive feelings towards the gaming environment. Removal of off-task, in-game features may decrease positive outcomes such as engagement.[100]

Assessing DGBL Games

For games to be effective learning tools, specific conditions must be met.[101] 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.[102] 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.[103]

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.[104] When the design of these learning mechanics is not aligned with the learning objectives of the game, then 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.[105]

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. In close collaboration with Relief International (RI)—a humanitarian relief agency providing remedial education services to Syrian children offering 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 that are newer arrivals to Jordan, and thus have lower levels of schooling, meeting the characteristics required for testing. 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 Feed the Monster [FTM] 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:

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
FTM Treatment	1.4%	20.4%	23.2%	21.1%	17.3%	16.5%	284
Endline	**5 years**	**6 years**	**7 years**	**8 years**	**9 years**	**10 years**	**TOTAL**
Control	1%	21%	19%	22%	19%	19%	200
FTM Treatment	2%	21%	23%	20%	19%	15%	199

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 to parents of the children in the treatment or control groups.

Participation among parents at baseline was lower than targeted (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 during baseline, and lack of desire to be surveyed.

	Baseline	Endline	Attrition
SDQ (parents)			
Control	123	47	76
FTM Treatment	125	65	60
EGRA (children)			
Control	383	200	183
FTM Treatment	284	199	85

Contamination

Contamination	
FTM Treatment	2 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 non-test subjects or control subjects downloading the apps. In Azraq, control and treatment groups were also separated by administrative villages, with Village 5 remaining physically separated from the other villages (fenced in 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. Nevertheless, it was discovered that some children/parents had downloaded the Antura and the Letters app (the other EduApp4Syria game). 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. Gender differentials for children were also considered, 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 app 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 among users (children). One hundred observations were conducted throughout the testing period to assess ease of use and game engagement.

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 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- to 7-year-old male user.
- **Younger Female Player.** This is a 5- to 7-year-old female user.
- **Older Male Player.** This is an 8- to 10-year-old male user.
- **Older Female Player.** This is an 8- to 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 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 data collectors with EGRA experience. They were selected from a pool of 70 trained and certified data collectors who met the 90% inter-rater reliability (IRR) gold standard, and worked on either the ACR—Little Thinking Minds EGRA-based impact evaluation or the USAID 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. This ensured strict adherence to data collection protocols, especially those 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 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, agreed 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 sessions must be considered.[106] Findings generally supported that more intensive interventions produced higher outcomes[107] but cautioned that more frequent or intense doses may not lead to better outcomes, particularly if the quality of the intervention is subpar.[108]

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 completion of the game itself. In considering the latter, dosage may be increased by adding or changing the levels and activities of the game, recognizing that dosage beyond completion of the game could have negative effects on engagement. However, adding levels to the game to increase dosage may make the game too large for easy storage or download, or have negative impacts on the game's simplicity. A follow-up study may glean more insights on ideal dosage in a household setting as opposed to classroom setting.

A survey of similar interventions shows an average implementation period of two months, with most falling within the range of 20 to 30 hours.[109] Studies with a lower dose (10 hours or less) over the same period did not have effective results.[110] 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, which are further explored in the risks section.

Annex 3: Baseline Child Demographic Survey Data Analysis Tables

Gender					
	Male	Female	Total	Male %	Female %
FTM	143	141	284	50%	50%
Control	142	141	283	50%	50%
Grand Total	**285**	**282**	**567**	**100%**	**100%**

Age													
	5 Years	6 Years	7 Years	8 Years	9 Years	10 Years	**Total**	5 Years	6 Years	7 Years	8 Years	9 Years	10 Years
FTM	4	55	62	73	50	40	**284**	1%	19%	22%	26%	18%	14%
Control	4	56	52	61	55	55	**283**	1%	20%	18%	22%	19%	19%
Grand Total	**8**	**111**	**114**	**134**	**105**	**95**	**567**						

When was the last time you attended school?				
	Feed the Monster	**Control**	**Feed the Monster %**	**Control %**
2007	0	1		
2013	0	0		
2014	0	1		
2015	9	8	3%	3%
2016	56	27	20%	10%
2017	218	224	77%	79%
I did not go to school	1	22		8%
Total	**284**	**283**	**100%**	**100%**

What was the last grade you attended in school?				
	Feed the Monster	**Control**	**Feed the Monster %**	**Control %**
1st Grade	127	160	45%	61.3%
2nd Grade	95	70	34%	26.8%
3rd Grade	56	30	20%	11.5%
4th Grade	5	1	1%	0.38%
5th Grade	0	0		
Total	**283**	**261**		

In which country did you attend school the last time?				
	Feed the Monster	**Control**	**Feed the Monster %**	**Control %**
Jordan	269	219	95.1%	84%
Syria	14	42	4.9%	16%
Total	**283**	**261**		

Who are you living with?				
	Feed the Monster	Control	Feed the Monster %	Control %
Parents	194	102	68.3%	36.04%
Father	65	169	22.8%	59.72%
Mother	22	11	7.75%	3.89%
Brothers	2	0	0.7%	
Grandmother	0	0		
Grandfather	0	1		0.35%
Uncle	1	0	0.4%	
Total	**284**	**283**		

Do you live with your parents?				
	Feed the Monster	Control	Feed the Monster %	Control %
Yes	270	277	95%	98%
No	14	6	5%	2%
Total	**284**	**283**		

Do you have a smartphone?				
	Feed the Monster	Control	Feed the Monster %	Control %
Yes	225	237	79%	84%
No	59	46	21%	16%
Total	**284**	**283**		

Are you able to use the smartphone?				
	Feed the Monster	Control	Feed the Monster %	Control %
Yes	169	184	75%	78%
No	56	53	25%	22%
Total	**225**	**237**		

Do you know how and where your parents charge their phones?				
	Feed the Monster	Control	Feed the Monster %	Control %
Yes	223	228	99%	96%
No	2	9	1%	4%
Total	**225**	**237**		

Can your parents charge their smartphone easily?				
	Feed the Monster	Control	Feed the Monster %	Control %
Yes	220	219	99%	96%
No	3	9	1%	4%
Total	**223**	**228**		

Can your parents charge their smartphone at any time?				
	Feed the Monster	Control	Feed the Monster %	Control %
Yes	219	220	98%	96%
No	4	8	2%	4%
Total	**223**	**228**		

Do you have books at home?				
	Feed the Monster	**Control**	**Feed the Monster %**	**Control %**
Yes	127	156	45%	55%
No	157	127	55%	45%
Total	**284**	**283**		

Do you read at home?				
	Feed the Monster	**Control**	**Feed the Monster %**	**Control %**
Yes	96	133	76%	85%
No	31	23	24%	15%
Total	**127**	**156**		

How often do you read?				
	Feed the Monster	**Control**	**Feed the Monster %**	**Control %**
Daily	41	75	43%	56%
Once a Week	25	22	26%	17%
Twice a Week	30	36	31%	27%
Total	**96**	**133**		

Do your parents or siblings read books with you?				
	Feed the Monster	**Control**	**Feed the Monster %**	**Control %**
Yes	76	120	60%	77%
No	51	36	40%	23%
Total	**127**	**156**		

Can you read alone?				
	Feed the Monster	**Control**	**Feed the Monster %**	**Control %**
Yes	78	100	61%	64%
No	49	56	39%	36%
Total	**127**	**156**		

Do you like coming to class?				
	Feed the Monster	**Control**	**Feed the Monster %**	**Control %**
Yes	282	281	99%	99%
No	2	2	1%	1%
Total	**284**	**283**		

Why do you like coming to class? (choose all that apply)				
	Feed the Monster	**Control**	**Feed the Monster %**	**Control %**
I just like it	40	26	14%	9%
Teachers	61	63	21%	22%
Students	26	14	9%	5%
Games	48	35	17%	12%
Learning	**224**	**234**	**79%**	**83%**

Annex 4: Baseline Parent Demographic Survey Data Analysis Tables

Gender		
	Feed the Monster	**Control**
Female	52%	44%
Male	48%	56%

Are you the head of your household?		
	Feed the Monster	**Control**
No	16%	31%
Yes	84%	69%

What is your educational status?		
	Feed the Monster	**Control**
Bachelor	12%	8%
High School	8%	20%
Higher Education		2%
Illiterate	27%	27%
Literate but incomplete high school	52%	42%
Vocational Training	2%	3%

Do you have a smartphone?		
	Feed the Monster	**Control**
No	28%	14%
Yes	72%	86%

Are you aware of the learning app on your phone?		
	Feed the Monster	**Control**
No	41%	43%
Yes	59%	57%

Are you able to use a smartphone?		
	Feed the Monster	**Control**
No	35%	37%
Yes	65%	63%

Do you let your children use your smartphone?		
	Feed the Monster	**Control**
No	23%	13%
Yes	77%	87%

Do you have access to charge your phone?		
	Feed the Monster	**Control**
No	0%	2%
Yes	100%	98%

Can you charge your phone easily and anytime?		
	Feed the Monster	**Control**
No	1%	20%
Yes	99%	80%

Do you have books at home?		
	Feed the Monster	**Control**
No	77%	62%
Yes	23%	38%

Do you read with your child?		
	Feed the Monster	**Control**
Never	7%	2.1%
No	29%	57.5%
Yes	64%	40.4%

How often do you read with your child?		
	Feed the Monster	**Control**
Daily	33%	42%
Once a Week	33%	11%
Twice a Week	33%	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 on the Attendance Sheet for each session*	16.3
2	**Class Duration** *Record minutes*	45
3	**Over-age students** *Record number: Number of students who are older than 10 years*	0.8
4	**Under-age students** *Record number: Number of students who are younger than 5 years*	0

	Level	Average # per classroom
5	**Number of students who finished the game** *Record number*	9.8
6	**Number of students who played the game more than two times** *Record number*	3
7	**Number of students who played the game more than five times** *Record number*	0.3
8	**Number of students who played the game more than six times** *Record number*	0.2
9	**Number of students who played the game and got bored after finalizing it for the first time** *Record number*	2.4

	Technology	# of instances overall	
		YES	NO
10	**Tablet 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**	6	4

Quality of Delivery		# of instances overall	
		YES	NO
14	Volunteers trained	10	0
15	Assistants trained	5	5
16	Volunteers support students: Volunteers attend to students who need technological help or 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	2	8
19	Volunteer is the same as the one initially trained	10	0
20	Volunteer adheres to the project as planned	10	0

Child Participation		Average score per classroom (scale of 5)
21	Child engagement	3.6
22	Child interest	3.7
23	Request for help (child)	2.1
24	Interest in character	3.7
25	Interest in music	3.7
26	Interest in visuals	4
27	Interest in the levels	3.8
28	Interest in activities	3.6

Observation Summary

No.	Indicator	# YES	% YES	# NO	% NO	# NA	% NA
29	Child is engaged in app (eye contact is on app, is playing with app, etc.)	98	96%	4	4%	0	0%
30	Child knows how to navigate (moves from stage to stage easily, finds different links, etc.)	100	98%	2	2%	0	0%
31	Child loses place on app frequently (more than three times)	1	1%	101	99%	0	0%
32	Child is smiling when playing (observed once)	66	65%	35	34%	1	1%
33	Child gets frustrated when using app (observed once)	7	7%	94	92%	1	1%
34	Child exhibits an understanding of the app	101	99%	0	0%	1	1%
35	Child likes receiving positive feedback from the app (smiles, calls out to volunteer for acknowledgment)	73	72%	28	27%	1	1%
36	Child loses interest after using app for 15 minutes	13	13%	84	82%	5	5%
37	Child asks for help using app	16	16%	86	84%	0	0%
38	Child listens to the story on the app	45	44%	48	47%	9	9%
39	Child laughs using the app	22	22%	80	78%	0	0%
40	Child likes the character of the app	91	89%	9	9%	2	2%
41	Child engages positively with other children using the app	86	84%	13	13%	3	3%
42	Child can follow verbal instructions of app easily	97	95%	3	3%	2	2%
43	Child can follow written words of app easily	78	76%	14	14%	10	10%
44	Child answers 50% or more of questions correctly	95	93%	5	5%	2	2%
45	Child knows when to click on button for next stage of app	100	98%	2	2%	0	0%
46	Child can hold tablet appropriately	101	99%	1	1%	0	0%
47	Child's fine motor skills assist in his/her playing the game	101	99%	1	1%	0	0%
48	Child is gaining a skill (moved levels, started to recognize a letter, etc.)	97	95%	5	5%	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.

9 Compas, B. (2006). Psychobiological processes of stress and coping: Implications for resilience in childhood and adolescence. Annals, NY: AcadSciences. Juster, R. P., McEwen, B. S., & Lupien, S. J. (2010). Allostatic load biomarkers of chronic stress and impact on health and cognition. *Neuroscience and Biobehavioral Reviews, 35*(1).

10 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.

11 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

12 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

13 Ibid.

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

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

16 Ibid.

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

18 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 SIGCHI Conference on Human Factors in Computing Systems*, p. 1059-1068.

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

20 Hamari, J., et al. "Challenging games help students learn: An empirical study on engagement, flow, and immersion in game-based learning." *Computers in Human Behavior, 54*, p. 170-179.

21 Ibid.

22 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.

23 Ibid., p. 12.

24 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.

25 Ibid.

26 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.

27 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.

28 Ibid., p. 18.

29 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 Erlabaum Associates.

30 Maamouri, Mohammed. "Arabic Literacy." University of Pennsylvania. 1999. https://www.ldc.upenn.edu/sites/www.ldc.upenn.edu/files/arabic-literacy.pdf. p. 11.

31 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.

32 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.

33 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.

34 Maamouri, "Arabic Literacy," p. 5.

35 Ibid., 6.

36 Ibid., 7.

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 Ibid.

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 Feed the Monster game developers.

46 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.

47 Rate of Change = (Endline-Baseline)/Baseline

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

49 Hogan, T. P., Bridges, M. S., Justice, L. M., & Cain, K. (2011). Increasing higher level language skills to improve reading comprehension. *Focus on Exceptional Children*, *44*(3), pp. 1-19.

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

51 Shanahan on Literacy. (2017). "Biography." Retrieved from http://shanahanonliteracy.com/biography

52 Shanahan on Literacy. (2009). Why it sometimes looks like teaching does harm [Blog post]. Retrieved from http://www.shanahanonliteracy.com/blog/why-it-sometimes-looks-like-teaching-does-harm#sthash

53 Ibid.

54 Available online at: http://www.sdqscore.org/Amber

55 Available online at: http://www.sdqinfo.org/d0.html

56 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.

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

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

61 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*

62 Adkins, S.S. (2017). "The 2017-2022 Global Game-based Learning Market." Serious Play Conference: Metaari Advanced Learning Technology Research.

63 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

64 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

65 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

66 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/

67 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

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

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

70 Ibid.

71 Kuruvilla, et al., quoted in SAIS. "A War for Information: Use of Mobile and OTT Services among Syrian Refugees." SAIS Perspectives. 2017. http://www.saisperspectives.com/2017issue/2017/5/30/rfvuesw9cw5owfgm5st5asa-1fh2302.

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

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

74 Wall, M., Campbell M.O., and Janbek, D. (2015). "Syrian refugees and information precarity." *New Media and Society, 19*(2), pp. 6-7, 10.

75 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

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

77 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.

78 Maamouri, Mohammed. "Arabic Literacy." University of Pennsylvania. 1999. https://www.ldc.upenn.edu/sites/www.ldc.upenn.edu/files/arabic-literacy.pdf. pp. 4-5.

79 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

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

81 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-childrenmiddle-east-and-north-africa-2015-en.pdf.

82 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

83 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

84 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

85 Ibid., 26.

86 Ibid., 34.

87 Ibid., 47.

88 Ibid., 54.

89 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.

90 Ibid.

91 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

92 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

93 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. p. 12.

94 Ibid., 27.

95 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

96 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.

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

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

99 Ibid., 45.

100 Ibid.

101 Mayer, R.E. "What Should Be the Role of Computer Games in Education?" 2016. *Policy Insights from the Behavioral and Brain Sciences 3*(1), pp. 20-26.; Plass, J.L., Homer, B.D., & Kinzer, C.K. (2015). "Foundations of Game-Based Learning." *Educational Psychologist, 50*(4).

102 Mayer, R.E. "What Should Be the Role of Computer Games in Education?" 2016. *Policy Insights from the Behavioral and Brain Sciences 3*(1)

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

104 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.

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

106 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.

107 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.

108 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.

109 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 Education 40* (1).

War Child Holland. (2016). "Research report: E-Learning Sudan Phases I and II."

110 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.*

Tsai, F. H., K. Yu, K.C., & Hsiao, H.S. (2012). "Exploring the factors influencing learning effectiveness in digital game-based learning." *Educational Technology and Society, 15*(3).


ALL
CHILDREN
READING:
A GRAND CHALLENGE
FOR DEVELOPMENT


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