http://doi.org/10.35847/DWoods.DGifford.PYovanoff.ASandoval.5.3.22

Research Article

The Effects of an Adult Literacy App on Word Decoding

Dawn M. Woods, Oakland University
Diane B. Gifford. Southern Methodist University
Paul Yovanoff, Southern Methodist University
Ashley Sandoval, Southern Methodist University

Abstract

In the United States, 16.6 million adults are unable to comprehend sentences or read common documents, thereby limiting their ability to participate in a literate society. Given the consequences of low literacy, we conducted a single-case study with adult multilingual learners to test the effectiveness of a game app, Codex: The Lost Words of Atlantis, on word decoding. The analyses of the pilot study showed moderate to large effect sizes on phonological decoding of taught and untaught words and large and statistically significant increases from pre-test to post-test on five distal reading measures. These results suggest the app may support adult multilingual learners in learning to decode common English words.

Keywords: adult literacy, multilingual learners, word decoding, gameplay, single-case research design

The relationship between English literacy in the United States and positive social outcomes including better health, increased educational and employment opportunities, and higher wages is well established (Rothwell, 2020). Yet,16.6 million individuals ages 16- to 65-years-old are identified as below Literacy Level 1 by the Program for the International Assessment of Adult Competencies (Mamedova & Pawlowski, 2019). Adults at or below Literacy Level 1 often do not have basic literacy skills supporting them to read short sentences or texts, complete common forms, or identify or make inferences from information (National Center for Education Statistics, 2017). Furthermore, within this level, individuals who are born outside of the United States are represented twice as frequently as adults born in the United States.

Adults at or below Literacy Level 1 often depend on adult literacy programs to increase their reading ability. However, access to adult literacy programs is already limited, and if available, are often at capacity (ProLiteracy, 2022). For example, in Texas, a state ranked 44 out of 50 for adult literacy (National Center for Education Statistics, 2017) and a state with a large population of residents born outside of the United States and identified as English Limited Proficient (54%) (Migration Policy Institute, 2019), 4.8 million adults need adult education programs (Texas Workforce Investment Council, 2018). Further, the Texas Workforce Investment Council (2018) estimates that this need will increase by 80% from 2020 to 2050.

To support individuals at or below Literacy Level 1

Correspondence: dawnwoods@oakland.edu

to gain access to literacy instruction, a group of Texas-based researchers collaborated to develop a smartphone/tablet-based literacy app. As part of the Barbara Bush Foundation Adult Literacy XPrize Global competition, the app called *Codex: The Lost Words of Atlantis* (from here, the app) was developed in 12 months and designed specifically for adult multilingual non-readers to acquire foundational English literacy skills (People ForWords, 2017).

Rationale for App Design

The App Design

A smartphone or tablet provides access to reading instruction through the app. Because of this, individuals learn and practice foundational English literacy skills in private, during short time periods, and at their own pace in order to mitigate larger social structures and policies that may impact access to adult education (Biglow & Vinogradov, 2011). In addition, research on gaming suggests that an adventure format may engage players to persist in improving upon their current performance levels (Wichadee & Pattanapichet, 2018; Woo, 2014). Therefore, the app is designed within the context of an adventure where players assume the identity of an enterprising archaeologist searching for clues to the forgotten language of the lost civilization of Atlantis.

Players begin in the country of Egypt and move through five levels of gameplay by completing literacy games focusing on different foundational reading skills including: letter-sound fluency; word identification fluency; and phonological decoding fluency using onset-rime. (See Table 1.) During each game, the sound or word is said aloud and the user selects it on the screen. (See Figure 1.) To support Spanish-speaking multilingual learners, a read-aloud dictionary supports meaning-making across Spanish and English languages. The skills and content in each level are reinforced through games that mimic real world tasks, such as reading mailing labels, street signs, or lists of objects.

FIGURE 1: The App in Use: Learning Sight Words

Notes: Image from *Codex: The Lost Words of Atlantis* (Southern Methodist University, n.d.)

Letter-Sound Fluency

Players begin each (new) level playing a game providing instruction and practice with letter-sound fluency. The rationale for starting with this skill is adults at or below Literacy Level 1 need instruction on phonological tasks required of proficient reading since they show reliance on experiential word knowledge and there may be a presence of double deficits in phonological processing and speed/rate of word processing compounding already weak phonological abilities (Sabatini, 2002). In addition, adults who are multilingual typically know that print carries meaning but may not be able to link or manipulate letters-sounds, unitize words, and in most cases, assign meaning to English words and phrases, other than basic words and symbols (Adrian et al., 1995; Kolinsky et al., 1987; Morais et al., 1979).

Sight Word Fluency

Adults at or below Literacy Level 1 show a reliance on experiential word knowledge, including sight words (Eskey, 2005). However, this knowledge is often unsystematic and incomplete. Thus, sight

TABLE 1: Literacy Skills Developed During Gameplay in Egypt

Level	Letter Sound Fluency *Match lower case letters with their common sounds.*	Word Identification Fluency *Match written and spoken sight words.*	Phonological Decoding Fluency *Blend and segment consonant-vowel-consonant (CVC) & consonant vowel consonant consonant (CVCC) words using onset-rime.*
1	a, f, m, p, s, t	40 sight words across 5 levels	-am, -ap, -ast, -at words
2	d, g, h, n o		-ad, -ag, -an, -og, -op, -ot words
3	i, k, l, w		-id, -ig, -ip, -ill, -in, -it words
4	b, c, r, u		-ab, -ib, -ob, -ub, -ug, -un, -up, -ut words
5	e, j, v, x, y		-et, -en, -ed, -eg, -ell, -ax, -ex words

words taught and practiced during gameplay may ensure emerging adult multilingual readers to fluently identify and understand connected text.

Phonological Decoding

Since phonological decoding is an area of weakness in adults at or below Literacy Level 1 (Greenberg et al., 1997, 2002; Thompkins & Binders, 2003), we designed a game to build and segment CVC words by onset-rime. Further, research indicates that adults benefit from explicit instruction on phoneme segmentation and manipulation (Adrian et al., 1995; Kolinsky et al., 2019; Kruidenier, 2002; Kruidenier et al., 2010; National Reading Panel, 2000), as well as practice of building and maintaining phonological representations of words (Dietrich & Brady, 2001). Therefore, gameplay initially focuses on explicit instruction of phonemes and phoneme segmentation to develop word decoding skills.

Purpose of the Current Study

In sum, the literature suggests that adults at or below Literacy Level 1 need systematic instruction of letter-sound fluency, word identification fluency, and explicit practice with phonemic segmenting and blending tasks. Because of the need for such an approach, we hypothesize the app could benefit adult multilingual learners at or below Literacy Level 1 and who may not have access to adult literacy classes. Therefore, our pilot study investigates two research questions:

1. Is there a relationship between playing the app and phonological decoding of taught and untaught words?
2. Will adult multilingual learners make literacy gains from pre-tests to post-tests on distal measures during the 8-week study?

Methods

Research Design

This 8-week study was conducted to understand the relationship between the app and phonological decoding of taught and untaught words, as well as the acquisition of foundational English literacy skills from pre-tests to post-tests. Since the population of our study has been historically marginalized, an experimental group design with a sufficient sample size was difficult to obtain. Given this, we used an experimental single-case research design (Kratochwill et al., 2010). Single-case design provides rigorous experimental control with each case serving as its own control. Multiple cases provide experimental replication and corresponding effect sizes are easily aggregated for estimation of overall causal effects (Burns, 2012).

The goal of this design was to provide experimental evidence to evaluate the change in targeted reading skills in response to the app (the "intervention"). Very simply, each participant provided a series of carefully obtained measures prior to using the app (the "baseline" phase), followed by a series of similar measures obtained while using the app (the "intervention" phase). As described below, analysis of the time-series data statistically contrasts the intervention data to the baseline data.

Based on an initial placement assessment, gameplay started at the participant's current skill level, followed by a series of very brief proximal measures of skill mastery, specifically mastery of phonological decoding of words taught (or practiced during gameplay) and skill transfer of learning to untaught words. After demonstrated mastery (after assessing at least five words), the app increments to the next skill level. These proximal measures were obtained during the baseline and intervention phases, resulting in a series of data points suitable for statistical analysis and estimation of experimental effects. In addition, distal measures were collected pre- (beginning of baseline) and post- (end of treatment), (i.e., letter-sound fluency, word identification fluency, phonological decoding fluency of isolated sounds and words) providing two data points for each measure. These data measured more global variables hypothetically related to use of the app.

With each of the nine participants treated as a single-case, we collected multiple-baseline data to test the effectiveness of the app across game levels and to examine the relation between using the app and progress on the proximal measure (e.g., Allor et al., 2018; Gast et al., 2014). For each individual case (i.e., participant), the multiple baseline design allowed us to experimentally study the effect of the app on reading across the different levels of gameplay. At each level, the baseline phase provides data for each case serving as their own control when the intervention is introduced. Based on placement tests, all participants began in Level 1. Each participant was treated as a unique case study and then the nine single-case results were aggregated for inference about the overall treatment effect.

Setting and Participants

After Institutional Review Board approval, we conducted an 8-week pilot study to test the effectiveness of the app. Because of budget and time constraints (due to our participation in the Barbara Bush Foundation Adult Literacy XPrize Global competition), we opted to use a convivence sample to understand the effectiveness of our app on the target audience: Spanish-speaking multilingual adults. To recruit participants, we provided flyers to local community centers in the Southwest. To be included in the study participants had to (a) be 18 years of age or older; (b) Spanish-speaking multilingual adults; (c) have not had the opportunity to develop basic literacy skills in English and in Spanish; (d) score at or below grade level 1.5 on a battery of standardized measures; (e) not be involved in any secondary or postsecondary education; and (f) have no known intellectual, hearing and/or sight disabilities. We recruited nine participants who were born outside of the United States and wanted to learn to read and write in English. For example, many of the participants, like Isabella and Luciana, wanted to learn to read to support their grandchildren and/or children who attended a local elementary school. Others, like Samuel and Sara, wanted to be able to better communicate with others in their workplaces. (See Table 2.)

TABLE 2: Participant Demographics

Participant*	Age	Gender	Highest Grade Level Completed	Technology Use Before Study
Sofia	60	Female	9th Grade	Cell Phone & Limited Game Play
Isabella	58	Female	2nd Grade	Cell Phone & Limited Game Play
Valentina	53	Female	9th Grade	Cell Phone & Limited Game Play
Samuel	56	Male	8th Grade	Cell Phone for calls only
Camila	20	Female	1st Grade	Cell Phone & Limited Game Play
Luciana	36	Female	8th Grade	Cell Phone & Limited Game Play
Jimena	54	Female	8th Grade	Cell Phone & Limited Game Play
Gabriela	49	Female	5th Grade	Cell Phone & Limited Game Play
Sara	52	Female	3rd Grade	Cell Phone & Limited Game Play

* **Note:** participants names are pseudonyms

Procedures

Participants were pre-tested using a competency-based standardized test of general reading ability and three one-minute assessments, which measured letter-sound fluency, word identification fluency, and phonological decoding fluency of isolated sounds and words. These measures were chosen to correspond directly to the skills taught in the game.

Participants received tablets and hotspots to access the app and online conferencing software. We found that use of the tablets and hotspots required extensive training due to participants' unfamiliarity with these technologies and we provided in-person workshops to overcome this barrier. Next, we trained participants to play the game and were asked to play the game four times a week for approximately 20 minutes across 8 weeks.

During the study, two multilingual data collectors contacted the participants four times each week via Zoom to test their word-reading ability using proximal measures on taught and untaught words (described below). These interactions were recorded. Upon study completion, we post-tested the participants using equivalent forms to the pre-test battery of standardized assessments. Participants were allowed to keep the tablets they used during the study to continue their learning.

Participant Measures

Screening

To be included in the study, participants needed to score at or below grade level 1.5 on relevant measures including: Letter Identification, Word Attack, and Passage Comprehension subtests from the Woodcock Johnson III Tests of Achievement (Woodcock et al., 2001). To provide an indication of overall lexical knowledge, the Picture Vocabulary subtest from the Woodcock-Munoz - Spanish Version (Woodcock & Munoz-Sandoval, 1993) was also administered. (See Table 3.)

Proximal Measures

Proximal measures were administered to see if participants could decode increasingly complex *taught* (words that they had exposure to during gameplay) and *untaught* decodable words for a given game level. For experimental purposes, each level of the game has a baseline and an intervention phase. Within the baseline and intervention phases of the five levels of gameplay, we created a proximal measure of 60 decodable words aligned to that level of gameplay and five alternate forms for each individual player. We used alternative forms to: (a) establish the pre-intervention baseline for *untaught* words that will be taught in the next level of gameplay, (b) determine mastery of *taught* words that players had exposure to within the current level of gameplay, and (c) observe any *transfer* effects to untaught words containing letter-sound combinations that players had previous exposure to in the game. With correct responses scored "1" or "0" otherwise, at least five data points were recorded for each of the five baseline-intervention measures (Kratochwill et al., 2010). (See Table 3.) Data collection continued until measures were obtained four times per week via Zoom Teleconferencing. Each testing session was untimed, and video recorded.

During the baseline phase of each level participants responded to measures of both taught and untaught words. The taught words were measures of the current phase effect, while the untaught words provided the baseline measure for the next level, i.e., untaught words are sampled from the next level. As noted above, each item is scored "1" correct or "0" otherwise, and a score of 85% on at least 5 data points of taught words was required prior to advancing to the next level.

Distal Measures

In addition to the short-term weekly measures, overall literacy gains were measured in person at the beginning and end of the study using the following measures: (a) the computer adaptive version of the Comprehensive Adult Student Assessment Systems (CASAS; Posey & Jacobsen, 2009); (b) Letter-Sound Fluency (Fuchs et al., 2001); (c) Word Identification Fluency (Fuchs et a., 2004); and (d) Decodable CVC Test (researcher-created) of isolated letter-sounds and words. (See Table 3.)

Survey

To provide us with information regarding the feasibility of the game app, we developed a participant survey. The survey was constructed using the five components of feasibility, as established by Bowen and colleagues (2009), including: (a) fidelity of implementation and dosage, (b) acceptability to end-users, (c) demand, (d) practicality, and (e) end-user recommendation for expansion or modification. This instrument allowed for open-ended as well as Likert scale responses. The data collectors conducted individual survey sessions, which were untimed and video recorded. (See Table 3.)

Data Analysis

Proximal Measures

Using the web-based Tau Calculator (Tarlow, 2016a), we determined if a baseline trend existed, i.e., was the participant performance changing prior to treatment. By design, we hypothesized no change during baseline prior to implementation of the baseline. If a trend did not exist (Tarlow, 2016b), a visual analysis of the data was performed to notice patterns in the trendlines for the baseline, taught, and transferable word-decoding data, as it was an important step in drawing conclusions about validity (Parker et al., 2011). The statistic Tau was then calculated to as the effect size index (Tarlow, 2016b).

Distal Measures

The Wilcoxon Signed Rank Test was used to analyze the pre-to-post distal measures. We did not make assumptions about the population

TABLE 3: Participant Measures

Source	Measure	Skills	Time
Screening	**Woodcock Johnson III[a]**		
	Letter-Word identification	Identify letters and words	Untimed
	Word Attack	Decode letters and words	Untimed
	Passage Comprehension	Read and understand text	Untimed
	Woodcock-Munoz[b]		
	Picture Vocabulary	Identify pictures in Spanish or English	Untimed
Proximal[g]	Taught words	Read words taught from current game level	
	Transferable words	Read words made up of previously taught letter-sounds	5 minutes
	Untaught words	Read words from next game level to establish baseline	
Distal	CASAS[f]	General reading knowledge related to life/work skills	Untimed
	Letter Sound Fluency[c]	Letter-sound recognition	1 minute
	Word Identification Fluency[d]	Sight word reading ability	1 minute
	Decodable CVC Word Reading Fluency[e]	Phonological decoding ability of isolated letter-sounds and words in closed syllable word patterns	1 minute
Survey	Feasibility of the App	Fidelity of implementation and dosage, acceptability to end-users, demand, practicality, and end-user recommendation for expansion or modification	Untimed

Notes: [a]Woodcock et al., 2001; [b]Woodcock & Munoz-Sandoval, 1993; [c]Fuchs et al., 2001; [d]Fuchs et al., 2004; [e]Researcher-created; [f]Posey & Jacobsen, 2009; [g]Researcher-created.

distribution and therefore used the small sample nonparametric alternative to the repeated measures t-test to determine if there were significant changes pre- to post-intervention.

Results

Proximal Measures: Relationship Between Gameplay and Literacy Outcomes

Research question 1 asked whether a relationship exists between playing the app and phonological decoding of taught and untaught words. Consistent with the Meets Standards recommendations as outlined by the What Works Clearinghouse (Kratochwill et al., 2010), at least five data points for the baseline and intervention conditions were collected and 85 percent word-reading accuracy was required before individuals could move to the next level of gameplay. To this end, the data for five baseline-intervention conditions were visually and statistically analyzed to investigate the game effect (Parker et al., 2011). First, we used the web-based Tau Calculator (Tarlow, 2016a) to test if a statistical trend existed within the participant's baseline series of data points. When a baseline trend did not exist, we performed a visual analysis. In these data, we noticed a marked change in reading proficiency once participants advanced to Levels 3, 4, and 5 of gameplay. To describe these intervention effects in detail, we examined three changes in the data: trend, variability, and performance (level). (See Figures 2 and 3).

Trend

First, we examined trendlines for (a) baseline (currently *untaught* words related to the next level of gameplay), (b) intervention (*taught* words), and (c) trendlines for *transfer* words. As trend reflects movement in the data over time, we looked at the visual changes in the trendlines between the baseline and intervention conditions, noting the direction of the data path and the immediacy of effect within and across participants. In Levels 1 and 2 (of the game), we observed the trendlines for baseline untaught words, intervention taught words, and transfer words were positive (with a few exceptions), indicating that participants may have improved without intervention. However, beginning in Level 3, the baseline trend for most participants' untaught words became flat or even negative, while the slope for taught and transfer word-reading ability became increasingly positive. This shift in baseline trend and subsequent acceleration of the intervention suggests that in Levels 3, 4, and 5, the game may have begun to support literacy development. Notably, in five of nine participants, Level 4 trendlines for taught words show an immediacy of effect between the baseline and intervention condition with no overlap. The slopes of the remaining four participants indicate a positive, but delayed response with some overlap. For Level 4 of the transfer words, one of nine show immediacy for taught words, while marginally delayed responses were observed in slopes of five participants. The transfer-word trendlines of the remaining three participants tended to be flat.

Variability

Next, we looked at the variability in the baseline, taught word, and transfer word trends. Upon visual analysis, we noticed a wide fluctuation of the trendlines across all participants. Because we had a strong a priori assumption that multilingual learners with low-literacy had limited experiential sight word recognition, variability in the baseline was somewhat anticipated; however, variability in the intervention was also observed, indicating that participants may not have been consistent in the pronunciation or identification of the letter-sounds when reading taught and transfer words. We note that this variability may also have come from extraneous random error associated with the participants' unfamiliarity with the technology.

Performance

Lastly, in addition to analysis of trend over time for each participant, we analyzed their performance (level) after controlling for trend (should one be present). For this analysis, we used the single-case statistic baseline corrected Tau (Tarlow, 2016b) to test statistical significance and compute corresponding effect sizes. Ideally, after controlling for baseline trend, baseline corrected Tau can be interpreted as the percent of non-overlapping data points after implementation of the game app. On another metric, it can be interpreted simply as an adjusted correlation between the phase and the measure. (See Tables 4 and 5.) A pattern emerged within the taught word data showing Level 4 of gameplay had statistically significant ($p < 0.037$) and moderate effect sizes greater than 0.6 for eight of the nine participants. Further, within the transfer word data, Level 5 of gameplay had statistically significant ($p < 0.035$) and moderate to large effect sizes (0.648 to 0.833) for five of nine participants.

Overall

With the effect sizes computed for each of the nine single-cases, we statistically estimated an overall effect by aggregating the effect sizes across cases. This procedure followed Tarlow's (2016b) recommendation for computing weights, and estimating a mixed model to compute the effect size Tau across all records for each participant and overall across each gameplay level. We found statistically significant (p = 0.000) and moderate effect sizes (0.721 and 0.497) for Levels 4 and 5 of the taught word measure and a statistically significant (p = 0.001) and moderate effect size (0.645) for Level 5 of the transfer word measure. These findings support the hypothesized relationship between participation in the gameplay and improved literacy outcomes, as well as confirm the prior visual inspection of the data where participants taught and transfer word-decoding abilities increased during the later levels of the game. We do conclude, however, further investigation of Levels 1, 2, and 3 is warranted.

FIGURE 2: Graphs of Proximal Taught Word Data for Each Participant

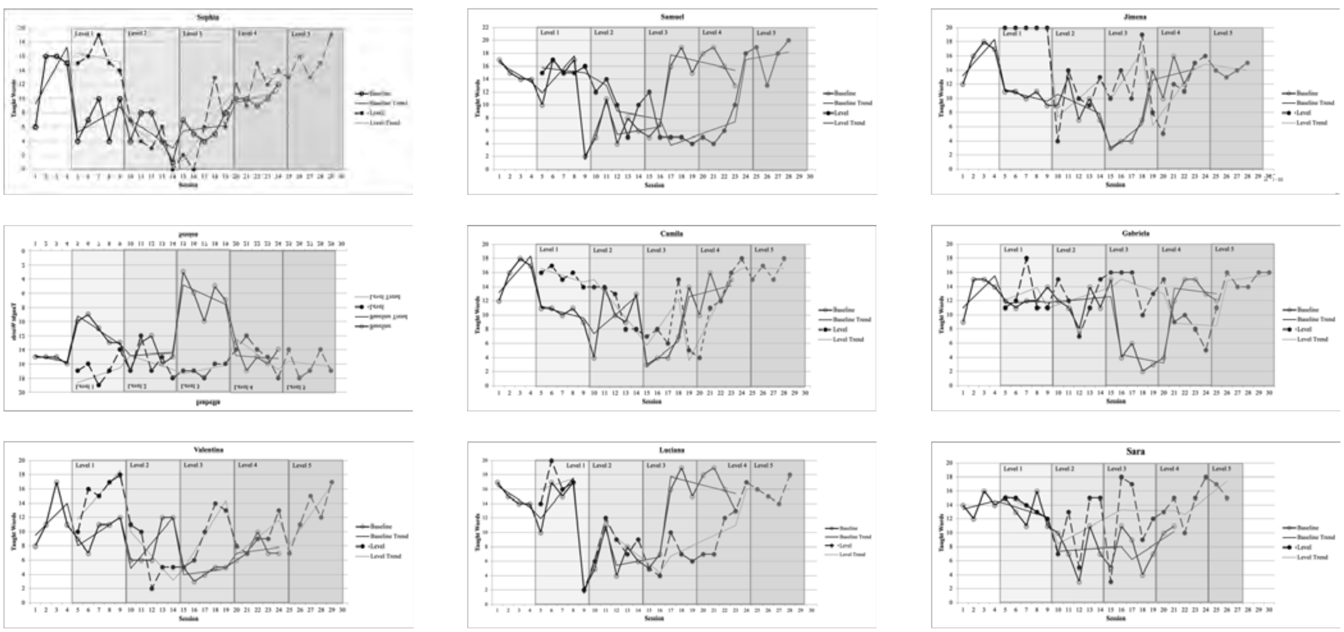

FIGURE 3: Graphs of Proximal Transfer Word Data for Each Participant.

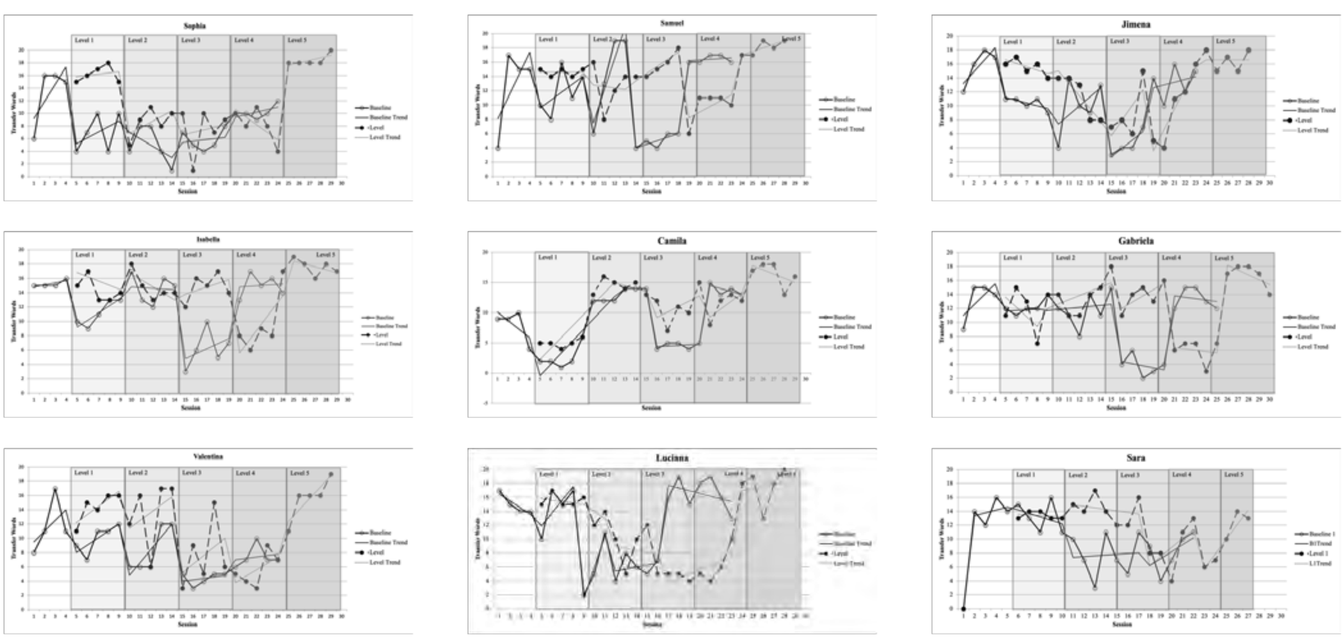

TABLE 4: Tau Effect Size Measures for Taught Words

Level	1		2		3		4		5		Combined*	
Participant	Tau	p-value	Tau	p-value	Tau	p-value	Tau	p-value	Tau	p-value	Tau	p-value
Sofia	0.082	0.899	-0.443	0.167	-0.031	1.000	**0.762**	0.012	**0.781**	0.011	0.412	0.000
Isabella	0.442	0.209	**0.640**	0.036	0.544	0.087	**0.754**	0.012	0.288	0.392	**0.581**	0.000
Valentina	0.268	0.459	-0.469	0.138	0.130	0.746	**0.781**	0.011	**0.641**	0.041	0.381	0.000
Samuel	0.040	1.000	**0.767**	0.019	0.000	1.101	**0.772**	0.011	-0.035	1.000	0.466	0.000
Camila	**0.679**	0.013	-308*	0.267	-0.533	0.191	**0.791**	0.017	**0.830**	0.015	0.465	0.000
Luciana	0.357	0.372	-0.645	0.108	0.101	0.880	0.422	0.160	-0.241	0.443	-0.006	0.939
Jimena	**0.877**	0.011	-0.352	0.286	0.356	0.317	**0.680**	0.037	0.158	0.670	**0.532**	0.000
Gabriela	-0.098	0.884	0.085	0.845	0.343	0.220	**0.693**	0.021	0.544	0.088	0.390	0.057
Sara	-0.04	1.000	-0.13	0.682	**0.613**	0.023	**0.671**	0.027	N/A	N/A	-0.014	0.022
Combined**	0.028	0.003	-0.005	0.989	0.219	0.387	**0.721**	0.000	**0.497**	0.000	-	-

Notes: *Estimates from mixed model computing the effect size across all case records.
** Estimates from mixed model computing the effect size for each level across case records.

TABLE 5: Tau Effect Size Measures for Transfer Words

Level	1		2		3		4		5		Combined*	
Participant	Tau	p-value	Tau	p-value	Tau	p-value	Tau	p-value	Tau	p-value	Tau	p-value
Sofia	0.327	0.375	0.281	0.396	0.593	0.056	0.443	0.167	**0.833**	0.009	**0.594**	0.000
Isabella	-0.374	0.304	**0.718**	0.020	0.031	1.000	0.427	0.172	**0.648**	0.035	0.400	0.000
Valentina	0.316	0.381	0.432	0.171	-0.195	0.589	0.197	0.588	**0.801**	0.010	0.440	0.000
Samuel	-0.179	0.686	0.117	0.804	0.000	1.098	**0.746**	0.018	**0.710**	0.028	0.417	0.000
Camila	0.023	1.000	**-0.553***	0.041	-0.465	0.222	**0.803**	0.017	0.553	0.104	0.169	0.078
Luciana	0.335	0.436	-0.258	0.593	0.264	0.539	-1.84	0.576	-0.184	0.576	-0.418	0.503
Jimena	-0.198	0.617	0.158	0.670	-0.149	0.713	0.545	0.105	0.593	0.056	0.244	0.001
Gabriela	-0.3	0.460	0.39	0.206	0.308	0.283	0.544	0.088	**0.656**	0.034	0.377	0.000
Sara	-0.258	0.515	0.238	0.413	0.195	0.511	0.213	0.528	N/A	N/A	0.108	0.022
Combined**	-0.027	0.512	0.184	0.532	0.105	1.00	0.338	0.00	**0.645**	0.001	-	-

Notes: *Estimates from mixed model computing the effect size across all case records.
**Estimates from mixed model computing the effect size for each level across case records.

Distal Measures: Pre-test to Post-test Literacy Gains

Next, we share our findings to research question 2: Will adult multilingual learners make literacy gains from pre-tests to post-tests on distal measures during the 8-week pilot study? Overall, we found all nine participants improved their scores on pre- and post-test computer adaptive administrations of the CASAS, affirming that adult multilingual learners made literacy gains on pre-tests and post-tests after participating in gameplay.

Six of the nine participants grew by 10 points or more. The CASAS moved eight of the nine from the Pre-Reading Literacy test (27R) at pre-test to the Beginning Reading test (81R) at post-test. (See Table 6.) Findings suggest that over the course of the study, adults at or below Literacy Level 1 may have been using the skills they had been working on in the game, applying and then transferring the skills to other contexts, particularly those that are critical for life and work reading.

Additionally, for all five measures, we obtained statistically significant gains using Wilcoxon Signed-ranks Test to compared pre-to-post scores. Referring to Table 7, the following results were obtained: CASAS ($p < .008$); Letter-Sound Fluency ($p < .008$); Word Identification Fluency ($p < .008$); CVC decoding of individual letter-sounds in words ($p < .011$); and CVC decoding of whole words ($p < .011$). Further, the effect size for each measure was greater than .6, interpretable as large (Cohen, 1988). These values indicate a possible positive game participation effect on literacy skill development.

TABLE 6: Pre- and Post-Test Reading Measures

	CASAS					Letter Sound Fluency			Word Identification Fluency			Correct Letter Sounds (CVCCLS)			Whole Word Fluency (CVCWWF)		
Participant	Test	Pre	Test	Post	Gain	Pre	Post	Gain	Pre	Post	Gain	Pre	Post	Gain	Pre	Post	Gain
Sofia	27R	194	81R	205	**11**	2	28	**26**	13	36	**23**	17	151	**134**	5	48	**43**
Isabella	27R	196	81R	206	**10**	9	12	**3**	19	40	**21**	72	145	**73**	23	45	**22**
Valentina	27R	181	81R	200	**19**	10	16	**6**	19	31	**12**	61	127	**66**	19	40	**21**
Samuel	27R	202	81R	205	**3**	1	26	**25**	6	32	**26**	18	107	**89**	5	34	29
Camila	27R	189	81R	191	**2**	3	16	**13**	7	10	**3**	3	66	**63**	1	21	**20**
Luciana	27R	174	28R	199	**25**	8	31	**23**	19	40	**21**	33	121	**88**	11	38	**27**
Jimena	27R	184	81R	189	**5**	9	16	**7**	26	30	**4**	124	123	**-1**	40	39	**-1**
Gabriela	27R	194	81R	205	**11**	0	14	**14**	10	35	**25**	14	95	**81**	4	28	**24**
Sara	27R	191	81R	203	**12**	2	46	**44**	15	19	**4**	52	97	**45**	16	33	**17**

TABLE 7: Wilcoxon Signed-Ranks Test for Distal Measures

	Pre-Intervention		Post-Intervention		Z	Sig. (2-tailed)	r
	Mdn	Range	Mdn	Range			
CASAS	191.00	28	203.00	17	2.668	.008	.629
LSF	3.00	10	16.00	34	2.666	.008	.629
WIF	15.00	20	32.00	30	2.670	.008	.629
CVCCLS	33.00	121	121.00	85	2.547	.011	.601
CVCWWF	11.00	39	38.00	27	2.547	.011	.601

Discussion

The purpose of this study was to determine if the app supported adult multilingual learners to develop English word decoding skills. Findings suggest that the app provides a viable approach to teaching adults at or below Literacy Level 1 how to read. We obtained significant and strong effects on proximal measures of targeted skills at levels 4 and 5 of the app, providing evidence of the participants' advancing literacy skills as they continued to learn letter-sound correspondences. Also, we found statistically significant gains on all distal measures supporting the hypothesis that participants learned as they played the app.

Relationship Between Gameplay and Literacy Outcomes

The first research question addressed a relationship between participation in the leveled gameplay and enhanced literacy outcomes on proximal word-reading measures. Using a single-case research design, we observed a relationship between participating in the intervention and improved reading scores on proximal measures of taught words. Participants across levels four and five showed significant improvement from baseline through intervention phases on word reading measures for taught and transfer words with combined effect sizes ranging from small to moderate. These data support an increased ability within each level of gameplay to recognize letter-sounds and unitize words directly taught in the app.

We questioned if repeated exposure to taught words could promote rote memorization of the practiced CVC patterns. To gain insight, we also investigated transfer word-reading. Transfer words are words that have similar CVC patterns but are not explicitly taught during gameplay. In these data, we also found a relationship between participating in the intervention and improved reading scores on proximal measures of transfer words. Seven out of nine participants showed significant improvement from baseline through intervention phases on word-reading measures of transfer words with effect sizes ranging from small to moderate. While it might have been possible for participants to begin to recognize taught words by sight, jeopardizing our conclusion of the app's intervention effect, an individual's enhanced ability to read transfer words makes this finding more unlikely. The transfer words required processing of different letter-sound combinations. Thus, the gains associated with these words support the hypothesis that participants were developing phonological processing abilities that likely contributed to accurate decoding and reading of text (Dietrich & Brady, 2001; Tunmer & Nesdale, 1985).

Despite these gains, it was important to note that it did take several weeks of gameplay to see significant literacy growth as measured by the effect sizes on the proximal measures across participants. This finding may be because it took time for participants to become familiar with the new technologies, or it may be because individuals had a reliance on word structure (i.e., experiential word knowledge) rather than on an understanding of phonological decoding (Thompkins & Binder, 2003).

As mentioned, not every participant scored statistically significant combined gains: Gabriela and Camila had inconsistent results. Gabriela did not show significant improvement from baseline to intervention for taught words but did show statistically significant growth on transfer words. We observed improvement beginning at Level 4 for taught words and at Level 5 for transfer words. During post-survey, Gabriela noted that "learning the letters' sounds and making the words" helped her to improve her ability to read. Camila, on the other hand, showed statistically significant progress from baseline to intervention reading taught words and not transfer words. Camila's growth strengthened at Levels 4 and 5 for taught words and at Levels 2 and then 4 for transfer words. She offered that the app helped her "to recognize words," but "in the start, it was hard..." She also mentioned that she "liked learning the sounds" of the English letters. Clearly, these cases support the hypothesis that participation with the app provides opportunity to learn letter/sound correspondence and therefore rely less on whole word knowledge and more on an ability to actively decode new words. These findings across participants tend to corroborate the improvement we saw from pre-test to post-test on the distal measures. (See Table 6).

Still, we wondered about Luciana who did not show significant improvement from baseline to intervention on any level of gameplay for taught and transfer words. Looking across data sources, she did show marked growth on the pre- and post-tests measuring letter-sound correspondences and decoding of CVC words. In the post-survey, she did not answer the question indicating whether the app helped her learn to read. Yet, she noted difficulty in being able to match sounds to the letters. Currently, Luciana's findings are inconclusive.

Pre-Test to Post-Test Literacy Gains

Our second research question was to examine whether multilingual adults at or below Literacy Level 1 would grow in their literacy learning across the 8 weeks of the pilot study. For all measures given, we found statistically significant increases in the means from pre-test to post-test. Moreover, results support that the participants were advancing in their word decoding skills and applied the skills they learned in the game to life- and work-related reading tasks (antidotally and as measured by CASAS).

As discussed, for both native speakers of English and multilingual adults at or below Literacy Level 1, several studies have found deficits in individuals' phonological processing and decoding ability, and particularly in explicitly linking letters to sounds (e.g., Adrian et al., 1995; Greenberg et al, 1997, 2002; Kolinsky et al., 2018; Thompkins & Binder, 2003). We found similar weaknesses in our participants. This evidence allowed us to draw inferences about how individuals might respond to the app with its carefully sequenced curriculum, opportunities for dedicated practice and feedback, and easy access to instruction. We hypothesized that the app would address critical gaps in letter-sound knowledge and in basic sight word identification for adults at or below Literacy Level 1 which, once filled, would propel them forward in their phonological decoding skills and ultimately advance their general reading scores. We deliberately chose three 1-minute measures that tested participants' ability to recognize letter-sound correspondences, to identify common sight words, and to decode CVC patterns and then blend them back into whole words, since these are also the skills that are directly taught in the game.

The results of our investigation supported our hypothesis. Findings suggest consistency in playing the app may impact participants' literacy skills by improving phonological decoding abilities. For each level of play, they learned and practiced a corpus of four to five new letter-sounds. Throughout, they were instructed to recognize and link each single letter to its regular or most common sound. Initially, for our Spanish-speaking participants, this proved to be most challenging, not only due to several articulation differences between English and Spanish phonemes, but also because they had never been exposed to English letter-sound relationships. Here, we infer that the app contributed to their fluency in recognizing individual letter-sounds (ranging from 3 to 44 additional letter-sounds across the study), but it was also especially instrumental in putting their newly found decoding skills into practice as they read more CVC words. (See Table 6.) Across eight of nine participants, we observed growth in individuals' ability to fluently segment letter-sounds and unitize words. The improvements ranged from 45 to 134 additional correct letter-sounds in one minute's time and from 17 to 43 additional whole words read in the same minute. Recognition of common sight words grew across participants as well, ranging from 3 to 26 more words read in one minute's time.

Yet, we wondered about Jimena who did not show growth in these areas because she improved only marginally in her letter-sound identification (9 to 16 letter-sounds) and sight word recognition (26 to 30 words). She scored one less letter-sound (124 to 123) and one less whole word (40 to 39) on the Decodable CVC measure from pre-test to post-test. In general, looking across her scores, a pattern emerged. Jimena knew many words by sight: at pre-test, she read 124 decodable words, far more than any other participant. As noted, research supports adults at or below Literacy Level 1 often have significantly weak phonological skills and thus may come to rely on experiential word knowledge (Greenberg et al., 2002). We believe that although Jimena was exposed to new letter-sounds at each subsequent level of gameplay, she did not specifically make connections to or internalize how these letter-sounds could be used to segment and blend words. She relied on her own personal sight word knowledge. This lack of growth supports this inference as does her comment in the post-survey data stating, "At the beginning, it [the app] wasn't challenging…" It is highly likely that the game never reached an appropriate difficulty level for Jimena, giving her a reason to use any newly acquired phonological decoding skills that she might have attained.

Across all individuals in our study, we sought to examine if there would be increases in the participants' ability to transfer their new knowledge to work- and life-related reading tasks. At the outset, we felt that growth in general reading scores would be a reliable indicator of whether the app was impacting the participants' ability to use their literacy skills to function in society and become independent readers. As the CASAS (Posey & Jacobsen, 2009) is a standardized instrument used by general stakeholders to test the attainment of real-life competencies and skills, we believed that this assessment would suit our purposes well. The CASAS links scale scores developed through item response theory to basic skill competencies. An individual's ability is then reported as a score along a fixed metric, ranging from 150 to 260. Any increase in scores over time represents the examinee's ability to perform at higher gradations of difficulty in their skills.

Over the course of the study, we observed that all participants improved upon their pre-test CASAS scores, and six of the nine improved by 10 points or greater with one participant scoring 25 points higher. If literacy acquisition constitutes mastery of a continuum of skills (e.g., Adams, 1990) beginning with explicit and systematic letter-sound instruction (Kolinsky et al., 2018), then our data and the increases we observed support the position that the app was effective in advancing the beginning reading skills of our participants. In addition, our results add to evidence supporting the notion that direct instruction of the phonological and orthographic structure of words may be an effective way to build the literacy skills of adults at or below Literacy Level 1 (e.g., Kolinsky et al., 2019; Kruidenier, 2002; Kruidenier et al., 2010). The fact that this instruction was accessible to individuals as an app downloaded to their personal devices may also be key to its success and a viable solution to issues surrounding accessibility of instruction.

Conclusion

In conclusion, the present study extends the research base about reading instruction for adult multilingual learners at or below Literacy Level 1, while adding to the knowledge of how gaming apps may support adult learning. Consistent with current research using explicit and systematic phonics instruction to teach phonological and orthographic structures of language (see Kolinsky et al., 2018; Kolinsky et al., 2019; Kruidenier, 2002; Kruidenier et al., 2010), we found positive outcomes in using the app. These results contribute to the body of evidence supporting the understanding that skills learned within a game environment may transfer to real life (Tobias et al., 2015). Evidence to support this was found in the increased CASAS (Posey & Jacobsen, 2009) scores since this test was designed to measure the attainment of real-life competencies and skills. Further, participants reported feeling more confident in their ability to read (and speak) English. This may be because gaming motivated players to persist in trying to improve upon their current performance levels (Wichadee & Pattanapichet, 2018; Woo, 2014). As such, the self-regulated learning offered through the app may have encouraged adult learners to move beyond their reliance on word structure to a new understanding of phonological decoding (Kolinsky et al, 2019; Thompkins & Binder, 2003).

While we provided a proof-of-concept showing how the app provides access to the knowledge and skills necessary to be able to teach adult multilingual learners at or below Literacy Level 1 how to read, we realize that we need to understand more about the affordances and constraints of the technology. Therefore, our next steps are to unpack this relationship by investigating survey and corresponding interview data, allowing us to determine how prior experience with technology may impact learner success. Moreover, our analysis will narrow to focus on understanding how participants took up the technological aspects of gameplay in relation to their literacy development.

Interpretation of our findings requires consideration of methodological limitations. Ideally, a stable baseline should be established prior to implementation of the intervention. However, due to the design, one limitation is that the baseline for proximal measures was not clearly established across all cases due to inconsistencies in the pronunciation of the CVC words. This limitation was mitigated by the authors' video recording the entire corpus of data and establishing scoring reliability across the measures. Further, we used the baseline corrected Tau statistic to consider the estimated baseline trend and its variability.

Another limitation was that the research was conducted using one small sample of multilingual adults at or below Literacy Level 1 from the same community center. Despite following the Meets Standards of the What Works Clearinghouse with fidelity, it was not clear if the participants relied on each other to support their literacy gains. Although all testing was done in their homes via Zoom teleconferencing, we cannot verify that they practiced their newfound skills elsewhere.

A final limitation was that the app was not a full curriculum; it was designed to support foundational learning of basic literacy skills by filling the gaps experienced by adults at or below Literacy Level 1. In structuring the scope and sequence, upholding the evidence that literacy follows a continuum of skills from the earliest identification of letter-sound correspondences to the processing of sub-word parts, words, sentences, and passages was foremost in our priorities as we sought to change nonreaders into readers.

References

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. MIT Press.

Adrian, J. A., Alegria, J., & Morais, J. (1995). Metaphonological abilities of Spanish illiterate adults. *International Journal of Psychology, 30*(3), 329-353. https://doi.org/10.1080/00207599508246574

Allor, J. H., Gifford, D. B., Jones, F. G., Al Otaiba, S., Yovanoff, P., Ortiz, M. B., & Cheatham, J. P. (2018). The effects of a text-centered literacy curriculum for students with intellectual disability. *American Journal for Intellectual and Developmental Disabilities, 123*, 474-494. https://doi.org/10.1352/1944-7558-123.5.474

Bigelow, M., & Vinogradov, P. (2011). Teaching adult second language learners who are emergent readers. Annual Review of Applied Linguistics, 31, 120-136. https://doi-org.huaryu.kl.oakland.edu/10.1017/S0267190511000109

Bowen, D. J., Kreuter, M., Spring, B., Cofta-Woerpel, L., Linnan, L. Weiner, D., Bakken, S., Kaplan, C. P., Squires, L., Fabrizio, C., & Fernandez, M. (2009). How we design feasibility studies. *American Journal of Preventive Medicine, 36*(5), 452-457. https://doi.org/10.1016/j.amepre.2009.02.002

Burns, M. K. (2012). Meta-analysis of single-case design research: Introduction to the special issue. *Journal of Behavioral Education, 21*(3), 175-184. https://doi.org/10.1007/s10864-012-9158-9

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Dietrich, J. A., & Brady, S. A. (2001). Phonological representation of adult poor readers: An investigation of specificity and stability. *Applied Psycholinguistics, 22*(3), 383-418. https://doi.org/10.1017/S014271640100306X

Eskey, D. (2005). Reading in a second language. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 563-580). Erlbaum.

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2004). Monitoring early reading development in first grade: Word identification fluency versus nonsense word fluency. *Exceptional Children*, *71*(1), 7-21. https://doi.org/10.1177/001440290407100101

Fuchs, D., Fuchs, L. S., Thompson, A., Al Otaiba, S. A., Yen, L., & Braun, M. (2001). Is reading important in reading-readiness programs? A randomized field trial with teachers as program implementers. *Journal of Educational Psychology, 93*(2), 251-167. https://psycnet.apa.org/doi/10.1037/0022-0663.93.2.251

Gast, D. L., Lloyd, B. P., & Ledford, J. R. (2014). Multiple baseline and multiple probe designs. In D. L. Gast & J. R. Ledford (Eds.), *Single case research methodology: Applications in special education and behavioral sciences* (pp. 251–296). Routledge.

Greenberg, D., Ehri, L. C., & Perin, D. (1997). Are word reading processes the same or different in adult literacy students and 3rd-5th graders matched for reading level? *Journal of Educational Psychology, 89*(2), 262-275. https://psycnet.apa.org/doi/10.1037/0022-0663.89.2.262

Greenberg, D., Ehri, L. C., & Perin, D. (2002). Do adult literacy students make the same word-reading and spelling errors as children matched for word-reading age? *Scientific Studies of Reading, 6*, 221-243. https://doi.org/10.1207/S1532799XSSR0603_2

Kolinsky, R., Cary, L., & Morais, J. (1987). Awareness of words as phonological entities: The role of literacy. *Applied Psycholinguistics, 8*, 223-232. https://doi.org/10.1017/S0142716400000278

Kolinsky, R., Leite, I., Carvalho, C., Franco, A., & Morais, J. (2018). Completely illiterate adults can learn to decode in 3 months. *Reading and Writing, 31*, 649-277. https://doi.org/10.1007/s11145-017-9804-7

Kolinsky, R., Carvalho, C., Leite, I., Franco, A., & Morais, J. (2019). How to teach fully illiterate adults to read. In D. Perin (Ed.), *The Wiley Adult Literacy Handbook.* John Wiley & Sons.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). *Single-case designs technical documentation*. What Works Clearinghouse. https://files.eric.ed.gov/fulltext/ED510743.pdf

Kruidenier, J. (2002). *Research-based principles for adult basic education reading instruction.* National Institute for Literacy. https://lincs.ed.gov/publications/pdf/adult_ed_02.pdf

Kruidenier, J. R., MacArthur, C. A., & Wrigley, H. S. (2010). *Adult education literacy instruction: A review of the research.* National Institute for Literacy. https://lincs.ed.gov/publications/pdf/adult_ed_2010.original2

Mamedova, S., & Pawlowski, E. (2019). *Adult literacy in the United States*. National Center for Education Statistics. https://nces.ed.gov/pubs2019/2019179/index.asp

Migration Policy Institute. (2019). *Texas: Language & education*. https://www.migrationpolicy.org/data/state-profiles/state/language/TX.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition, 7*, 323-331. https://doi.org/10.1016/0010-0277(79)90020-9

National Center for Education Statistics (2017). *The Condition of Education 2017* (NCES 2017- 144). U.S. Department of Education. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2020224

National Reading Panel (2000). *Report of the National Reading Panel - Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. National Institute of Child Health and Human Development.

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining Nonoverlap and Trend for Single-Case Research: Tau-U. *Behavior Therapy*, *42*(2), 284–299. http://doi.org/10.1016/j.beth.2010.08.006

People ForWords (2017). *Codex: The lost words of Atlantis*. Unpublished technical report.

Posey, V., & Jacobsen, J. (2009). *Comprehensive Adult Student Assessment Systems – CASAS.* California Department of Developmental Services.

ProLiteracy. (2022). *Adult literacy and basic education: The facts*. https://www.proliteracy.org/Adult-Literacy-Facts

Sabatini, J. (2002). Efficiency in word reading of adults: Ability group comparisons. *Scientific Studies of Reading*, *6*, 267-298. https://doi.org/10.1207/S1532799XSSR0603_4

Southern Methodist University (n.d.). *Codex: The Lost Words of Atlantis.* https://www.smu.edu/simmons/About-Us/XPrize/People-For-Words#Contact

Tarlow, K. R. (2016a). *Baseline corrected Tau calculator*. http://www.ktarlow.com/stats/tau.

Tarlow, K. (2016b). An improved rank correlation effect size statistic for single-case designs: Baseline corrected Tau. *Behavior Modification*, *41*(4), 427-467. https://doi.org/10.1177/0145445516676750

Texas Workforce Investment Council (2018). *Annual Report: Fiscal year 2019*. https://gov.texas.gov/uploads/files/organization/twic/2019-Annual-Report.pdf

Thompkins, A. C., & Binder, K. S. (2003). A comparison of the factors affecting reading performance of functionally illiterate adults and children matched by reading level. *Reading Research Quarterly*, *38*, 236-258. https://doi.org/10.1598/RRQ.38.2.4

Tobias, S., Fletcher, J. D., & Chen, F. (2015). Digital games as educational technology: Promises and challenges in the use of games to teach. *Educational Technology*, *55*(5), 3-12. https://www.jstor.org/stable/44430402

Tunmer, W. E., & Nesdale, A. R. (1985). Phonemic segmentation skill and beginning reading. *Journal of Educational Psychology*, *77*(4) 417-427. https://psycnet.apa.org/doi/10.1037/0022-0663.77.4.417

Wichadee, S., & Pattanapichet, F. (2018). Enhancement of performance and motivation through application of digital games in an English language class. *Teaching English with Technology*, *18*(1), 77-92. https://cejsh.icm.edu.pl/cejsh/element/bwmeta1.element.desklight-7aeb770c-7cdc-41c3-8a3b-668a67a9f088/c/ARTICLE5.pdf

Woo, J. C. (2014). Digital game-based learning supports student motivation, cognitive success, and performance outcomes. *Educational Technology & Society*, *17*(3), 291-307. https://www.jstor.org/stable/pdf/jeductechsoci.17.3.291.pdf

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement.* The Riverside Publishing Company.

Woodcock, R. W., & Munoz-Sandoval, A. F. (1993). *Woodcock-Munoz Language Survey – Spanish Form.* The Riverside Publishing Company.