Proof of Concept:

A Literature Review with In-App Comparisons

Speech Blubs

Helping Kids Speak!

Table of Contents:

INTRODUCTION

Speech Blubs is the leading speech therapy app for kids. Since its launch in the summer of 2017, it has had over 5 million downloads in countries across the world, has offerings in 5 different languages and dialects, and boasts impressive reviews from parents and professionals alike. At the time of writing, Speech Blubs is the most downloaded speech therapy app globally, has over 13,000 5-star ratings, and is the #2 app for kids under 5 in the US app store, right behind YouTube Kids. But how is this app made? How can it claim to be a speech therapy app? Is such a thing even possible?

This paper seeks to synthesize the existing literature to demonstrate the therapeutic benefits of the Speech Blubs app. The paper will consist of four sections: (1) Video Modeling (2) App Content and Mechanics (3) Gamification (4) Structure and Outside Support. The first section will explain the app's central mechanic: video modeling. Speech Blubs' bread and butter, this is the premise that underlies all of the app's functionality; at its simplest, it is the idea that children learn best from videos of other children. The second section will give a brief overview of some of the application's central features. It will discuss what children are taught, how, and why, along with justification for the therapeutic validity for each of these examples, along with specific examples from the app. The third section will discuss the gamification of the app, specifically, the importance of ludic learning, and the twofold appeal and effectiveness of basing learning in pleasure, particularly for young children. Finally, the last section will describe what Speech Blubs does to support its customers outside of the app experience, from resources and help to design and limiting screen time.

(1) VIDEO MODELING

Speech Blubs' foundational principle is Video Modeling. It is the lynchpin of the app's design, and almost omnipresent in the activities that make up the app's content. This section will explain the concept video modeling is founded on, mirror neurons, before elaborating on the efficacy of video modeling with children, for children – the mechanic that sets Speech Blubs apart from its competitors. Before understanding video modeling in Speech Blubs, it is critical to understand Social Learning Theory, mirror neurons, and the broader field of video modeling.

Social Learning Theory

Bandura's Social Learning Theory (1977) provides the conceptual underpinnings for video modeling. The theory hinges on the belief that human behavior is primarily learned by observing and repeating after, or modeling others. This imitation can be done live, recorded, or imagined, but what is critical is that the observed behavior can be replicated by the observer, upon observation (Corbett, 2005). Bandura identifies four distinct processes that mediate and facilitate observational learning: **attentional, retentional, production and**

motivational (Bandura, 1986). The **attentional** process describes the vicarious acquisition that takes place during perception. The **retentional** process occurs when the sensory observations recorded in the previous phase are understood, symbolically processed, and internally coded in order to be retained. The **production** process encapsulates the reproduction of the initially-observed behavior; indeed, it is when symbols become actions. Lastly, the **motivational** process describes the time in which learning occurs and is reinforced, either externally or by positive outcomes. It is this process that is critical to the adoption of an acquired skill.

The design of Speech Blubs' primary mechanic, the vocabulary-teaching sticker books, line up with these four phases exactly. The exercise starts with a child-teacher demonstrating the pronunciation of a target word several times. During this time, the child observes the word-production and observes the shapes the teacher's mouth makes, along with the sounds they are producing. The high repetition rate triggers the retentional phase. Certain "guessing exercises" give the child the opportunity to test retention by asking them to select an image tile with a picture of the target word on it once the word has stopped being repeated. The production process takes place when the app prompts the child to try and say the word. The display changes, and the app awaits audio input from the user. The motivational process is triggered by the app's positive reinforcement. The user gets a moderate degree of reinforcement from just having reached that phase of the exercise in the form of engaging visuals – celebratory bubbles float out from the microphone icon. If the child makes sounds, more bubbles come out faster. If the child says a word correctly or almost correctly, the bubbles go green. Regardless of correctness, they get verbal affirmation and the exercise marked as completed.

Mirror Neurons

But what is the actual science behind Bandura's Social Learning Theory and why video modeling works? The success of these methods can be explained by mirror neurons, discovered by Giacomo Rozzolatti and his team in the 1980s. A mirror neuron is a neuron that fires both upon performing and observing one same action. The idea is then that when a child observes somebody else speaking, their mirror neurons will fire, and encode the word in their own brains, so they can say it as well. While the original literature talks about physical actions, Rizzolatti (2009) finds that mirror neurons can be located in Broca's area, the part of the brain responsible for language production and speech processing. Therefore, it is likely that mirror neurons could be responsible for word production. Indeed, transcranial magnetic stimulation (TMS) experiments have produced evidence for a mechanism in the brain that translates heard phonemes into motor programs (Rizzolatti, 2009), providing an explanation for how infants were able to discriminate between sounds they couldn't yet produce (Hickok, 2010). This research provides a sound basis for the validity of video modeling, particularly in how it pertains to teaching children speech.

Furthermore, mirror neurons were also found in areas of the brain that mediate emotion-related behaviors, and brain imaging studies found that individuals' brains activated the same way not only when they felt, but also when they observed, disgusting or painful stimuli. These findings are highly suggestive of mirror neurons being involved in emotional sensation and recognition. This is especially pertinent for Speech Blubs, as it uses video modeling in its feeling-recognition section of the application. It has also been shown that synchronous observation and execution facilitates the formation of motor memories better than just one without the other. Of course, observation and action are each productive ways to learn, but combining them appears to give even stronger results (Rizzoalti, 2009). It is for this reason that Speech Blubs encourages simultaneous observation and production, and leaves room for one in the absence of the other.

It must be noted that mirror neurons are not without their controversy. The Rizzolatti theory has been the subject of much criticism (Heyes and Catmur, 2021; Hickok, 2010), notably, a reality-check as to what mirror neurons can actually do. Concisely, mirror neurons are not "telepathic" and cannot discern the intention behind an action one sees performed, however, they do help discern movement and motion, and with time, acquire their mirror properties through sensorimotor learning. To use the metaphor of playing an instrument, mirror neurons can help teach one how to play by observing others, and can lend an understanding of how the instrument works and interacts with the player, and even the quality of the player, but mirror neurons will not instantly transmit the knowledge to play into the observer's brain. The understanding of mirror neurons that the Speech Blubs app relies on makes full understanding of these limitations, among many others, and is very careful to only make decisions based on confirmed and replicable data.

Video Modeling

Wilson (2012) provides a succinct summary of what is known about video modeling, and where the literature calls for development.

What we know about video modeling[a]	*Questions that remain about video modeling*
• Video modeling has produced positive effects in a variety of areas (e.g., functional skills, communication, social skills) in children with autism spectrum disorders (ASD) in a variety of settings (e.g., home, clinic, school) (Ayres & Langone, 2005; Delano, 2007; Rayner, Denholm, & Sigafoos, 2009; Shukla-Mehta, Miller, & Callahan, 2010). • Skills gained through video modeling are often maintained and generalized (across settings, materials, and people) (Delano, 2007; Rayner et al., 2009). • Video modeling is a socially valid and noninvasive intervention procedure (Delano, 2007). • Video modeling has a high degree of practicality based on consistency, ease of use, and availability of technology (Delano, 2007). • A variety of model types have been used effectively in video modeling (Delano, 2007; Shukla-Mehta et al., 2010). • Some skills are suggested to facilitate greater success with video modeling (e.g., imitation, ability to attend) (Delano, 2007; Shukla-Mehta et al., 2010). • Repeated viewing of video models < 5 min in length is suggested to produce the greatest effects (Shukla-Mehta et al., 2010).	• For whom is video modeling best suited and how can success be best predicted? • How can we predict what model type is best suited to an individual? • What aspects of video modeling or accompanying strategies best promote generalization and maintenance of skills? • What are the effects of video modeling in isolation and when combined with a variety of intervention approaches? • What effects do narration, verbal prompts, or voice-overs have when they are incorporated into the video model? • Are video models more effective when the target behavior is presented as a whole or in segments (i.e., a series of videos)? • What are the effects of video modeling on adolescents and adults with ASD? • What are the critical components of video modeling (e.g., related to content, length, model type, number of models/viewings)? • What other skills and behaviors could video modeling impact (e.g., problem-solving behaviors)?

Digital video modeling has been found to have numerous benefits, even beyond live modeling. In addition to taking "approximately one-third of the time with one-half the cost when compared to in vivo procedures" (Corbett, 2005) video modeling is purported to be more effective than live modeling. Charlop-Christy's results suggest that video modeling in fact leads to more rapid learning and subsequent adoption of tasks and behaviors than live modeling, and furthermore, it was effective in promoting generalization. The idea supporting this theory is that video modeling can produce a greater breadth of settings that are more challenging to replicate in real life, or a clinic or classroom setting. Furthermore, the instructor has greater control over the modeling, as they can adjust models, sound, quality, lighting, and re-record until they are satisfied with the content. Another benefit is that children can watch the same video as many times as necessary. Repeating the exact same video removes irregularities and distractions and doesn't require repeated exertion from the model. Dowrick, among the first to prove the efficacy of video modeling, finds that some of video modeling's success can be attributed to it being more stimulating than live modeling; it is a change from the habitual work environment, and the association video has with recreation further entices and motivates the child (Dowrick, 1986). This higher salience due to the intrinsic motivation of digital stimuli is corroborated by Puca & Schmalt (1999), and Bandura in his Social Learning Theory.

The method appears to be especially effective for children with ASD (Autism Spectrum Disorder). Numerous decades of scholarship support video modeling's capacity to help children on the spectrum target a breadth of skills and behaviors across language, social behavior, play, academics, and adaptive skills, including, but not limited to expressive labeling, independent play, spontaneous greetings, oral comprehension, conversational speech, cooperative and social play, and self-help skills (Corbett, 2005; Bellini & Akullian, 2007; Hitchcock, Dowrick, & Prater, 2003, Wilson, 2012, Charlop-Christy, 2000). It has been demonstrated that people with autism show strengths in processing visual information over verbal, which makes visual cues a particularly effective form of instruction (Quill, 1997; DeMyer et al., 1974; Shah & Frith, 1983; Happe 1994a; Freeman et al., 1985; Asarnow et al., 1987; Lincoln et al., 1988; Corbett, 2005). Video modeling has the benefit of being able to focus in on details and eliminate distractions. This can be especially beneficial for children with ASD since it limits the stimuli they need to focus on, thereby limiting their stimulus overselectivity. Furthermore, the videos can showcase social interaction and behaviors without the actual stress or demands of in-person social behavior. Learning from a video allows them to concentrate on the content without the stressors of social interaction that inherently come with learning from a live model (Charlop-Christy, 2000). Furthermore, live modeling may even make generalization more difficult, since the child and the model are physically part of an environment that binds the behavior in ways that the digital realm does not (Stokes & Baer, 1977; Charlop-Christy, 2000). A 1989 study by Charlop and Milstein used video modeling to teach autistic children conversational speech skills; after watching the modeling videos, they found all children to rapidly learn speech skills which were furthermore maintained at the 15-month follow-up. In essence, video modeling proves to be more effective than live modeling for children with ASD because of its systemic and simple format, capacity to gain and keep attention, and natural predisposition to suit the social needs of children on the spectrum. However, these findings can also be generalized beyond autism. Speech Blubs' primary audience (children aged around 1-3) tend to be shy; new social experiences can be experienced as stressful, not unlike children with autism. It is for this reason that speech therapists who specialize in this age group often work with parents; it is very tricky to gain the trust and cooperation of a toddler.

Speech Blubs, which uses predominantly video modeling to convey the app's content, keeps the strengths of this medium at the forefront of their design. The video modeling is made to be simple and clear; the user gets a large-format view of the teacher's face on a high-contrast, low-distraction background, where they hear the same word repeated multiple times in two different paces (one to target phonemes, one to target the overall pronunciation). The organization of the exercises is consistent throughout the app, once again minimizing distraction and maximizing ease of understanding and efficient concept adhesion. The modeling is supported by visual cues for the target word through images, graphics, sounds, and acting out.

Video Modeling with Children, for Children

While it is logistically much simpler to find adult actors, Speech Blubs specifically sought out child actors to perform the video modeling, since it is more effective for children to learn from people that look and sound as close to them as possible. Smith (2010) confirms that the characteristics of the model impact the observer's absorption of the target content. Generally speaking, one is more likely to imitate someone more similar to themselves, or whom they perceive as more similar to themselves, than someone they can identify with less (cf. Bandura, 1977; Kazdin, 2001). This finding proves to hold especially true of children learning from other children their age (Kelly, 1982). As has been discussed, video modeling came out of the mimicry and modeling literature. Dowrick, a pioneer in the field of video modeling, in fact proved the efficacy of the method with children, including those with spina bifida, hyperactivity, and developmental disabilities (Dowrick, 1991; Dowrick & Dove, 1980; Dowrick & Raeburn, 1977; Dowrick & Raeburn, 1995; Charlop-Christy, 2000). Meanwhile, Egel (1981) successfully used live modeling to teach autistic children various discrimination tasks, and Tryon and Keane (1986) promoted imitative, independent play for children with autism using peer modeling. Evidently, a plethora of research supports peer modeling as a worthwhile intervention for teaching a variety of skills to children, perhaps notably those with ASD.

Furthermore, Speech Blubs allows for specific personalization, with your child's picture or avatar being used in the app so they can see themselves learning. The child follows themselves through every exercise, and watches their character interact with teachers and the game. This high degree of personalization proves helpful not only in retaining child interest, but also in modeling target skills and behaviors.

(2) IN-APP CONTENT

One of the most critical questions is what content is actually in the Speech Blubs App. This section will runt through a couple of key content items and mechanics, such as repetition, animal sounds, positive reinforcement, feeling identification, informational videos, song sections, personalization, and explain the benefits that each offers. The app seeks to provide interactive and instructive learning opportunities that feel social, instead of academic. To give a little bit of an idea, ASHA, the American Speech–Language–Hearing Association, provides guidelines on how one should interact with children in order to best teach them speech. Some of their tips include pointing out colors and shapes, using gestures, modeling good speech, speaking clearly, cutting out pictures of favorite or familiar things and putting them into categories (eg, things to ride on, things to eat, and things to play with), or making silly images by mixing and matching pictures, naming body parts while identifying what they do, and explaining uses for familiar objects (ASHA).

Every single one of these interventions can be found in the Speech Blubs app. Speech Blubs child models clearly identify and model a wide variety of words, including shapes, colors, animals, feelings, relatives, toys, numbers, among many others. The children, still peers of the target demographic but slightly older in order to ensure better pronunciation, were vetted by professional speech and language pathologists in order to ensure pronunciation correctness and verbal clarity. Furthermore, the child teachers pantomime emotions, animals, and other appropriate target words. In terms of the overall app, it is almost exactly the digital version of ASHA's cutout recommendation, as it collects "stickerbooks" of familiar and favorite words (including vehicles, food, toys, and body parts) and presents them in categories. Each exercise within those categories begins with a video model of the pronunciation, then most go on to show silly images where children can see themselves with a face filter on, that is related to the topic they are learning about. The exercises also contain informational videos that contextualize the target word, either providing a fact, asking a question, making an observation, or some combination of the latter.

Repetition

The idea that repetition aids learning is far from novel. The child's brain is born receptive to knowledge, synapses are formed, and the most heavily-used are strengthened; the strongest ones will remain, and the weaker will be lost. Parents, as the source of some of the primary interactions babies have, have a colossal impact on their child's early brain development. By the very nature of synapses, repeated use will strengthen them. Ergo, repeated exposure, visualization, and repetition will strengthen a child's retention and ability to produce and store information on a word (Queensland Government, 2020). A 2011 study by Horst showed that 3-year-old children learned more new words from shared storybook reading when the same three storybooks were read repeatedly, as opposed to when multiple different storybooks were read (Horst, 2013). It appears that repetition helped solidify the novel words. Furthermore, repeating the same words appeared to allow children to understand them in different contexts, and develop a deeper, more robust understanding of the word. Not only were they able to know the target word more reliably, but they also understood it more thoroughly. Horst finds that contextual repetition helps lower the attentional demands of word learning, meaning that increased familiarity with the context of the word helps children pick up on it better, which in turn allows for better analysis and internalization (Horst, 2013). The first step in "learning" a word is connecting the sound to what it refers to. Next, children are able to make that connection in different contexts, and it is after repeated recognition that they begin to produce the word themselves. Naturally, with increased repetition comes a more robust understanding of the word. Getting to this stage of word mastery is an iterative process, that relies heavily on repetition (Yu and Smith, 2007; Horst and Samuelson, 2008; McMurray et al., 2012; Horst, 2013). Dodge (2009) finds that children need repetition and reinforcement in order to feel safe and confident. The familitariy is in fact a source of comfort and self-esteem as well as it is a pedagogical tool.

Each exercise in the Speech Blubs app repeats the target word extensively and in different contexts; either just repeated exposure, in a recognition quiz, or with an informational video. The app also doesn't remove completed exercises, rather keeps them open and available for repetition. Much of the exercise layout isn't ranked, or organized in an order of progression, rather, it is an open plan that encourages exploration, play, and easily being able to find the words a child would want to go back to, or a parent would like to work on.

Animal Sounds

Animal sounds have long been a method to get children to start producing phonemes that grow into building blocks for words. Some of the American Speech–Language–Hearing Association's first recommendations for new parents are to say sounds like "ma," "da," and "ba" and to try and get the baby to repeat them back. They also recommend that babies be taught about animal sounds and connect the sounds to animals. This not only aids sound-production, but also learning about new concepts, such as animals (ASHA). Adamson (2019) finds that while many children as young as 12 months old will be alert to and try to replicate novel sounds to their parents, the childrens' sound production skyrocketed when the novel sounds were repeated back at them.

Speech Blubs' early sounds section does just this. It features children repeatedly mimicking an animal sound while pantomiming the animal. Next, the child is prompted to either repeat the sound or match it to the animal that makes it, before being shown an informational video about said same animal. The initial repetition of the novel sound, paired with a call-and-response repetition of the animal phonemes which is further reinforced with visual confirmation indeed hits all the markers of a maximally-beneficial animal sound section.

Positive Reinforcement

Positive reinforcement is baked into every step of the exercises in Speech Blubs. Studies show that warm and accepting interactions with children engendered reciprocity and engagement on the part of the child (Westerlund and Lagerberg, 2008; Safwat, 2014). At its simplest, positive reinforcement is just giving attention to desired behavior while attempting to ignore inappropriate behavior. Furthermore, the most effective positive reinforcers are those that are quick and easily attainable (Sigler, 2005). In the context of the Speech Blubs app, they do not attempt to punish or ignore inappropriate behavior, rather, they reward any interaction with the app (Sesame Street, 2012). This taps into the warmth and acceptance that encourages child participation, which then hopes to encourage sound and word production. The app is an environment where the child cannot be scolded or criticized, but instead can be enriched and encouraged to start speaking.

Feeling Identification

Identifying and recognizing feelings in oneself and others is an important part of growing up and socializing. It is also something many children struggle with, in particular, children with Autism Spectrum Disorder (CDC, 2022). Studies have used "social robots" or simplified robots that demonstrate what emotions look like. Pop (2013) found that the robots were successful in helping children with ASD recognize emotions. She further finds there to be advantages to using robots. To quote directly from the paper,

> "(i) the embodiment of the robot is offering human-like social cues and, at the same time, it keeps its object-like simplicity; (ii) robots have the capacity to gradually increase the complexity of the tasks, and (iii) they are predictable, controllable, and enable errors to be made safely, thus giving the possibilities to train a wide range of social and communication behaviors. A number of research groups have examined the responses of children with ASD to several types of robots. Robots may act as a mediator between the child and the therapist, and can be used in play therapy and can elicit interaction and joint attention episodes between a child and an adult" (Pop, 2013).

Pioggia (2007) discusses a different social robot, FACE, who is meant to resemble the user as much as possible. He finds that using a visualization that more closely resembles the user has better results (Pioggia, 2007). In his study, he also finds that after his users interacted with the social robot, they demonstrated a decrease in the CARS (Child Autism Rating Scale) emotional response score of between 1 and 0.5 points, knowing that the scale is out of 4. This body of research is representative of the growing value and validity of RAT (Robot Assisted Therapy) to use situation-based emotions in order to help the ASD children to understand the social context in which emotion may appear. This effort extends beyond robots, however. *The Transporters*, a show designed to help children with autism learn to recognize and understand emotions, has some of their success attributed to the fact that children learn from real faces on an intrinsically motivating medium (Baron-Cohen, 2009).

Interestingly, this is the same method Speech Blubs uses to teach children how to recognize and understand emotions, both in themselves and others. The feelings are laid out in a grid format. Once an exercise is selected, it will show a video of a social story followed by a video of a child model (or multiple successive child models) face- and body-acting the feeling. Users are then prompted to match the mimicry and name of the emotion to a picture of a different child expressing it. Upon completing that exercise, another social story is played, and the original child mimics the feeling once more.

Informational Videos

One of the first issues that arose in digital gamified interventions for children is that apps would intersperse games between therapy sessions instead of integrating learning into games (Malinverni, 2016). This holds true for speech interventions, where it is critical to embed the learning into gamified content in order to engage deeper retention (this will be further discussed in the following section). In order to engage that deep learning, it is critical to understand the user, in the instance of Speech Blubs, a child typically under 5 years old. Children of this age cannot read, so clear pictures and icons are critical, however, while these children can see in three dimensions, they cannot yet visualize a 3D experience on a 2D screen. This ability typically doesn't develop until age five (Gelman, 2014). These data tell us that users need a highly simplified, literal interface. For ease of understanding, designers should focus on using color and detail for more important elements in front, whereas simple shapes and muted colors should suffice for background, or mere contextual information (Gelman, 2014). Furthermore, in these types of visually-optimized experiences that combine therapy with pleasure, sound needs to be engineered to communicate, inform, and instruct, as opposed to simply entertain (Gelman, 2014).

It is with these principles in mind that Speech Blubs came up with their information videos. Short, instructional videos with simple demonstrative graphics that reinforce meaning and context for the target word, all while creating more exposure. The videos also ask questions while providing further information, hitting upon the ASHA suggestion to reinforce concepts via question asking and answering (ASHA). They are crafted to succinctly elaborate on what a word means while entertaining the child watching it through simple but dynamic visuals, and clear and concise script. The length of the videos is limited to a couple dozen seconds in order to keep engagement alive and reduce chances of losing the child to complacent watching. The information comes right after an engaging activity, and is followed up by prompts for more engagement. With these actualizations of the ideals outlined in the literature, the informational videos in Speech Blubs are crafted to help children boost language skills without them even being able to notice they are learning.

Song Sections

Singing is important for children's development in many ways. Exposure to song and instruments shows children cause-and-effect logic, reasoning, patterns, numbers, changes in pitch and sound, and new words. The underlying goal that connects all of these benefits is pattern recognition (Dodge, 2009). Human beings are hardwired to recognize patterns and make deductions from them (Salimpoor, 2013), especially in song, as it is one of humanity's oldest ways to pass down generational knowledge (Masataka, 1999). It's no coincidence that "parent speak" or "motherese" is composed of all the basic building blocks of music (intervals, melody, simple songs and singing...) and is very consistent across cultures (Welch, 2001).

Something about singing really sticks with humans. Perhaps it's the physical, soothing response humans have to it. Dr. Sandra Trehub shows that singing calms babies for twice as long as talking, even when they don't understand the words (or recognize the voice singing them). The rhythms and melodies are soothing in ways that speech isn't (Trehub, 2015). In another study, Trehub also found that babies with high levels of cortisol (a hormone affiliated with stress and arousal) had those levels reduce when they listened to singing. Conversely, babies with lower cortisol levels had their cortisol levels increase when listening to singing. Because the babies in the study all appeared content during the experiment, Trehub concluded that singing has a modulating effect on cortisol, keeping it in just the right range for positive attention. This evidence suggests that singing is most useful for helping babies regulate emotion and focus attention (Suttie, 2016). Now that we have established how the medium primes children for positive emotions and attention-paying, what do children actually stand to learn from songs? Songs can improve language skills, helping children increase vocabulary, phonological awareness, and reinforcing known words. They also promote math and counting skills by introducing rhythm, count, and beats.

Although speech and musical skills develop independently they can be used to strengthen each other; musical and linguistic skills have been shown to be linked in young preschoolers, regardless of their non-verbal abilities and verbal memory. The timing and melodic skills they possess are associated with language development, and are not just by-products of other cognitive skills and processes. It is not surprising then, that the rhythmic aspect of musical ability predicts phonological awareness and the melodic aspect of musical ability predicts language grammar (Politimou, 2019). Indeed, the highly organized, rhythmic, and logical structure of music mimics the universal ideas that underlie human language, and primes babies to understand what exactly is coming out of their parents' mouths, and how to dissect it. The same study responsible for the aforementioned finding also reveals that informal music exposure at home can positively impact a 3-4 year old's complex language skills. This is not to say that parents who don't play music for their children have locked them into a life of failure and parents who do are raising future CEOs, neurosurgeons, and world leaders — the Mozart effect has been thoroughly disproven. This finding just suggests that increased exposure to music and musical experiences helps to familiarize oneself with language, help with sound production and fluency, and strengthen connections between musical and linguistic skills at a crucial developmental stage (Politimou, 2019).

This information yields that successful kids' songs will have repetition, games, and funny sounds and lyrics. These elements draw children in to catch their attention, and then promote language development skills, gross and fine motor skills, and most importantly, fun.There are five pillars to attractive music — catchiness, adaptability, repetition, relatability and transgression — and they're the same for humans of all ages (16). Combining pleasure with pedagogy is the key to tapping into children's millennia of hardwiring (Dodge, 2009; Underwood, 2020).

The key to teaching through song is repetition (Dodge, 2009). In a private meeting in 2021, Sandra Trehub described how children that are not yet speaking may not understand that lyrics are composed of meaningful units — words — instead, they intake it as highly pleasant repetitive music. From listening, they're getting melody, intonation, rhythm, and words, not grammar and the nitty-gritty of language. That's what makes the medium so accessible to a wide array of children. Indeed, some of the first signing kids do is completing the last word of a line being sung to them. For kids to derive some meaning from the lyrics, they need to be very explicitly shown what word maps on to what real-life counterpart. The song itself needs to meet certain criteria for it to impact a child. First, the song must be highly familiar to the child — they connect with the song more than the singer or the instrumentals, so no matter how it's presented, children will love hearing a voice singing a pattern they know. Kids really respond to the predictability and familiarity of a song they know, so when they correctly guess what is coming next, it gives them a real sense of joy.

Speech Blubs' singing section was designed with these principles in mind. The joint-singing model is meant to encourage and leave space for children to participate, while the graphics allow children to understand context and connect new words with their meanings (not to mention reinforce words they already know). The child models sing familiar, repetitive songs in happy, approachable voices that children can relate to and feel comforted by. The videos are under a minute long, provide opportunities for children to see themselves and sing along, repeat the song multiple times, and show visualizations (with lyrics for the older children) in order to solidify the understanding. Speech Blubs understands that participating in a sing-along does not look the same for all children. Children that aren't speaking or are still unfamiliar with the song might stare, dance along, babble, or hum, whereas speakers might start using words and phrases. The app does not punish or reward one behavior over another.

(3) GAMIFICATION

A key part of Speech Blubs' design is its gamelike mechanics and interface. Intentionally, it is not a textbook, parent manual, or video playlist that will run on uninterrupted, rather, a collection of gamified exercises, interspersed with mini-games. In order to understand why this was the elected method, it is critical to understand how children learn, and how to best promote that kind of learning.

Puzzles and Riddles

Anybody who has spent time around infants and toddlers will be able to confirm that they hardly if ever pick up a textbook or attend a lecture in order to learn a concept, let alone language. Their methods of learning are altogether different. The AAP, American Academy of Pediatrics, maintains that children learn best through "observation and active engagement" rather than "passive memorization or direct instruction" (Yogman, 2018). They recommend children learn from guided play; defined as an activity that the child initiates, but that falls within a setting delineated by an adult, in order to address a learning goal. Examples might include a children's museum exhibit, a Montessori task, or the Speech Blubs app). Guided play can also include adults supplementing exploration with guiding questions and comments in order to maximize learning, without limiting the child's joy or pleasure (Yogman, 2018). ASHA, the American Speech–Language–Hearing Association, provides myriad guided play suggestions for how to teach babies speech, primarily focused on question asking and answering. Suggestions include: "Play the yes–no game. Ask questions such as, *Are you Marty?* and *Can a pig fly?* Have your child make up questions and try to fool you. Ask questions that include a choice. *Do you want an apple or an orange? Do you want to wear your red shirt or your blue shirt?*" (ASHA). While there is yet to be consensus on the definition of play, there is mounting agreement that play can only be “intrinsically motivated, entails active engagement, and results in joyful discovery” (Yogman, 2018). These expert findings and recommendations all highlight the importance of child-motivated problem-solving.

It is for this reason that Speech Blubs is organized in such a way as to promote discovery. The app’s colorful and icon-rich interface does not lead a child down a prescribed path (unless that’s the mode of progression a parent chooses for children not quite at the independent exploration phase), rather presents them with animals, topics of interest, and enticing, recognizable images that spark curiosity. The exercises include riddles, questions and answers, recognition quizzes, and prompt active engagement, curiosity, and ultimately learning. The breadth of topics and modes of engagement seek to allow children to explore, encounter novelty, and reaffirm the familiar in order to use their learning both in and outside of the app. The gamified element, central to pleasure, and intrinsic motivation, is the key to lifelong learning that Speech Blubs looks to foster with its design.

Ludic Learning

The efficacy of intrinsically-motivated pleasurable education can be explained by the concept of ludic learning. Ludic learning is quite simply a pedagogy that uses games to teach. The AAP confirms in a widely-cited study that “Play is not frivolous: it enhances brain structure and function and promotes executive function (ie, the process of learning, rather than the content), which allow us to pursue goals and ignore distractions” (Yogman, 2018). In fact, the app looks to break the false dichotomy between play and formal learning by showing how in fact, pedagogies that are enjoyable and entertaining can yield the best results. Active engagement and engrossment in an activity are purported to build “executive functioning skills and [contribute] to school readiness”. The literature is increasingly impassioned about the inextricable link between play and learning, citing heightened curiosity as the main catalyst for memory, learning, collaboration, and problem-solving skills that ludic learning frameworks spark in children. Indeed, these are skills and neural networks that children will not only need today and tomorrow, but in all likelihood skills that they will rely on and build upon for the rest of their lives (Yogman, 2018). However, not all games are built the same. An exhaustive literature review yields that there are a collection of key game elements essential to educational game design, such as narrative context, rules, goals, rewards, multisensory cues, and interactivity, seem necessary to stimulate desired learning outcomes (Dondlinger, 2007).

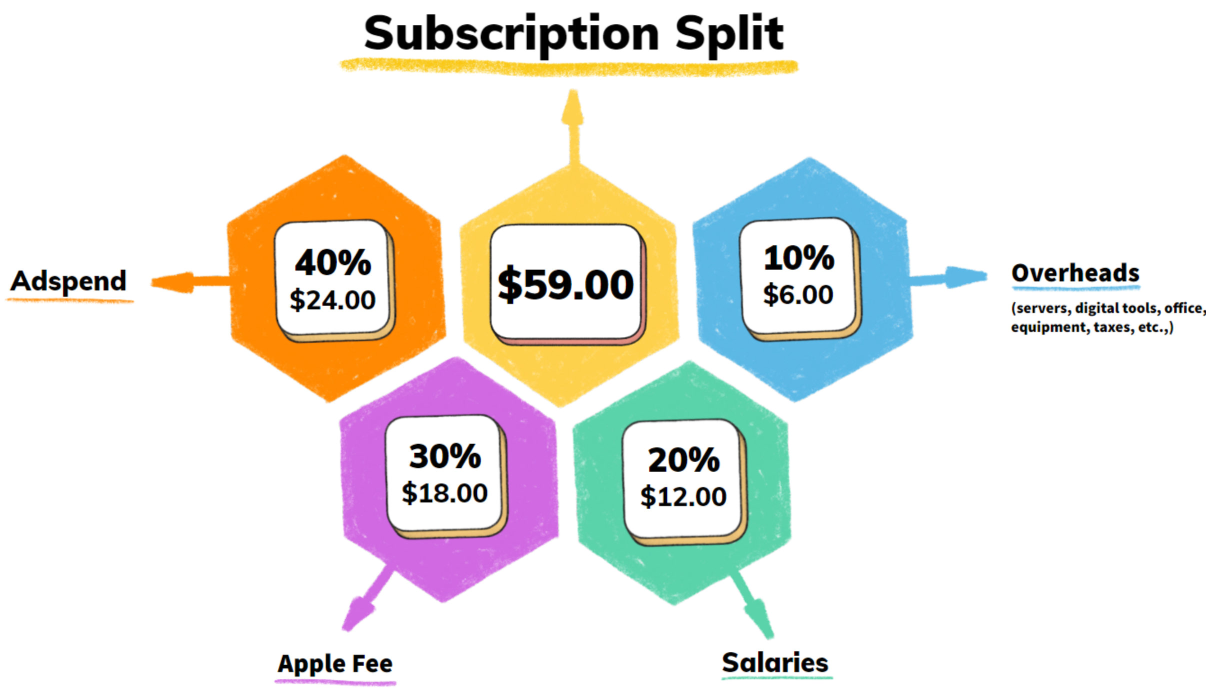
While Speech Blubs isn’t a video game in the traditional sense, it can nonetheless use the learnings from educational video game design to assess how supportive it is of learning. The app is laden with reward. The children get ample positive reinforcement for working on their speech at all, and are praised for any progress or effort they can make. Speech Blubs will never penalize a child not meeting milestones, not trying hard enough, or otherwise not meeting a performance indicator. Instead, there are different gradarions of reward for the learning goal, and incorrect responses are treated as a step in the pedagogical process; something that can be learned from. The app gives multisensory cues: visual, audio, and physical to stimulate interest, entice engagement, and reaffirm feedback. A careful interaction of the three modes of stimuli seek to draw the user's attention to different elements of the app usage experience, from educational content connected to their real-world experiences, engagement prompts, and reward and reinforcement of the user's behaviors. Each exercise is densely packed with coded information and feedback that promotes prolonged interest and attention span in the educational content. Speech Blubs is deeply aligned with the ludic learning pedagogy; all of the design elements, from loading screens, to color palette, vibrations, and sound choice are made to encode therapeutic information and skills in a fun, gamified experience. A proverbial spoonful of sugar to help the medicine go down, the gamified Speech Blubs experience is designed to make educational benefits shine.

(4) STRUCTURE AND SUPPORT

This section seeks to cover the most overarching elements of the app's design: general design and usability, and parental support. Arguably one of the most important parts of helping children learn speech is the community and support they have access to. No child can learn speech in a vacuum, so their loved ones and the resources those grown-ups have at their disposal are the tools a child counts on. Speech Blubs knows this, and strives to support parents, guardians, and all grown-ups with helping their little ones talk. One of the simplest and most effective ways Speech Blubs does this is with robust guidance for parents in the form of blogs, resources, and in-app support. Another, perhaps less obvious, way of supporting parents and speech communities is the app's structuring to be "smart screen time", or a positive, productive use of digital technologies. This includes specializing in design for children, with a focus on therapeutic benefit.

So, where does my money go?

Speech Blubs is a subscription-based service. This is how the expenses break down:



Is it what you expected? While $60 a year may come as a surprise to some customers, the monthly subscription cost is carefully calculated to cover everything we need to keep the vocabulary coming every month. Speech Blubs' commitment to quality and efficiency try and keep the app as affordable as possible, which is why we love to share data like this, so you can see we put our money where our mouth is.

General design and useability

The general design and usability is a wide-ranging concept that hinges on creating a hands-on expiration experience made to be structured, systematic, and intuitive. It should engender emotions of fun and enjoyment while providing doses of education seamlessly integrated into the play design, not served up alongside it. Children should be able to feel welcomed and led through the game (hence why they are themselves featured in the app alongside child teachers that are their peers). Ideally, this creates a feeling of dialogue and identification between the child and the app (Sesame Street, 2012). These children are just beginning to use technology, and therefore do not use it the way grown-ups do. Extra considerations are required in order to ensure that an app meets their needs, both for use and for learning (Gelman, 2014). Oftentimes, shortcomings come from the lack of user involvement in design, something Speech Blubs looks to overcome by frequently user- and expert-testing the application (Malinverni, 2016). Indeed, Malinverni helped come up with a model to integrate the expertise of clinicians, contributions of children and experience of designers in order to design playful experiences that are effective in terms of therapeutic objectives and are enjoyable and engaging for children. The methodology follows four steps:

(1) The elicitation of requirements from experts aimed toward properly defining the therapeutic goals, the structure of the experience and the therapeutic techniques (2) The elicitation of contributions from children aimed toward identifying their interests, motivations and preferences (3) The integration of contributions from experts and children in order to define the mechanics of the game, its elements and the experience as a whole (4) The exploratory evaluation of game suitability with children (Malinverni, 2016). Indeed, this is the exact methodology used by the developers of the Speech Blubs app in order to ensure it's value, quality, and enjoyment by the target audience.

The general design and usability can be broken into the following subsections: interactive design, screen design, text design, visual layout, visual design, and audio design. Each will be elaborated on below.

According to Sesame Street's 2012 workshop on making digital content that benefits children, good interactive design will begin with greetings. In Speech Blubs, the child is welcomed by Finny, an anthropomorphized fish that guides the child through speech activities. The children are then supposed to receive instructions that state the objective, as well as how to accomplish it. In Speech Blubs, child teachers model, give instructions that state the objective as well as how to achieve it, and nudge children towards the correct answer if the app detects a couple seconds of inactivity. This period of inactivity is referred to as a "time out", and usually indicate that a child is having trouble completing an exercise. In this instance, the app has built-in support to support learning and perseverance, whereas a child may have given up without the extra support. Sometimes, a child will get a question wrong. The Speech Blubs app follows the design guidelines that dictate gentle feedback; a first wrong answer will be met with a phrase that identifies the wrong choice and offers encouragement. The second wrong answer will either repeat the feedback from before, or identify the wrong choice, restate the objective, offer a hint, and provide encouragement. In either case children are never scolded or punished. Wrong answers are seen as a learning opportunity no less valuable than a correct answer. Positive reinforcement and payoffs are very important to children, as they keep them motivated and interested. These celebrations can be verbal, visual, or auditory, but are always included in the Speech Blubs experience. Even if there are no explicitly right or wrong answers, users get encouragement and reaction to input. At the end of each exercise, users are brought back to the activity selection screen from where they can choose to repeat or to try something new (Sesame Street, 2012).

Screen design for children must be conscientious of the limitations of the users, in terms of dexterity, literacy, and visual metaphor. Children under 5 cannot easily tell what the important parts of an interface are; they click everywhere to see what happens, and interpret that exploration and uncertainty as part of the game (Gelman, 2014). It is then in the designer's best interest to create a strong visual distinction (e.g., color, line weight, art style) between what can and cannot be interacted with – navigation should look clickable, but not be too interesting as to detract attention from the actual focus of the activity (Gelman, 2014; Sesame Street, 2012). In Speech Blubs, navigation buttons are big, colorful, and attractive, but stay still unless they must be pressed. When they become the critical next interaction, then they take on attention-catching characteristics so that the child avoids being stuck in an exercise. For children, hierarchy must be visually explicit in ways that adults do not need, since visual filtration skills do not mature until later. Gesturally, children can best intuit tapping, drawing, swiping, dragging, and sliding, whereas they struggle most with pinching, tilting/shaking, multi-touch, flicking, and double-tapping (Sesame Street, 2012). This can help designers such as those behind Speech Blubs to ensure smooth user interactions and minimize technical hangups, while also providing a little more protection for sections children shouldn't be able to access, such as the parent section, which requires multiple complex taps to get to. In the children's section, activities only ask for large broad gestures or single taps, and all on-screen elements are sized-up so that children can easily manipulate what they are meant to. The app is mindful of little ones (especially the tiniest among them) not yet navigating with one finger, so allows for whole-hand input and imprecise selection. Furthermore, the navigation buttons are on the bottom left and right corners of the interface, relying on placement more than iconicity to convey visual metaphor (Gelman, 2014). This way, children are unlikely to bump them by accident, and can more easily understand metaphorical progression. Inside the exercises themselves, the goal of gameplay is made visually explicit. The back and forth between teacher and child is accompanied by visual reminders of the target word, and visual reinforcement when the target action is accomplished. The visuals are accompanied by verbal statements of the goal. Any "hot spots" a child must touch are large, and spatially distinct from other hot spots. Zooming out, the app is organized in "stickerbook" sections by theme for easy comprehension and navigation. Not everything is visible on one screen, so border elements peek in from the edges of the screen, implying that more is discoverable by scrolling. One final screen-type that was not yet discussed was the camera screen. In Speech Blubs, children are frequently given the opportunity to see themselves and take a selfie complete with on-theme face filters. In order for this operation to be successful, children need to coordinate holding the device with two hands (or propping it up), aiming it at themselves, and triggering the capture button (Sesame Street, 2012). To solve this, Speech Blubs's capture button is in the middle of the screen, and the face filters follows the child as long as they are on screen. This makes it much easier for the child to capture an image of themselves.

Text on screen must be used sparingly, knowing that the majority of children in this age group cannot read. Design experiences that include text are presumed to be used with adults, or to teach early literacy skills. Fonts should be simple and clear (such as Zaner-Bloser and similar), and steer clear of serifs (Sesame Street, 2012). Speech Blubs primarily uses text for adult-facing content and instructions. The sentence-building section and the sing-along section are two parts of the app where users can find text. Both of these sections use high-contrast text that approximates Zaner-Bloser as much as possible. Any branded lettering is kept colorful, simple, and outside of the exercises aimed at pre-literate children.

The visual layout of the app should adhere to certain principles that help ease of use. The menu, or a way to exit the exercise, should be available at all points of the exercise. In Speech Blubs, there is a subtle exit button always available in the top left corner of the interface, cleared from accidental bumps. The exit button also requires confirmation of leaving, presenting the user with a home icon and a return arrow, so even children can make the decision to stay or return. The content is designed to be scanned from left to right and top to bottom, the way that is most intuitive for western children (who make up the vast majority of users). A more technical issue concerning the visual layout is the edges of the screen; oftentimes, children rest their wrists on the edges of tablets, so if hot buttons are placed there, they can easily be triggered on accident (Sesame Street, 2012). Speech Blubs keeps hot buttons clear from these areas that are often subject to accidental contact.

The visual design is closely related to the visual layout, but includes more invisible concerns. Once again, visual instructions are instrumental to success, this includes highlighting, wiggling, and other forms of visual indication that were elaborated on earlier in this section. Critically, the visual components of instruction complement the auditory ones, in order to optimize understanding. These indications extend to interactivity indicators; on-screen cues such as wiggling, enlargement, and a pointing hand all help children understand that they are supposed to interact with the part of the app the audio feedback is telling them to touch. With this differentiation, hot buttons are distinguished from “cold” content that doesn’t change size or wiggle, doesn’t get pointed to, and doesn’t stand out in terms of color and graphic design. Conveniently, icons stay the same throughout the app, be it progression buttons, or exercise prompts and layouts. This way, children find themselves in more and more situations where they know what is expected of them (Sesame Street, 2012). This allows them to focus on learning instead of being mentally taxed by navigation.

Audio design was touched upon above, but merits its own section. One of the most important aspects of audio design is concision. Children typically do not pay attention to audio instructions alone, so keeping the essential audio (e.g., specific task or game instructions) minimal and precise gives it the best chance of being useful. Furthermore, designers should put the specific instructions at the end of the sentence, instead of at the beginning. Combining the primary audio with any visual support possible, as has been illustrated above, is the preferred way to reinforce audio instruction. Another area where extra precision is rewarded is in describing hand motions: tapping, pinching, and sliding are not self-explanatory, and should be explained plainly as well as reinforced visually. When it comes to longer audio, as is present in Speech Blubs, it is recommended that the audio be interruptible. However, the very point of Speech Blubs is to expose children to specifically-crafted audio inputs in order to boost expressive and receptive language skills, so it does the child a disservice if they can interrupt it. Of course, the audio will not play along forever; a certain amount of audio is required to play, and any subsequent repetition is interruptible and voluntary. Other audio output, such as sound effects, are often triggered by a child's touch. Touching a "cold" spot will return a simple sound, whereas children can "tickle" teachers and themselves and hear giggles. The sounds are consistent throughout the app experience. Another instance of sound output is when children get an answer wrong. In this instance, they are met with a sound from a small selection of sounds, including but not limited to flatulence, a weak horn, and "oh no!". Sound is only ever used in Speech Blubs when it is the direct form of therapy, reinforces instruction, or serves some other explicit purpose in the app. Noticeably, Speech Blubs does not use background music. Background music can enhance engagement for a child, however, it can also detract from the actual gameplay (Sesame Street, 2012). In this particular instance, total auditory focus on the gameplay is of the utmost importance, as that is where the therapeutic value lines.

These sections all detail how Speech Blubs sought to optimize its therapeutic delivery. Ultimately, the best apps will promote exploration and discovery, have flow but allow for non-linear play, and provide rewarding visual and auditory feedback with every interaction, after clear and concise cues and directions (Gelman, 2014). Gelman offers a succinct table illustrating key particularities that users the age of Speech Blubs' primary demographic that the creators of Speech Blubs factored into every design decision and concept creation. The table is attached below:

2–4-year-olds...	This means that...	You'll want to...
Focus on details instead of the "big picture."	They can't distinguish main elements of an interface from the details.	Create a very clear visual distinction between interactive items and design extras.
Can rank items by only one characteristic at a time (i.e., color, shape, and so on).	They get overwhelmed when there are too many variables competing for their attention.	Pick a smallish set of easily identifiable elements (like colors) and use them consistently throughout your design.
Can only associate a single function with an item or object.	If an item expands or makes a sound on rollover, they'll believe that's the sole purpose of that item and won't know to click on it.	Limit the behavior of your navigation elements to navigation (for example, don't have them pop up or make noise).
Can only see items on a screen in two dimensions, not three.	Everything on a screen looks like it's in a single, flat plane to them.	Make your foreground items much clearer and more detailed than stuff in the background.
Are just learning to think abstractly.	They are unable to understand icons and symbols that are second nature to adults.	Use icons that are highly representative of the task you're trying to communicate.
Use sound to identify items in their environment.	They get confused when different sounds have different meanings (for example, a police siren and an ambulance siren).	Make sure that every sound you use has a specific meaning and function.
Are starting to develop their own identity.	They develop a sense of self at around age 2, complete with gender identity, which forms very early.	Create a design that allows for gender identification without forcing kids down a specific gender path.

(Gelman, 2014)

Guidance for Parents

One of the first resources parents come across is the milestone assessment; a tool iteratively designed by speech therapists, synthesizing and rephasing the best from over 20 milestone assessments for parents to get an idea of where their children are in terms of communication milestones. Broken up into 5 age buckets that target our primary users (between 7 and 49 months), the assessment runs through hearing, pragmatics, play, comprehension, and talking. A whopping 96% of our parents complete this assessment and get the associated report which gives parents professional tips and support in areas where they might want to pay some extra attention. The assessment stays available for parents to retake whenever they wish.

Within the app itself, there is a parent's section, protected, and discreetly tucked away from the child's user interface. This section, the "Parent's Academy" has a help section, SLP videos, tips, milestone breakdowns, activities, and blogs that parents can consult in a streamlined experience (Sesame Street, 2012). The Parents Academy averages around 2,000 views a day with grown-ups enjoying just under a minute of bite-sized information in the average session.

There is also a corresponding Speech Blubs website that is even more popular with parents. Complete with filterable, professionally-written blogs, and more, the website gets around 1,500 views a day, but boasts a 4+ minute average session time. That number reveals that grown ups go to the blog as a resource and like what they see; it's achieving its goal of answering questions, and being there as a trustworthy resource.

Smart Screen Time

Studies show that children are not only more engaged, but can also gain more from smart screen time, or screen time that is carefully-curated, age appropriate, and beneficial to engage in. As was discussed above, video instruction can teach a breadth of skills either as an intervention on its own (MacDonald, Clark, Garrigan, & Vangala, 2005; Nikopoulos & Keenan, 2004b, 2007) or paired with prompting or reinforcement (Keen, Brannigan, & Cuskelly, 2007; Murzynski & Bourret, 2007; Reeve, Reeve, Townsend, & Poulson, 2007; Mehta, 2005). The benefit is not in the sheer presence of the screen, of course, but rather in how it is used. There is a real difference between using evidence-based video practices such as video modeling, and letting children watch any form of recorded video content, and that difference is smart screen time. It is at this point abundantly clear that children require active engagement in order to benefit from any activity, and that digital interventions can lend themselves especially well to that joy, activation, and interest, even more so than in traditional physical interventions, However, that intervention must be intentionally designed, such as with smart screen time. This observation was confirmed by Hoque (2009), who details how smart screen time can engender a heightened state of attention and absorption, which when combined with the sensory and perceptual needs of a particular client, can result in an excellent, active, and productive use of technology.

This is what Speech Blubs has sought to do. With thorough design ideation, extensive expert input, and constant reference to the literature around speech interventions and childhood technology usage, the app seeks to harness technology's potential and drastically minimize adverse effects. Furthermore, it recommends a 10-minute session a handful of times a week, not endless unsupervised play. The content, recommendations, layout, parent advice, and content are all based on scientific findings and methods proven to put children first and help them meet their goals.

CONCLUSION

In conclusion, Speech Blubs is thoroughly researched and heavily founded in scientific research. The design of the app, which was made in consultation with hundreds of Speech and Language pathologists, looks immediately and clearly to research to ensure the validity and efficacy of the app, down to the littlest details.

Video Modeling, the underpinnings of the Speech Blubs app, have been found to be highly effective for teaching language, and can even surpass the abilities of live language teachers. This is especially true of children, amplified further when the child-teacher resembles them, or in some way provides some relatability or familiarity to latch on to, such as with Speech Blubs' ability to personalize the app experience. Inside the app itself, children are exposed to repetition, animal sounds, positive reinforcement, feeling identification, informational videos, and songs, that have all been designed in line with the science to help children start understanding and producing speech and language. This is done through gamification. Critically, Speech Blubs makes their app fun for the target user group, allowing pleasure and enjoyment that capture attention for longer sessions and prime the brain for longer periods of learning. Together, these benefits of playful learning help children reap more benefits than they would have from dryer content. Finally, Speech Blubs heavily uses Speech and Language Pathologists to help provide support to parents, guardians, and families of the children that use the app. They also adhere to smart screen time guidelines, designing the app to be a quality use of time for children, and encouraging minimal usage. It is for this reason that the app's design must be so packed with proven efficacy – the idea is to expose children briefly but consistently and concisely, so they can learn most effectively by having fun with the app.

Works Cited:

- Adamson, L. B., Bakeman, R., Suma, K., & Robins, D. L. (2019). Sharing sounds: The development of auditory joint engagement during early parent–child interaction. Developmental Psychology, 55(12), 2491-2504. doi:https://doi.org/10.1037/dev0000822
- American Speech-Language-Hearing Association (ASHA). “Activities to Encourage Speech and Language Development.” *American Speech-Language-Hearing Association*, American Speech-Language-Hearing Association, https://www.asha.org/public/speech/development/activities-to-encourage-speech-and-language-development/.
- Asarnow JR, Carlson GA, Guthrie D. Coping strategies, self-perceptions, hopelessness, and perceived family environments in depressed and suicidal children. J Consult Clin Psychol. 1987 Jun;55(3):361-6. doi: 10.1037//0022-006x.55.3.361. PMID: 3597949.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. Psychological Review, 84(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191
- Bandura, A., & National Inst of Mental Health. (1986). *Social foundations of thought and action: A social cognitive theory.* Prentice-Hall, Inc.
- Baron-Cohen S, Golan O, Ashwin E. Can emotion recognition be taught to children with autism spectrum conditions? Philos Trans R Soc Lond B Biol Sci. 2009 Dec 12;364(1535):3567-74. doi: 10.1098/rstb.2009.0191. PMID: 19884151; PMCID: PMC2781897.
- Bellini, S., & Akullian, J. (2007). *A Meta-Analysis of Video Modeling and Video Self-Modeling Interventions for Children and Adolescents with Autism Spectrum Disorders. Exceptional Children, 73(3), 264–287.* doi:10.1177/001440290707300301
- CDC. “Signs and Symptoms of Autism Spectrum Disorders.” *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 28 Mar. 2022, https://www.cdc.gov/ncbddd/autism/signs.html.
- Charlop MH, Milstein JP. Teaching autistic children conversational speech using video modeling. J Appl Behav Anal. 1989 Fall;22(3):275-85. doi: 10.1901/jaba.1989.22-275. PMID: 2793634; PMCID: PMC1286179.
- Charlop-Christy, Marjorie H.; Le, Loc; Freeman, Kurt A. (2000). *A Comparison of Video Modeling with In Vivo Modeling for Teaching Children with Autism. , 30(6), 537–552.* doi:10.1023/a:1005635326276
- Corbett, Blythe A, and Maryam Abdullah. “Video Modeling: Why Does It Work for Children with Autism?” *American Psychological Association*, American Psychological Association, https://psycnet.apa.org/fulltext/2014-52006-002.html.
- DeMyer, M.K., Barton, S., Alpern, G.D. *et al.* The measured intelligence of autistic children. *J Autism Dev Disord* **4,** 42–60 (1974). https://doi.org/10.1007/BF02104999

- Dodge, Diane Trister, et al. *The Creative Curriculum for Preschool*. Teaching Strategies, 2009.
- Dondlinger, Mary Jo. "Educational Video Game Design: A Review of the Literature". *Journal of Applied Educational Technology,* vol. 4 no. 1, Spring/Summer 2007, pp. 21-31.
- Dowrick, P. W., & Raeburn, J. M. (1977). Video editing and medication to produce a therapeutic self-model. *Journal of Consulting and Clinical Psychology, 45*(6), 1156–1158. https://doi.org/10.1037/0022-006X.45.6.1156
- Dowrick, P. W., & Dove, C. (1980). The use of self-modeling to improve the swimming performance of spina bifida children. *Journal of Applied Behavior Analysis, 13*(1), 51–56. https://doi.org/10.1901/jaba.1980.13-51
- Dowrick, P. W. (1986).*Social Survival for Children: A Trainer's Resource Book*, Brunner/Mazel, New York.
- Dowrick, P. W. (1991). *Practical guide to using video in the behavioral sciences.* John Wiley & Sons.
- Dowrick, P. W., & Raeburn, J. M. (1995). Self-modeling: Rapid skill training for children with physical disabilities. *Journal of Developmental and Physical Disabilities, 7*(1), 25–37. https://doi.org/10.1007/BF02578712
- Egel, A.L., Richman, G.S. and Koegel, R.L. (1981), NORMAL PEER MODELS AND AUTISTIC CHILDREN'S LEARNING. Journal of Applied Behavior Analysis, 14: 3-12. https://doi.org/10.1901/jaba.1981.14-3
- Freeman EW, Boxer AS, Rickels K, Tureck R, Mastroianni L Jr. Psychological evaluation and support in a program of in vitro fertilization and embryo transfer. Fertil Steril. 1985 Jan;43(1):48-53. doi: 10.1016/s0015-0282(16)48316-0. PMID: 3965315.
- Gelman, Debra Levin. Design for Kids - Rosenfeld Media, 2014. https://rosenfeldmedia.com/wp-content/uploads/2014/11/DesignforKids-excerpt.pdf.
- Happé, F. G. E. (1994a). Wechsler IQ profile and theory of mind in autism. Journal of Child Psychology and Psychiatry, 35, 1461-71.
- Heyes, Cecilia, and Caroline Catmur. "What Happened to Mirror Neurons?" *Perspectives on Psychological Science*, vol. 17, no. 1, Jan. 2022, pp. 153–168, doi:10.1177/1745691621990638.
- Hickok G. Eight problems for the mirror neuron theory of action understanding in monkeys and humans. J Cogn Neurosci. 2009 Jul;21(7):1229-43. doi: 10.1162/jocn.2009.21189. PMID: 19199415; PMCID: PMC2773693.
- Hickok, Gregory. "The role of mirror neurons in speech and language processing." *Brain and language* vol. 112,1 (2010): 1-2. doi:10.1016/j.bandl.2009.10.006
- Hitchcock, C. H., Dowrick, P. W., & Prater, M. A. (2003). Video Self-Modeling Intervention in School-Based Settings: A Review. Remedial and Special Education, 24(1), 36–46. https://doi.org/10.1177/074193250302400104

- Horst, J. S., Scott, E. J., and Pollard, J. P. (2010). The role of competition in word learning via referent selection. *Dev. Sci.* 13, 706–713.
- Horst, J. S., Parsons, K. L., and Bryan, N. M. (2011a). Get the story straight: contextual repetition promotes word learning from storybooks. *Front. Psychol.* 2:17. doi: 10.3389/fpsyg.2011.00017
- Horst, J. S. (2013). *Context and repetition in word learning. Frontiers in Psychology, 4.* doi:10.3389/fpsyg.2013.00149
- Hoque, Mohammed et al. "Exploring Speech Therapy Games with Children on the Autism Spectrum." 10th Annual Conference of the International Speech Communication Association, INTERSPEECH 2009, September 6, 2009 - September 10, 2009.
- Kazdin, A. E. (2001). Behavior modification in applied settings (6th ed.). Belmont, CA: Wadsworth.
- Kelly, J. A. (1982). Social skills training: A practical guide for interventions. New York: Springer.
- Lincoln, A. J., Courchesne, E., Kilman, B. A., Elmasian, R. & Allen, M. (1988). A study of intellectual abilities in high-functioning people with autism. Journal of Autism and Developmental Disorders, *18*, 505-24
- MacDonald, R., Clark, M., Garrigan, E., & Vangala, M. (2005). Using video modeling to teach pretend play to children with autism. *Behavioral Interventions*, *20*(4), 225–238. https://doi.org/10.1002/bin.197
- Malinverni, L., et al., An inclusive design approach for developing video games for children with Autism
- Spectrum Disorder, Computers in Human Behavior (2016), http://dx.doi.org/10.1016/j.chb.2016.01.018
- Masataka, N. (1999). Preference for infant-directed singing in 2-day-old hearing infants of deaf parents. Developmental Psychology, 35(4), 1001–1005. https://doi.org/10.1037/0012-1649.35.4.1001
- McMurray, B., Horst, J. S., and Samuelson, L. K. (2012). Word learning as the interaction of online referent selection and slow associative learning. *Psychol. Rev.* 119, 831–877.
- Mehta, Smita & Miller, Trube & Callahan, Kevin. (2010). Evaluating the Effectiveness of Video Instruction on Social and Communication Skills Training for Children With Autism Spectrum Disorders: A Review of the Literature. Focus on Autism and Other Developmental Disabilities - FOCUS AUTISM DEV DISABIL. 25. 23-36. 10.1177/1088357609352901.
- Nikopoulos, Christos & Keenan, Mickey. (2004). Effects of Video Modelling on Training and Generalisation of Social Initiation and Reciprocal Play by Children With Autism. European Journal of Behaviour Analysis. 5. 10.1080/15021149.2004.11434227.

- Nikopoulos CK, Keenan M. Using video modeling to teach complex social sequences to children with autism. J Autism Dev Disord. 2007 Apr;37(4):678-93. doi: 10.1007/s10803-006-0195-x. PMID: 16897375.
- Pioggia, G., Sica, M. L., Ferro, M., Igliozzi, R., Muratori, F., Ahluwalia, A., & De Rossi, D. (2007). Human-Robot Interaction in Autism: FACE, an Android-based Social Therapy. RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication. doi:10.1109/roman.2007.4415156
- Politimou, N., Dalla Bella, S., Farrugia, N., & Franco, F. (2019). Born to Speak and Sing: Musical Predictors of Language Development in Pre-schoolers. Frontiers in Psychology, 10. doi:10.3389/fpsyg.2019.00948
- POP, C. A., SIMUT, R., PINTEA, S., SALDIEN, J., RUSU, A., DAVID, D., … VANDERBORGHT, B. (2013). *CAN THE SOCIAL ROBOT PROBO HELP CHILDREN WITH AUTISM TO IDENTIFY SITUATION-BASED EMOTIONS? A SERIES OF SINGLE CASE EXPERIMENTS. International Journal of Humanoid Robotics, 10(03), 1350025.* doi:10.1142/s0219843613500254
- Queensland Government. “The Power of Repetition.” *Early Childhood Education and Care*, 30 Oct. 2020, https://earlychildhood.qld.gov.au/early-years/activities-and-resources/resources-parents/read-and-count/the-power-of-repetition.
- Quill, K.A. Instructional Considerations for Young Children with Autism: The Rationale for Visually Cued Instruction. *J Autism Dev Disord* **27,** 697–714 (1997). https://doi.org/10.1023/A:1025806900162
- Rizzolatti, Giacomo; Fabbri-Destro, Maddalena; Cattaneo, Luigi (2009). *Mirror neurons and their clinical relevance. Nature Clinical Practice Neurology, 5(1), 24–34.* doi:10.1038/ncpneuro0990
- Safwat, Rasha Farouk; Sheikhany, Aya R. “Effect of parent interaction on language development in children”; April 2014; p.225-263; DOI: 10.4103/1012-5574.138488
- Salimpoor, V. N., van den Bosch, I., Kovacevic, N., McIntosh, A. R., Dagher, A., & Zatorre, R. J. (2013). Interactions Between the Nucleus Accumbens and Auditory Cortices Predict Music Reward Value. Science, 340(6129), 216–219. doi:10.1126/science.1231059
- Sesame Street. Best Practices: Designing Touch Tablet Experiences for Preschoolers, 2012. https://joanganzcooneycenter.org/wp-content/uploads/2020/02/SesameWorkshop-2012.pdf.
- Shah, A., & Frith, U. (1983). An islet of ability in autistic children: A research note. *Child Psychology & Psychiatry & Allied Disciplines*, *24*(4), 613–620. https://doi.org/10.1111/j.1469-7610.1983.tb00137.x

- Sigler, E.A., Aamidor, S. From Positive Reinforcement to Positive Behaviors: An Everyday Guide for the Practitioner. *Early Childhood Educ J* **32,** 249–253 (2005). https://doi.org/10.1007/s10643-004-0753-9
- Suttie, Jill. “Why Parents Sing to Babies.” Greater Good, 9 Jan. 2016, https://greatergood.berkeley.edu/article/item/why_parents_sing_to_babies.
- Tryon, Adeline S., and Susan P. Keane. "Promoting Imitative Play through Generalized Observational Learning in Autisticlike Children." *Journal of Abnormal Child Psychology*, vol. 14, no. 4, 1986, pp. 537. *ProQuest*, https://www.proquest.com/scholarly-journals/promoting-imitative-play-through-generalized/docview/1300096313/se-2?accountid=14496.
- Underwood, Paul L. “Why Do Kids Love Terrible Music?” The New York Times, The New York Times, 15 Apr. 2020, https://www.nytimes.com/2020/04/15/parenting/music-for-children.html.
- University of Montreal. "Singing calms baby longer than talking: New study shows that babies become distressed twice as fast when listening to speech compared to song." ScienceDaily. ScienceDaily, 28 October 2015. <www.sciencedaily.com/releases/2015/10/151028054532.htm>.
- Welch, Graham. (2001). The importance of singing in child development. 5 to 7 Educator. 1. 35-37. 10.12968/ftse.2001.1.6.16857.
- Westerlund, M. and Lagerberg, D. (2008), Expressive vocabulary in 18-month-old children in relation to demographic factors, mother and child characteristics, communication style and shared reading. Child: Care, Health and Development, 34: 257-266. https://doi.org/10.1111/j.1365-2214.2007.00801.x
- Wilson, Kaitlyn. (2012). Incorporating Video Modeling Into a School-Based Intervention for Students With Autism Spectrum Disorders. Language, speech, and hearing services in schools. 44. 10.1044/0161-1461(2012/11-0098).
- Yogman M, Garner A, Hutchinson J, et al; AAP COMMITTEE ON PSYCHOSOCIAL ASPECTS OF CHILD AND FAMILY HEALTH, AAP COUNCIL ON COMMUNICATIONS AND MEDIA. The Power of Play: A Pediatric Role in Enhancing Development in Young Children. Pediatrics. 2018;142(3):e20182058
- Yu, C., and Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychol. Sci.* 18, 414–420.

Speech Blubs
Helping Kids Speak!