JOURNAL OF NEW APPROACHES IN EDUCATIONAL RESEARCH
2023, VOL. 12, NO. 1, 40-61, e-ISSN: 2254-7339
https://doi.org/10.7821/naer.2023.1.1269

ORIGINAL

Experimental Analysis of the Effectiveness of a Cyber-physical Robotic System to Assist Speech and Language Pathologists in High School

Eldon Glen Caldwell-Marin[1], Miguel Cazorla[2] and José María Cañas-Plaza[3]

[1]University of Costa Rica, Costa Rica
[2]University of Alicante, Spain
[3]Universidad Rey Juan Carlos, Spain

ABSTRACT

This research focuses on whether the use of a cyber-physical robotic system (CPRS) to assist Speech and Language Pathologists (SLP) in a Special Education service is beneficial. The research method is based on a quasi-experiment with a 2k design and a two-way ANOVA, implemented with real high school students over 10 weeks. It was found that the use of this CPRT could improve, preliminarily and as an initial exploratory finding, therapeutic speech effectiveness up to 11.3 percentage points with a statistical confidence of 95%, when SLPs work with students with mild articulation disorder and a restricted time for therapy, but especially when the technology is used without time constraints. It is concluded that assistive CPRT could be a causal factor of improvement in specific treatments performed by SLPs, with the statistical evidence being sufficiently significant (95%) to maintain scientific and educational interest in this research line in the future.

Keywords ASSISTIVE ROBOTICS, SPECIAL EDUCATION, SPEECH AND LANGUAGE THERAPY, CYBER-PHYSICAL SYSTEMS, EDUCATION TECHNOLOGIES

Received 2022-10-10
Revised 2022-10-14
Accepted 2022-11-25
Published 2023-01-15

Corresponding Author
Eldon Glen Caldwell-Marin,
eldon.caldwell@ucr.ac.cr

University of Costa Rica, PO box: 11501-2060, Costa Rica.

DOI https://doi.org/10.7821/naer.2023.1.1269
Pages: 40-61

1 INTRODUCTION

Rates of speech disorders across the world remain high. The reported prevalence ranges widely between 1% and 25%, with variations according to the type of disorder and age groups, and the highest prevalence being found in younger population (Choo, Smith, & Li, 2022).

This research focuses on articulation disorders, that is, alterations in the articulation of phonemes that may occur due to the absence or alteration of some of these, or as a result of improper replacement by others (Chaminade et al., 2012; Kim et al., 2013; Robles-Bykbaev et al., 2018).

The research problem was defined as follows: Is it beneficial to use robotic cyber-physical technology to assist the practice of special educators specialized in speech and language therapy for students with articulation disorders in terms of the therapeutic effectiveness of speech over time?

As a first phase, a robotic system was developed based on a physical and virtualized robot (cyber-physical robot platform) that executes a therapeutic exercise for mild articulation disorder and which was programmed for mobile devices (Caldwell, Morales, Solis, Cazorla, & Cañas, 2021). It is useful for three types of functional disorder therapies in professional practice: "substitution therapy", "omission therapy" and "phoneme change" (Flagge, 2013; García-Vergara, Brown, Park, H, & Howard, 2014; Pennington, Saadatzi, Welch, & Scott, 2014).

The platform developed was preliminarily validated in two tests, one with pathologists and another with real patients. We conclude that the possible practical use of a robotic platform such as this has been preliminarily evidenced and that pathologists are highly confident it can become a useful tool in the treatment of different types of mild functional disorders.

This paper concerns the second phase of our research, which is the experimental exploration of benefits of the platform with real users in therapeutic sessions and with statistical rigor. The first goal is to experimentally analyze the platform as a tool to assist educators specialized in speech and language therapies in terms of its therapeutic effectiveness. The second aim is to determine whether there is any difference in the effectiveness achieved with the use of the physical robot and the application for mobile devices, compared to the use of the application for mobile devices alone.

In the following sections, the theoretical framework and related research works are first described. Subsequently, the robotic platform model, the design of the experiment and the groundings of its statistical rigor are presented. Finally, we describe the results of the experiment implemented, suggesting conclusions and future research opportunities.

2 THEORETICAL FRAMEWORK

This section presents a brief outline of the theoretical framework and a review of works related to this research. Various conditions may cause the need for speech or language therapy, such as hearing problems, cognitive delays (intellectual, reasoning), weakness in oral musculature, birth conditions, such as cleft lip, autistic spectrum disorders, motor or respiratory problems, swallowing disorders and traumatic brain injuries (Costa, Lehmann, Dautenhahn, Robins, & Soares, 2015; Costescu, Vanderborght, & David, 2015; Raul & Ahyea, 2016).

Commonly found needs are those related to aphasia, apraxia or dysarthria of all types and degrees, dementia, head trauma, cognitive impairments, and typical aging. Specifically, in the area of voice, therapies are oriented towards dysphonia, vocal education and disabilities related to medical procedures such as laryngectomy. In relation to swallowing and speech, disorders include dyslalia, dysphagia, atypical or adapted swallowing and oral breathing.

A functional disorder is one where diction errors are typically similar to those produced by children during the stages of language acquisition. Certain analytical dimensions are postulated as causes of functional disorders, among which there are psychological, environmental, hereditary and cognitive factors (ASHA, 2018).

Typically, in the field of speech therapy, different types of articulation disorders can be identified. For example, developmental disorders refer to anomalies of articulation that occur in the early stages of speech development and are considered normal, since the child is in the process of acquiring language and still cannot accurately reproduce all of the phonemes. Audiogenic disorders are caused by hearing impairments, in which the child cannot receive, and therefore discriminate between similar sounds. In addition, organic disorders are caused by organic alterations, such as genetic predisposition, insufficient psychomotor control, environmental problems, muscular tension or cognitive deficits.

Specifically, this research focuses on the possibility that robotic cyber-physical technologies can assist therapies to treat functional articulation disorders. A number of factors are postulated as causes of functional disorders, including psychological factors, environmental factors, hereditary factors and cognitive factors.

The literature review performed reflects a growth in the use of robotics and information and communication technologies in applications that seek the achievement of therapeutic objectives in the field of speech and language skills (Caldwell, Cazorla, García, Azorin, & Zamora, 2017; Estevez, Terrón-López, Velasco-Quintana, Rodríguez-Jiménez, & Álvarez Manzano, 2021). However, many of these are oriented towards robot-like features, behavior analysis, models, teaching or providing feedback or learning stimuli (Begum, Serna, & Yanco, 2016; Cabibihan, Javed, Ang, & Aljunied, 2013; Diehl et al., 2014; Scassellati, Admoni, & Mataric, 2012; Seong, Jin, & Hyun, 2016).

In addition, it is reported that the effectiveness of the treatment depends on the perseverance of the individual and, on the other hand, the effectiveness of the human-robot interaction depends on the degree of an individual's acceptance of the robot. This correlates with their human appearance, that is, if the robot assistant appears friendly, people are more likely to accept it as an assistive technology (Calderita et al., 2015).

This type of research has also been conducted in other fields in which social and assistive robotics is developed; for example, in the care of older adults and in the physical rehabilitation and care of people with cognitive disabilities (Gregor, 2016). J. C. Pulido et al. (2017), for instance, present an application for rehabilitation of upper limbs without contact and with robotic autonomy and teleoperation for children with physical disabilities (NAOTherapist), which has been positively evaluated. However, research on social and assistive robotics has often focused on individuals on the autistic spectrum (Tapus et al., 2012), and thus there remains a great opportunity to apply it to other therapeutic fields of speech and language.

Deeper analysis of this literature review can be found in Caldwell et al. (2017, 2021); Marge et al. (2022); Vázquez-Villasuso and Diaz-Monterrey (2015).

3 DESCRIPTION OF THE ROBOT AND MOBILE DEVICE BASED SYSTEM

This section describes the core technical foundations of the speech-language cyber-physical robotic platform and its essential characterization.

First, Kansei Robotic Flow (KRF) is commonly used to determine the behavioral characteristics of a robot or robotic platform based on a chosen archetype (Marge et al., 2022; Pakrasi, 2018). A robot archetype is understood as a common blueprint for character traits that can be found in stories across cultures and are thus easily identifiable (Scassellati & Tsui, 2016).

KRF was used as a systematic approach to define the personality traits of the robot behavior and the mobile device application that emulates the robot (Marge et al., 2022; Pakrasi, 2018; Pakrasi, Laviers, & Chakraborty, 2018). Five special educators certified in speech and language therapies were interviewed to apply the KFR method and as Zero Archetype, the "human-robot interaction based on trust/friendship traits" (HRI/TF) was defined with two "Kansei" emotions: happiness and perseverant feeling.

The next phase was the engineering process. Spanish was the language selected because scientific research in the field in this language is considerably more limited than that in other languages, such as English. In addition, it was decided to work on the implementation of "substitution", "omission" and "phoneme change" functional articulation disorders due to their high frequency of use (Liu, Li, Chen, Goh, & Sung, 2014).

Figure 1 shows the general architecture of the system where two key interactions define core processes related with the users.

One interaction is between pathologists and the physical robot and mobile devices software. The second interaction is between students (speech language users who must to interact with supervision) and the physical robot and mobile devices software. Pathologists define topic of the exercise, threshold, phonemes and complexity level of the words and pictograms. These parameters can be changed remotely if the pathologist decides to make adjustments.

Figure 2 shows an overview of the schematic roles of the architecture components. Pathologists work with speech language users (SLU) managing the selected technology (robot with app or the app alone). If the selected technology is the app alone, users can operate the app with the pathologist training. On the other hand, speech language users can run the mobile devices application with a username and password that parents or tutors must to control, because the therapeutic exercises should be supervised.

It is important to clarify that the platform follows international speech therapeutic practice and standards. For example, the database with words associated with pictograms is ARAWORD (an element of the ARASUITE system) which is an internationally recognized framework developed in Spain by the Government of Aragon (http://www.arasaac.org/software.php?id_software=2).

As shown in Figure 1, the design of the Robotic Platform Model is based on a humanoid NAO v.6, chosen for its flexibility, Python language compatibility and friendly appearance. Specifically, it is based on CMU Sphinx (Carnegie Mellon University, CMU, https://cmu

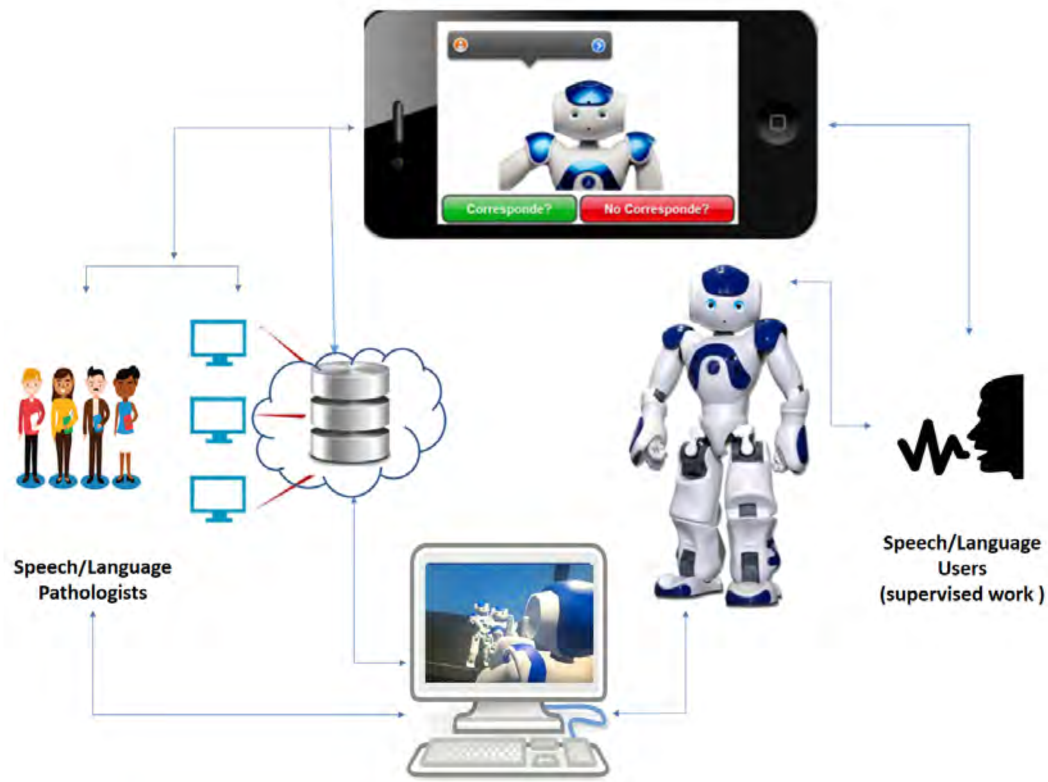

Figure 1 General architecture of the system

sphinx.github.io/wiki/download/) with an integration toolset for Apache CORDOVA (https://cordova.apache.org/), which works with an Internet connection and is compatible with Python and NAO robot platform. Apache CORDOVA, owned by Adobe Systems, is a platform that uses tools such as JavaScript and HTML5, among others, as an alternative framework when hybrid architectures are required.

The speech/language users can interact directly with the physical robot, as well as with the emulated robot, using a mobile device application. When the user speaks, the physical robot has the functionality of voice recognition. It can assist an educator with a speech and language specialization (or Speech Language Pathologist) by finding previously defined speech failures to provide feedback about words previously organized by themes and therapeutic goals. In addition, the user can interact with the mobile device application and a similar process is executed, but which uses a virtual emulation of the Robot designed with Kansei robotic archetype parameters.

The speech recognition algorithm was programmed using an Application Programming Interface (API) and an integration of Python and Apache CORDOVA. This algorithm was programmed to be able to identify specific phonemes provided by the pathologist. As in the humanoid robot in the physical world, this API works with a "threshold" assignment to perform the recognition. Python is used in order to control the robot and execute specific movements or therapeutic routines as an assistant of the speech and language pathologist

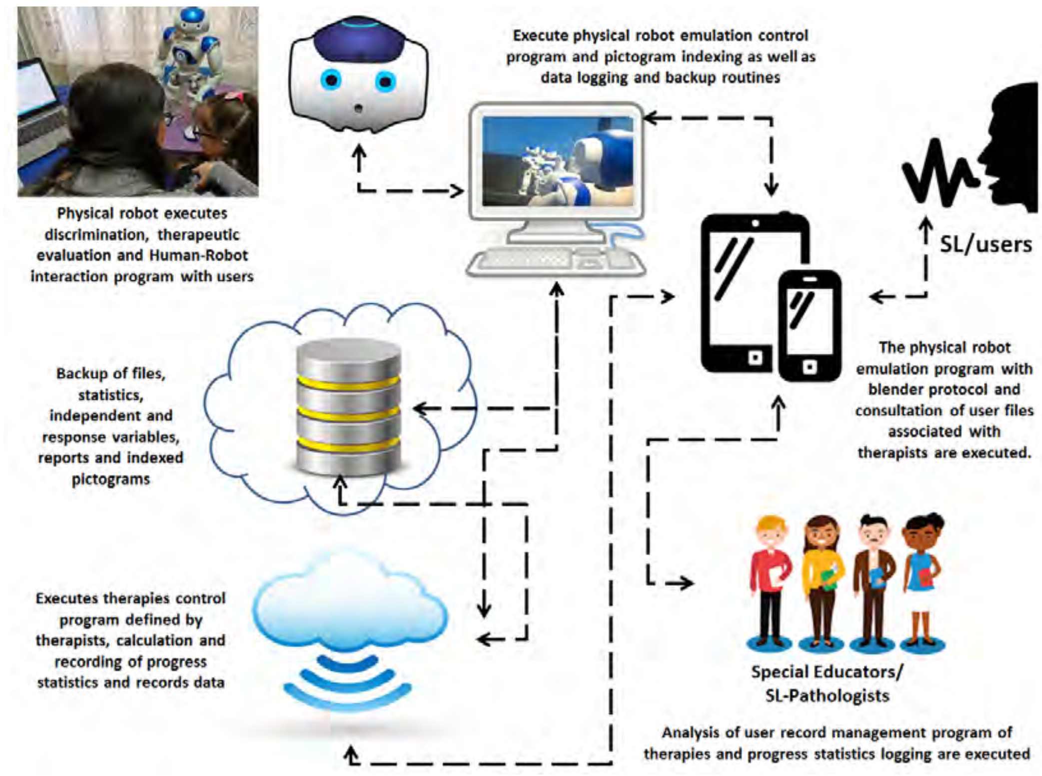

Figure 2 Schematic roles of the Robotic Platform components

(or special educator).

Further details about the platform design and its functional validation can be found in Caldwell et al. (2021).

4 RESEARCH METHOD AND IMPLEMENTATION APPROACH

This section describes the foundations of the design of a full factorial 2k quasi-experiment, as a research method to explore possible advantages of the platform developed. It is worth clarifying that the platform is designed to assist the special educator (speech and language pathologist or SLP) in their work, and not in the sense of substituting robots for humans. The platform can remotely give specific assistance, but special educators or speech and language pathologists must diagnose, and define and supervise the therapeutic process, analyzing the data that can be generated through the system.

The experiment was carried out in Costa Rica, at the Heredia Special Education Centre. We worked with students (young men and women) supervised by their parents and professional therapists responsible for their treatment. They were all aged between 15 and 20 years. The selection of participants was based on homogeneous diagnoses of mild functional articulation disorders, a socioeconomic status above the poverty line (to ensure they had sufficient Internet services and mobile devices to participate in the experiment) and their being able to read and identify pictograms correctly. All the students gave their consent and, due to legal requirements, the parents and tutors of the students signed their consent for participation advised by the SLPs. Additionally, this research project is registered at the University of Costa Rica and complies with all the respective ethical requirements.

In this research, a specific therapeutic exercise was defined by the SLPs and was executed as part of a therapeutic strategy defined for each participant in the experiment. This basic exercise consists in presenting an individual with a sequence of pictograms that are related one by one with a word that contains specific phonemes to improve pronunciation or functional articulation in a speech process. The patient must listen to the pronunciation and then pronounce the word to be evaluated.

Regarding the design of experiments, factorial designs can be unifactorial or multifactorial (two or more factors) and in research with humans, "intrasubject" designs are common, where the same individuals participate in all the conditions of the variables, as are "intersubject" designs where the participants are different. Likewise, mixed designs can be defined with intra and intersubject procedures (Creswell & Creswell, 2017) .

The 2k design responds to a scientific tradition where its use is widespread, taking into account different variants in positivist research in fields such as biology, chemistry, agronomy and engineering (G. Pulido, Vara, & Salazar, 2012). Such experimental designs are also widely recognized and used in the social sciences, especially since the publication of Harley & Harley in the 60s (W F Harley & Harley, 1968).

Factorial designs are uncommon in the literature on research in special education. However, such experiments are effective and highly accepted when conducting comparative and exploratory studies in the field of speech and language therapy, specifically when the demand of the statistical requirements can be met (Collins, Dziak, & Li, 2009; Glogowska, 2011; Jackson & Cox, 2013; Kirk, 2014; Medina-Varela & López-Reyes, 2011; A. A. Montgomery, Peters, Little, & L, 2003).

Like all scientific research methods, regardless of the underlying epistemology, the experimental method and particularly the complete factorial design has advantages and disadvantages that have been well documented in various studies in speech and language therapy (Lewison & Carding, 2003; Smith, Williams, & Karen, 2017).

Multifactorial designs are the most widely used because they have the advantage of allowing for the study of more complex and interactive phenomena such as those that frequently involve humans. Consequently, they are typically selected for studies that involve human behavior and especially when the aim is to simultaneously evaluate the effects of the factors. According to Medina-Varela and López-Reyes (2011), the factorial design "allows the interaction effect (the combined effect of both variables) to be assessed, that is, it allows the main effect of A, that of B and the combined effect of both to be determined". Another advantage reported is that such experiments enable a better use of resources (Smith et al., 2017).

According to A. A. Montgomery et al. (2003), the completely randomized 2k design "is particularly useful in the initial stages of experimental work, when the effect of many factors is likely to be investigated. This design provides the lowest number of runs with which k factors can be studied in a complete factorial design.

D. C. Montgomery (2002) also explains that in a factorial design with two levels, the levels of the factors are conventionally called "low" and "high". Additionally, also by convention, the effects are denoted using capital letters A, B and AB.

In addition, ν is used to denote that two factors are in the low level and, on the other hand, "ν**,** α**,** β **and** $\alpha\beta$" represent the total **n** replications implemented with the combination of the treatments.

The average effect of a factor can be calculated by means of the change in the response variable produced by a change in the level of that factor averaged for the levels of the other factor.

Effect A at the low level of B is [α - ν]/**n** and the effect of A with the high level of B is [$\alpha\beta$ - β] / **n**.

By averaging these two quantities, the main effect of A is obtained: the average principal effect of B is found from the effect of B with the low level of A and the effect of the AB interaction is defined as the average difference between the effect of A with the high level of B and the effect of A with the low level of B.

The combinations for this factorial design are established and identified as follows, becoming the responses for n replicas:

(0,0) : ν

(1,0) : α

(0,1) : β

(1,1) : $\alpha\,\beta$

Therefore, the average effects of A and B are (Figure 3):

$$A = \frac{1}{2n}(\alpha\beta + \alpha - \beta - \nu) \quad (1)$$

$$B = \frac{1}{2n}(\alpha\beta + \beta - \alpha - \nu) \quad (2)$$

Figure 3 Average effects of A and B

Additionally, average effect AB (average difference between the effect of A at level 1 of B and the effect of A on level 2 of B) can be calculated as follows (Figure 4):

$$AB = \frac{1}{2n}(\alpha\beta + \nu - \alpha - \beta) \quad (3)$$

Figure 4 Average effect AB

Where, again, ν, α, β and $\alpha\beta$ represent the total n replications carried out with the combination of the treatments.

Specifically, for this research project, the factors and levels were defined as follows:

Factor 1: Time of exposure to therapeutic exercise

- Level 1: Fixed time of 30 minutes maximum per session; 2 sessions per week in previously established time. (0)
- Level 2: Fixed time of 30 minutes maximum per session; two or more sessions per week at the time the user wants. (1)

Factor 2: Technology

- Level 1: Collaborative Humanoid Robot (CHR) and, in addition, a virtual emulation of the CHR using a mobile device application. (0)
- Level 2: Virtual emulation of the CHR using a mobile device application. (1)

In addition, the combinations for the experimental design can be described as follows:

(0,0) : ν

Group that uses the collaborative humanoid robot (CHR) and its virtual emulation with a fixed time of 30 minutes maximum per session and 2 sessions per week at a previously established time. One of the sessions will always be face-to-face with the supervision of the special educator who supervises the therapeutic process.

(1,0) : α

Group that uses a virtual emulation of the collaborative humanoid robot (CHR) with a fixed time of 30 minutes maximum per session and two sessions per week at a previously established time. One of the sessions will always be face-to-face with the supervision of the special educator who supervises the therapy.

(0,1) : β

Group that uses the collaborative humanoid robot (CHR) and its virtual emulation with a fixed time of 30 minutes maximum per session and two or more sessions per week in a schedule that the user determines. One of the sessions will always be face-to-face with the supervision of the special educator who supervises the therapy.

(1,1) : $\alpha\,\beta$

Group that uses a virtual emulation of the collaborative humanoid robot (CHR) with a fixed time of 30 minutes maximum per session and two or more sessions per week at the time the user determines. One of the sessions will always be face-to-face with the supervision of the special educator who supervises the therapy.

The response variable ($\phi\mathbf{p}$) is the speech effectiveness in weekly periods from an initial level previously established by a special educator certified in speech therapy and in accordance with an accepted international standard. This effectiveness is calculated as the total number of words pronounced correctly in weekly periods divided by the total number of words reviewed by the user. In addition, the ratio of words pronounced correctly per minute of therapy ($\rho\mathbf{t}$) is analyzed.

Furthermore, statistical data is collected for performance analysis, such as words failed and their frequency of failure per period of time (week, day, minute). Additionally, indicators such as number of sessions, time lapse of active therapy per session, total time lapse of active therapy per period of time (day, week, month, etc.) number of different words used, number of sessions per period (day, week, month) are also collected and analyzed.

Regarding the number of replicas, it is typical to define this based on the test of the differences between the treatment means and the Fisher (F0) statistic. The number of replicas needed is primarily influenced by four variables that are required for the calculations Kuehl (2001), as follows:

1. The variance (σ^2)
2. The size of the difference (which has physical significance) between the two means.
3. The level of significance of the test or the probability of type I error.
4. The power of the test or the probability of detecting the difference between means (1-probability of type II error).

Using these calculation parameters, the number of replicas is estimated as follows (Figure 5):

$$r \geq 2\,[Z_{\alpha/2} + Z_{\beta}]^2\,(\sigma/\gamma)^2 \qquad (4)$$

Figure 5 Estimation of the number of replicas of the experiment

Where:

- **r**: number of replicas
- γ: the size of the difference (which has a physical meaning) between the two means
- σ: population deviation
- α: level of statistical significance or type I error probability
- β: type II error probability
- **Z**α**/2**: statistic for a standard normal distribution at a level of significance α
- **Z**β: statistical for a standard normal distribution at a probability level β

It is also possible to estimate the number of replicas if the coefficient of variation (CV) is known, by dividing the population deviation by the population mean multiplied by 100. The CV replace σ equation (4) (Kuehl, 2001).

Table 1 shows the number of replicas estimated according to the above criteria and parameters (Kuehl, 2001). As can be seen, Table 1 shows that using a number of replicas between 7 and 26, we can get inferences with 95% of confidence (α=0.05), with percentage differences between averages of 10% and 20% and covering good probability values **(1-** β**)** for coefficients of variation of 5 % and 10% (considering parameters for experiments with human beings).

For a completely randomized 2k experiment, it is recommended to establish the number of replicas presented in Table 2. This is widely used because it is constructed based on Table 1 (Kuehl, 2001) and the statistical study of reliability and internal validity of multiple 2k design experiments reported in the literature on factorial design of experiments (G. Pulido et al., 2012).

Table 1 Number of replicates required for a given coefficient of variation and probability (1- β) to obtain a significant difference of% γ between two treatment means, with a bilateral test at a level of significance α (Kuehl, 2001)

		α= **0.05**		α =**0.01**	
		% γ		**%** γ	
%CV	1- β	10	20	10	20
5	0,8	4	1	6	2
	0,95	7	2	9	3
10	0,8	16	4	24	6
	0,95	26	7	36	9

For the proposed quasi-experiment, it was decided to work with groups of 4 users (replicas), entailing 16 runs for this experiment. This number of runs is the maximum recommended presented inTable 2 for a completely randomized experiment with two factors and two treatments, rendering this proposal robust and yielding a well recommended level of statistical significance for this type of study.

Table 2 Recommended replicas in 2k designs (J. C. Pulido et al., 2017)

Design 2k	Recommended replicas	Number of runs
k= 2	3 o 4	12,16
k= 3	2	16
k= 4	1 o 2	16,32
k= 5	Fraction 25-1 or 1	16,32
k= 6	Fraction 26-2 or fraction 26-1	16,32
k= 7	Fraction 27-3 or fraction 27-2	16,32

The sample of 16 users (4 groups with 4 students of the Special Education Secondary School) comes from different randomly selected male and female by special educators certified in speech and language therapy (each student is linked to one previously assigned pathologist). The systematic process to select and assign individuals to each group is shown inFigure 6. As observed, a double randomization is executed for the assignment of users to each group in such a way that the assumption of independence of the observations can be met, and such that the methodical parsimony does not distort the evidence of the homoscedasticity assumption in the analysis of variability (ANOVA) to be performed in the experiment. In addition, a group of control patients were selected and assigned with 10 weeks of conventional therapy.

The special educators (speech language pathologists) It is worth noting that each user has no interaction with other users nor do the therapists interact with each other, thus avoiding extraneous variables being introduced into the experiment.

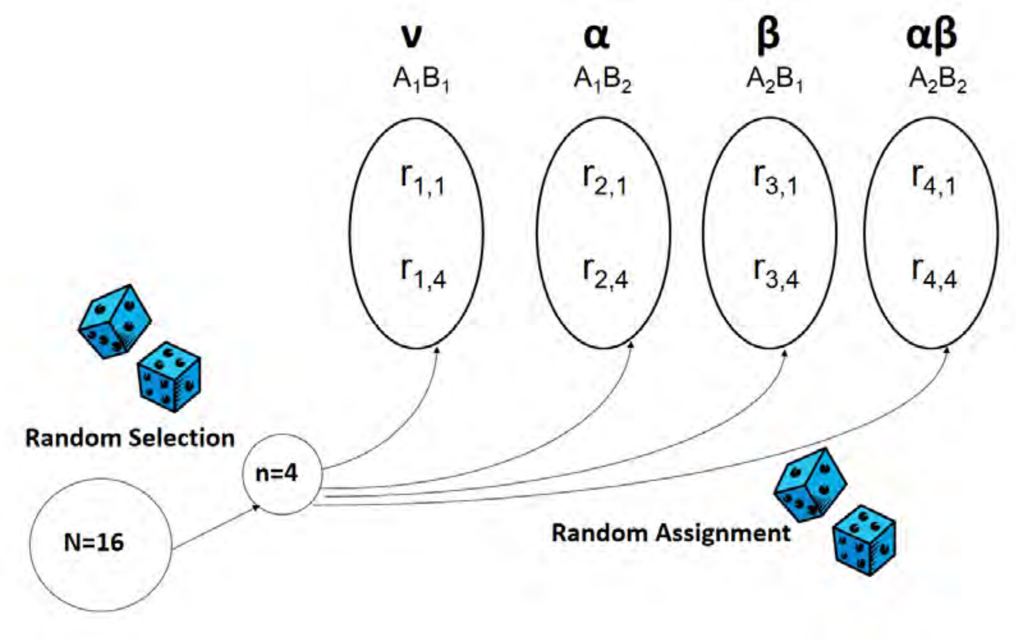

Figure 6 Systematic process for full factorial 2k experiment sample assignment

Professional recommendation is that therapy should be repeated cyclically for at least 10 weeks to see plausible results in advancing speech skills. Accordingly, the experiment is conducted over that period of time. This allows for the analysis of variability (ANOVA) with the weekly observations and the results obtained in the response variable for the total of 10 weeks.

5 RESULTS AND DISCUSSION

The analysis of the experimental results is presented in this section. It should be noted the results refer to the factorial design cyclically repeated over the 10 weeks. The data is analyzed from the following perspectives:

1. The differences between groups of individuals (which are characterized by the treatments associated with the causal variables) in each of the 10 weeks with respect to the response variable (therapeutic effectiveness).
2. The evolution of the response variable over time, that is, over the 10 weeks of experimentation with the four groups of participants.
3. All the participants were supervised and followed the therapeutic strategy of a special educator certified as a speech and language pathologist authorized by law, who was trained in the use of the assistive robotic platform.

With the data obtained for 10 weeks, a non-parametric test known as the "Mann Whitney U Test" was performed, which allows two statistically small samples to be compared when it is not possible to verify normality or homoscedasticity. This test is equivalent to the comparison of t-Student averages, which are typically used for normality assumptions.

The null hypothesis is the equality of the groups, that is, the assumption that therapeutic effectiveness is no different for the groups that worked with the restricted time (two therapy sessions per week) and the groups that worked without time restriction.Table 3 shows the results of this test and, as can be seen, the value p> α, that is, p-value > 0.05, and so, in this case, the null hypothesis cannot be rejected.

Table 3 Mann-Whitney Test for Factor 1 in 10 weeks.

	Mann-Whitney Test for Factor 1 over 10 weeks	Sum of ranks
2 therapy sessions per week (max.30 min per week)	10	92
Time without restrictions in the use of mobile device APP	10	118
TOTAL	20	210

Note: 105.00 expected value; 13.23 standard deviation; 0.344 p-value (two-tailed)

Table 4 shows the results of this test for Factor 2. The null hypothesis is also the equality of the groups, that is, the assumption that the therapeutic effectiveness is not different for the groups that worked with the robot and the application and the groups that worked only with the application, regardless of the time of use. Unlike the previous case, it was found that "p-value" <α, that is, "p-value" <0.05, and so, in this case, the null hypothesis must be rejected.

Table 4 Mann-Whitney Test for Factor 2 over 10 weeks

	n	Sum of ranks
Robot with app	10	55
APP	10	155
TOTAL	20	210

Note: 105.00 expected value; 13.23 standard deviation; 0.0002 p-value (two-tailed)

The results of the first week's experiment are shown inTable 5 and represent the initial state of the experiment on which the findings are analyzed and data collection performed.

The ANOVA inferential analysis shows that there is insufficient statistical evidence to reject the hypothesis of equality in effectiveness due to the time in the use of the robotic platform compared to the use of the mobile device application, considering a 95% statistical confidence. This is because the "p-value" (0.33) is greater than 0.05 when the data for Factor 1 (time in the use of technological resources) are correlated with the robot and application scenario and the mobile device application scenario.

The two-way ANOVA analysis shows there is sufficient statistical evidence to reject the hypothesis of equality in the effectiveness of the use of the application for mobile devices compared to the use of the robotic platform (robot and application).

Table 5 Two-Way ANOVA Results for Week # 1.

		Factor 2 *p-value = 3.89E-06*		
		Robot with app	**App**	**Global Average**
	Two therapy sessions per week (max. 30 min. per week)			
		35,70	60,50	**48,10**
Factor 1				
p-value= 0.330	Time without restrictions in the use in the use of mobile device App	29,30	29,80	**44,60**
		29,30		
	Global Average	32,50	60,20	**46,30**

Note: 4 replications per cell; Interaction p-value= 0.428

This means that, with a 95% statistical confidence ("p-value" = 3.89E-06; less than 0.05), groups that used the application alone started the experiment with a higher level of therapeutic effectiveness. This is shown in Figure 7, where "mean level" of "App" is higher than "Robot with App" on the graph.

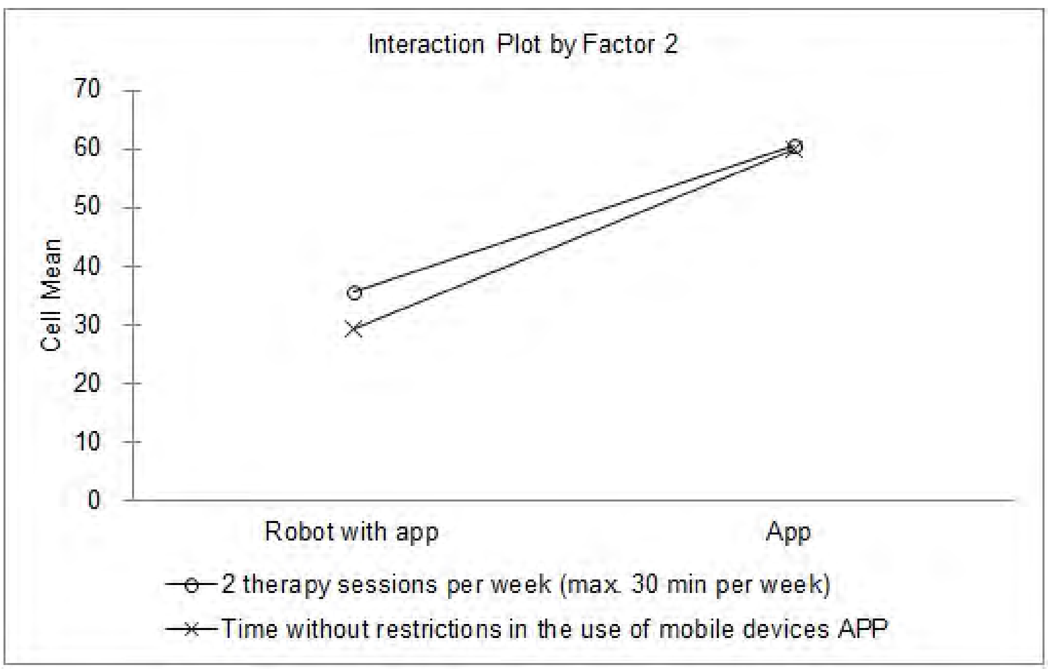

Figure 7 Interaction Plot by Factor 2: Technology Tool

Figure 8 shows the overall effectiveness for 10 weeks associated with the exposure time factor using mobile devices application. An improvement in effectiveness of approximately

12.4 percentage points can be observed taking into account the complete data series from week 1 to 10 for groups that have no time restrictions and approximately 8 percentage points for groups that use technology with the restriction of 2 sessions per week (the absolute deviation of "mean of means" was calculated following Central Limit Theorem principles).

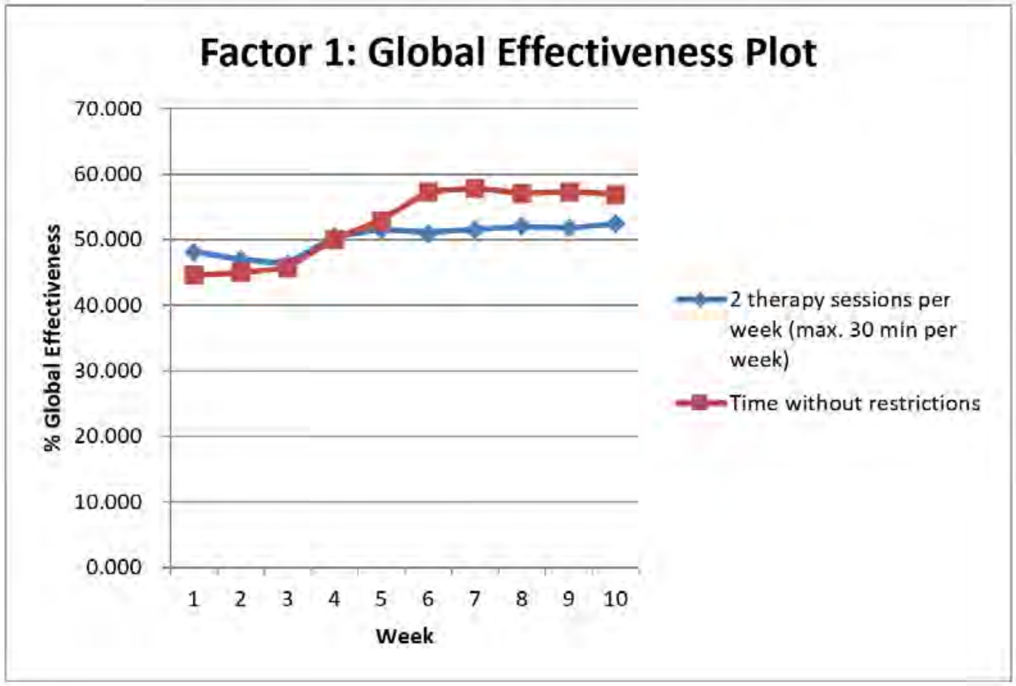

Figure 8 Overall effectiveness graph for factor 1 (time of exposure to therapeutic exercise) in weeks 1-10

Furthermore, Figure 9 shows the overall effectiveness for factor 2 (use of technologies), where an increasing trend in the effectiveness of the groups that use the physical robot with the mobile device application can be observed. This allows for an approximate increase of 13.2 percentage points, meaning an appreciable result of improvement in speech skills of the patients in these groups. It is important to clarify that the randomization of the systematic process to select and assign individuals to each group cause a specific level of effectiveness of the group from the beginning, but the trend of improvement is the core of the analysis. In this specific case, in groups that use only the mobile device application, the effectiveness is randomly higher than the use of robot with app, but remains stable over 10 weeks, due to the two factors having a statistical interaction and the time restriction possibly significantly affecting the use of the mobile device application.

Figure 10 shows the evolution of the groups that used "robot with app" over the 10 weeks. In both groups, there is an increasing trend in effectiveness for any time of use and the group that increased the most over time was the group without restriction of time of use of the platform (from 29.3% to 44.8% for an improvement of 15.5 percentage points). In the case of the group that worked with restricted time, the improvement in effectiveness was a little lower (11 percentage points) but still significant in that it rose from 35.7% to 46.7% in

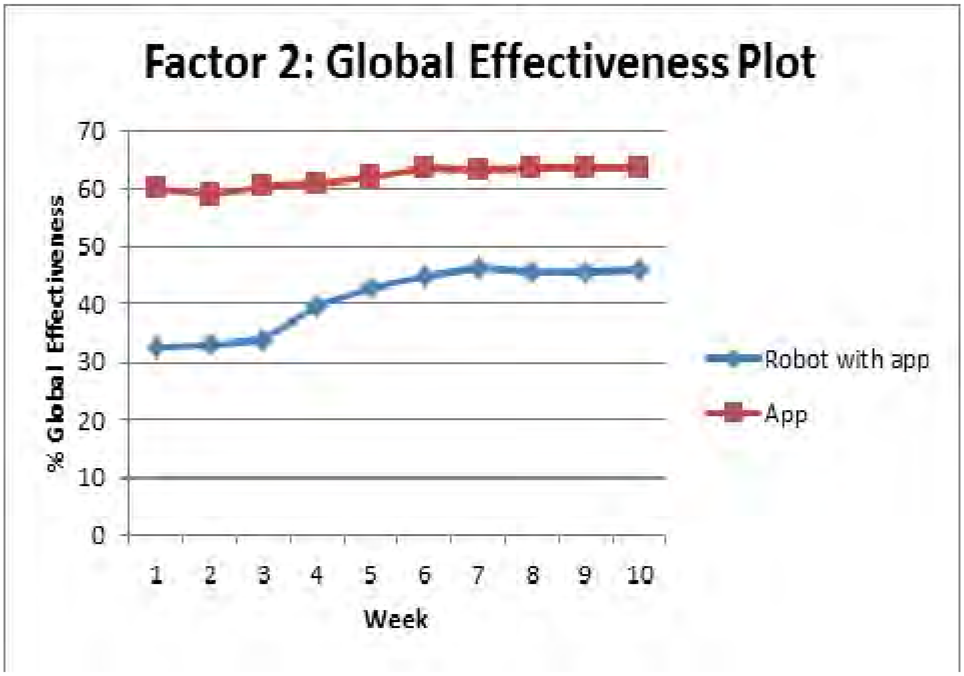

Figure 9 Overall effectiveness graph for Factor 2 over weeks 1-10

only10 weeks.

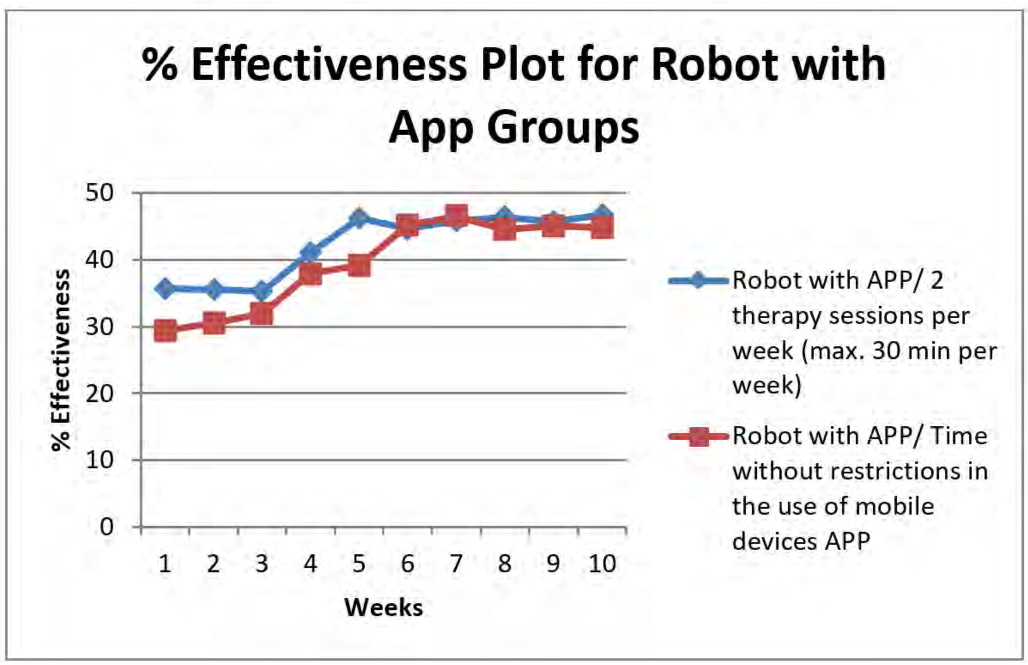

Figure 10 Effectiveness of the groups that used robot / app over weeks 1-10.

Figure 11 shows the evolution of the groups that used "only the app" over the 10 weeks. In this case, there is an increasing trend in effectiveness for the group without time of use restrictions (from 59.8% to 69.1% for an improvement of 9.2 percentage points), while the group with a restricted time showed no increasing trend.

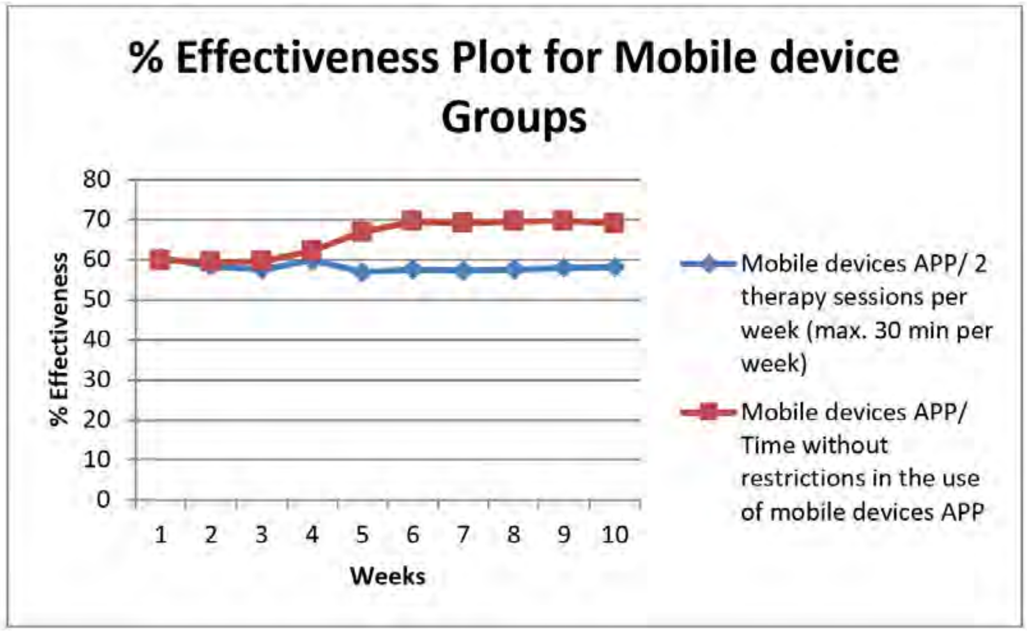

Figure 11 Effectiveness of the groups that used only the mobile device app over weeks 1-10

Table 6 shows the analysis of variance (ANOVA) for the 10 weeks of experimentation. From week 5, behavioral changes are evident in the experimental groups and this evolution is clearer in the following weeks up to week 10.

Table 6 Two-Way ANOVA: results for week # 1 to 10

	Week 1	**Week 2**	**Week 3**	**Week 4**	**Week 5**	**Week 6**	**Week 7**	**Week 8**	**Week 9**	**Week 10**
Source	*p-value*	*p-value*	*p-value*	*p-value*	*p-value*	*p-value*	*p-value*	*p-value*	*p-value*	*p-value*
Factor 1	0,330	0,474	0,795	0,894	0,537	0,019	0,028	0,037	0,021	0,044
Factor 2	3,89E-06	5,73E-07	5,22E-07	1,30E-06	1,69E-06	4,04E-06	1,94E-05	2,64E-06	1,45E-06	1,27E-06
Interaction	0,428	0,279	0,309	0,278	0,002	0,029	0,045	0,008	0,012	0,007

In addition, behavioral changes are evident in the experimental groups and this evolution is clearer in the following weeks until week 10. The groups that use the technology without time restrictions have an increasing tendency in their therapeutic effectiveness, which means that they improve for any amount of exposure time whether they use the physical robot with the application or only the computer application.

Finally, a control group was implemented with the same selection and assignment criteria, working over 10 weeks using the traditional method of speech therapy, 2 sessions per week, 30 minutes per session. The effectiveness of this group randomly begun on the range of 60% and 70% and remained the same over the experimental period, which means that no trend in therapeutic effectiveness (positive or negative) was found in this group. In other words, compared to the results obtained with the randomized control group, the statistically reliable changes evidenced in the experiment (with 95% of statistical confidence) were sufficiently meaningful to maintain scientific and educational interest in the use of assistive robotic technologies in the future.

This means that having found a difference between groups from week 5 onward, using supporting technologies, is a significant experimental finding.

6 CONCLUSIONS

The research problem focuses on whether the use of a cyber-physical robotic system (CPRS) to assist to assist the practice of special educators specialized in speech and language therapy for students with articulation disorders is beneficial.

After the design and implementation of a full factorial 2k quasi-experiment cyclically repeated over 10 weeks, some conclusions can be drawn in terms of the therapeutic effectiveness as a response variable, and two casual factors: time restriction and technological resources, with 95% statistical confidence:

1. Robotic and cyber-physical technologies may be experimental causal factors (that is, they may cause statistically significant differences in the central tendency) of greater therapeutic effectiveness on speech skills of patients, when they assist speech and language pathologists or specialized special educators in professional practice. This is important given that experimental research in this type of assistive robotics applications is not so common.
2. Over 10 weeks, the use of the platform may be advantageous for speech and language therapeutic practice to improve effectiveness. In this specific case, this improvement could be up to 11.3 percentage points, with this change being significant considering that traditional therapy without the use of these technologies and represented in the control group of patients in that same period of time, exhibits no statistically significant change.
3. Time restrictions in some types of speech therapy strategies could be a key factor. In this specific case, the longer the entire platform is used or only the application for mobile device, the better is the expected effectiveness. The positive results are significantly greater if the time of exposure to the therapeutic exercise through the assistive technology is not restricted to a number of weekly sessions and, in contrast, can be flexible and at any time of the day, reaching an improvement of 10.4 percentage points in 10 weeks. In addition, there is an increasing trend in effectiveness for the group that used only the mobile device application without time restrictions (from

59.8% to 69.1% for a 9.2 percentage points improvement) while the group with a restricted time showed no trend.

4. The use of these assistance technologies could have positive impacts but time is needed to see an improvement trend. In this specific case, it did not generate a statistically significant change during the first 4 weeks of use, but, from 5 weeks onward, statistically significant changes can be expected. A very important element of this research is that this finding is statistically proven at a 95% confidence level.

Regarding scientific methodologies, factorial experiment designs with real patients and data analysis over 10 weeks are unusual in the literature on research in speech therapy with robotic applications. However, it is possible to conclude that this kind of experiment and, specifically, the full factorial 2k designs are effective and highly accepted when conducting comparative and exploratory studies in the field of speech and language therapy, especially when the statistical requirement demands can be met.

Finally, in this specific case, compared to the results obtained with the randomized control group, the statistically reliable changes evidenced in the experiment were sufficiently meaningful to maintain scientific and educational interest in the use of assistive robotic technologies in the future. For example, it was clearly found that in groups using only the mobile device application, the effectiveness remains stable, due to the two factors having a statistical interaction and the time restriction possibly significantly affecting the use of the mobile device application.

This finding opens up a research question for future works. In addition, a key opportunity would be to improve and clinically validate the use of this platform with a virtual reality experience and in combination with a digital twin of the physical robot to improve the real life experience of the users.

7 AUTHORS' CONTRIBUTIONS

1. **Eldon Glen Caldwell Marín** served in the research and methodology roles, developing the design of the scientific method and conducting the implementation of the experiments and data/evidence collection and curation. In addition, he served in the writing-original draft role working on the creation of the published document. Also, he did the software development including the implementation of the computer code and supporting algorithms and testing.
2. **Miguel Cazorla** served in the supervision role, executing the leadership responsibility for the research including mentoring to the core team. He too, served in the conceptualization role, collaborating in the formulation of overarching research goals and, in addition, served in the writing / original draft role both in the preparation of the document and in its substantive translation.
3. **Jose Maria Cañas-Plaza** served in conceptualization role giving orientation to the evolution of overarching research goals and, in addition, played the formal analysis role, reviewing the application of statistical models and formal tests with the required

parsimony to give robustness to the discussion of results and conclusions. Likewise, he served in the writing/review editing role, specifically he did the critical review and technical criteria to solve the peer reviewers requirements.

ACKNOWLEDGEMENTS

We are deeply grateful for the collaboration of the certified speech and language pathologists and special educators, Emilia Solís Sánchez, Adriana Jones Morales and Silvia Soto Rojas. In addition, we appreciate the support of the Heredia Special Education Center staff in Costa Rica and the contributions of professor Carlos Andrés Morales and Eldon Wesley Caldwell Salazar.

REFERENCES

American Speech Language Hearing Association. (2018). *Aphasia Definitions*. Retrieved from http://www.asha.org/public/speech/disorders/ChildSandL.htm

Begum, M., Serna, R. W., & Yanco, H. A. (2016). Are Robots Ready to Deliver Autism Interventions? A Comprehensive Review. *International Journal of Social Robotics*, *8*(3), 157–181. https://doi.org/10.1007/s12369-016-0346-y

Cabibihan, J. J., Javed, H., Ang, M., & Aljunied, S. M. (2013). Why robots? A survey on the roles and benefits of social robots in the therapy of children with autism. *International Journal on Social Robotics*, *5*(4), 593–618.

Calderita, L. V., Bustos, P., Mejías, C., Fernandez, F. S., Viciana, R., & Bandera, A. (2015). Socially Interactive Robotic Assistant for Therapies. *Revista Iberoamericana de Automática e Informática industrial*, *1*(1), 99–110. https://doi.org/10.1016/j.riai.2014.09.007

Caldwell, E., Cazorla, M., García, J., Azorin, J., & Zamora, M. (2017). An exploratory critical review on assistive robotics applied to autism spectrum: employability challenges. *2017 International Symposium on Industrial Engineering and Operations Management (IEOM)* (pp. 24–25). Retrieved from http://www.ieomsociety.org/ieomuk/

Caldwell, E., Morales, C., Solis, E., Cazorla, M., & Cañas, J. M. (2021). Designing a Cyber-physical Robotic Platform to assist Speech and Language Pathologists. *Assistive Technology: the official journal of RESNA*. https://doi.org/10.1080/10400435.2021.1934609

Chaminade, T., Fonseca, D., Rosset, D., Lutscher, E., Cheng, G., & Deruelle, C. (2012). FMRI study of young adults with autism interacting with a humanoid robot. *The 21st IEEE International Symposium on Robot and Human Interactive Communication* (pp. 380–385). https://doi.org/10.1109/ROMAN.2012.6343782

Choo, A. L., Smith, S. A., & Li, H. (2022). Prevalence, severity and risk factors for speech disorders in US children: the National Survey of Children's Health. *Journal of Monolingual and Bilingual Speech*, *4*(1), 109–126. https://doi.org/10.1558/jmbs.20879

Collins, L. M., Dziak, J. J., & Li, R. (2009). Design of Experiments with Multiple Independent Variables: A Resource Management Perspective on Complete and Reduced Factorial Designs. *Pychology Methods*, *14*(3), 202–224. https://doi.org/10.1037/a0015826

Costa, S., Lehmann, H., Dautenhahn, K., Robins, B., & Soares, F. (2015). Using a Humanoid Robot to Elicit Body Awareness and Appropriate Physical Interaction in Children with Autism. *International Journal of Social Robotics*, *7*(4), 265–278. https://doi.org/10.1007/s12369-014-0250-2

Costescu, C. A., Vanderborght, B., & David, D. O. (2015). Reversal learning task in children with autism spectrum disorder: a robot-based approach. *Journal of Autism Dev.Disorder*, *45*(11), 3715–3725. https://doi.org/10.1007/s10803-014-2319-z

Creswell, J. W., & Creswell, J. D. (2017). *Research Design: Qualitative, Quantitative and Mixed Methods Approaches* (7th ed.). Sage Publications.

Diehl, J., Crowell, C. R., Villano, M., Wier, K., Tang, K., & Riek, L. (2014). *Clinical applications of robots in autism spectrum disorder diagnosis and treatment: a comprehensive guide to autism*. Springer International Publishing. Retrieved from https://doi.org/10.1007/978-1-4614-4788-7_14 https://doi.org/10.1007/978-1-4614-4788-7_14

Estevez, D., Terrón-López, M. J., Velasco-Quintana, P. J., Rodríguez-Jiménez, R. M., & Álvarez Manzano, V. (2021). A case study of a robot-assisted speech therapy for children with language disorders. *Sustainability*, *13*(5), 2771–2771. https://doi.org/10.3390/su13052771

Flagge, N. M. (2013). Trastornos del lenguaje: diagnóstico y tratamiento [Language disorders: diagnosis and treatment. *Revista Neurología*, *57*(1), 85–94. https://doi.org/10.33588/rn.57S01.2013248

García-Vergara, S., Brown, L., Park, W., H, & Howard, A. M. (2014). *Engaging Children in Play Therapy: The Coupling of Virtual Reality Games with Social Robotics*. Springer. https://doi.org/10.1007/978-3-642-45432-5_8

Glogowska, M. (2011). Paradigms, pragmatism and possibilities: mixed-methods research in speech and language therapy. *International Journal of Language and Communication Disorders*, *46*(3), 251–260.

Gregor, W. (2016). Employment, Disabled People and Robots: What Is the Narrative in the Academic Literature and Canadian Newspapers? *Societies Journal*, *6*(15), 2–16. https://doi.org/10.3390/soc6020015

Jackson, M., & Cox, D. R. (2013). The Principles of Experimental Design and Their Application in Sociology. *Annual Review of Sociology*, *39*(1), 27–49. https://doi.org/10.1146/annurev-soc-071811-145443

Kim, E. S., Berkovits, L. D., Bernier, E. P., Leyzberg, D., Shic, F., Paul, R., & Scassellati, B. (2013). Social robots as embedded reinforcers of social behavior in children with autism. *Journal of Autism Dev. Disorders*, *43*(1), 1038–1049. https://doi.org/10.1007/s10803-012-1645-2

Kirk, R. E. (2014). *Completely Randomized Factorial Design with Two Treatments, In Experimental Design, Procedures for the Behavioral Sciences*. SAGE Publications.

Kuehl, R. O. (2001). *Design of Experiments: Statistical Principles for research design and analysis* (2nd ed.). Thompson Learning Ed.

Lewison, G., & Carding, P. (2003). Evaluating UK research in speech and language therapy. *International Journal of Language and Communication Disorders*, *38*(1), 65–84. https://doi.org/10.1080/13682820304815

Liu, L., Li, B., Chen, I. M., Goh, T., & Sung, M. (2014). Interactive robots as social partner for communication care. *2014 IEEE International Conference on Robotics and Automation*. https://doi.org/10.1109/ICRA.2014.6907167

Marge, M., Espy-Wilson, C., Ward, N. G., Alwan, A., Artzi, Y., Bansal, M., & Yu, Z. (2022). Spoken language interaction with robots: Recommendations for future research. *Computer Speech & Language*, *71*. https://doi.org/10.1016/j.csl.2021.101255

Medina-Varela, P. D., & López-Reyes, A. (2011). Análisis Crítico del Diseño Factorial 2k sobre casos aplicados. *Scientia Et Technica*, *17*(47), 101–106.

Montgomery, A. A., Peters, T. J., Little, P., & L. (2003). Design, analysis and presentation of factorial randomised controlled trials. *BMC Medical Research Methodology*, *3*(1), 26–42. https://doi.org/10.1186/1471-2288-3-26

Montgomery, D. C. (2002). *Experiments Analysis and Design* (2nd ed.). LIMUSA-Wiley Co.

Pakrasi, I. L. (2018). *Towards expressive mobile robots* (Unpublished master's thesis). University of Illinois.

Pakrasi, I. L., Laviers, A., & Chakraborty, N. (2018). A design methodology for abstracting character archetypes onto robotic systems. *5th International Conference on Movement and Computing (MOCO)* (pp. 28–30). https://doi.org/10.1145/3212721.3212809

Pennington, R., Saadatzi, M. N., Welch, K. C., & Scott, R. (2014). Using robot-assisted instruction to teach students with intellectual disabilities to use personal narrative in text messages. *Journal of Special Education Technology*, *29*(4), 4958–4972. https://doi.org/10.1177/016264341402900404

Pulido, G., Vara, H. D. L., & Salazar, R. (2012). *Análisis y Diseño de Experimentos* (3rd ed.). McGraw-Hill.

Pulido, J. C., González, J. C., Suárez-Mejías, C., Bandera, A., Bustos, P., & Fernández, F. (2017). Evaluating the child-robot interaction of the NAOTherapist platform in pediatric rehabilitation. *International Journal of Social Robotics*, *9*(3), 343–358. https://doi.org/10.1007/s12369-017-0402-2

Raul, F., & Ahyea, A. (2016). Differentiating Language Difference and Language Disorder: Information for Teachers Working with English Language Learners in the Schools. *Journal of Human Services: Training, Research and Practice*, *2*(1), 1–22.

Robles-Bykbaev, V., Velásquez-Angamarca, V., Mosquera-Cordero, K., Calle-López, D., Robles-Bykbaev, Y., Pinos-Vélez, E., & León-Pesántez, A. (2018). A proposal of a virtual robotic assistant and a rule-based expert system to carry out therapeutic exercises with children with Dyslalia. *2018 IEEE Third Ecuador Technical Chapters Meeting (ETCM)* (pp. 15–19). https://doi.org/10.1109/ETCM.2018.8580302

Scassellati, B., Admoni, H., & Mataric, M. (2012). Robots for use in autism research. *Annual Rev. Biomed. Eng*, *14*(2), 275–294. https://doi.org/10.1146/annurev-bioeng-071811-150036

Scassellati, B., & Tsui, K. (2016). Co-Robots: Humans and Robots Operating as Partners. *Handbook of Science and Technology Convergence* (pp. 427–439). Springer International Publishing. https://doi.org/10.1007/978-3-319-07052-0_27

Seong, C., Jin, & Hyun, A. D. (2016). Socially Assistive Robotics in. *Autism Spectrum Disorder. Hanyang Medical Review*, *36*(2), 17–26. https://doi.org/10.7599/hmr.2016.36.1.17

Smith, C., Williams, E., & Karen, B. (2017). A systematic scoping review of speech and language therapists public health practice for early language development. *International Journal of Language and Communication Disorders*, *52*(4), 407–425. https://doi.org/10.1111/1460-6984.12299

Tapus, A., Peca, A., Aly, A., Pop, C., Jisa, L., Pintea, S., … David, D. O. (2012). Children with autism social engagement in interaction with NAO, an imitative robot. *Interact Stud*, *13*(3), 315–347. https://doi.org/10.1075/is.13.3.01tap

Vázquez-Villasuso, V., & Diaz-Monterrey, M. (2015). Hablando acerca de lenguaje. *Revista Cubana de Tecnología de la Salud*, *6*(4), 121–125.

W F Harley, J., & Harley, W. F. (1968). The Effect of Hypnosis on Paired-Associative Learning. *Journal of Personality*, *36*(6), 147–172. https://doi.org/10.1111/j.1467-6494.1968.tb01478.x