Impact of an Embodied Conversational Agent on Loneliness and Emotional States in Community Dwelling Older Adults: Initial Quantitative and Qualitative Evidence

Feliciano Villar*[1] and Rumen Manolov[2]

[1] Department of Social Psychology and Quantitative Psychology, Faculty of Psychology, University of Barcelona, Spain

[2] Experimental Health Psychology, Faculty of Psychology and Neuroscience, Maastricht University, the Netherlands

* **Corresponding author**

Prof. Feliciano Villar, PhD

Department of Cognition, Development and Educational Psychology; Faculty of Psychology; University of Barcelona, Passeig de la Vall d'Hebron 171, 08035 Barcelona, Spain

E-mail: fvillar@ub.edu

ORCID: 0000-0003-0830-7034

Rumen Manolov, PhD

Department of Social Psychology and Quantitative Psychology; Faculty of Psychology; University of Barcelona, Passeig de la Vall d'Hebron 171, 08035 Barcelona, Spain

E-mail: rrumenov13@ub.edu

ORCID: 0000-0002-9387-1926

Author Contributions

Conceptualization: FV. Design: FV and RM. Data collection: using the conversational bot and research assistants. Qualitative data analysis: FV. Quantitative data analysis: RM. Manuscript preparation: FV and RM.

Statements and Declarations

Declaration of Conflict of Interest

The authors declare that they received funding from the company that developed AiMA, the ECA tested in this study. However, the authors have no vested interest in obtaining any particular results, beyond reflecting reality as accurately as possible. Therefore, the data collection and analysis procedures follow the rigorous standards of single-case methodology.

Ethical Approval and Informed Consent (to Participate and for Publication)

Ethical approval was obtained from the University of Barcelona Bioethics Committee (IRB 00003099) on 10th October 2024. Informed consent was obtained from the participants.

Data Availability

The quantitative data, R code, and results are available at https://osf.io/nm625/.

Abstract

The current study explores the implementation of an embodied conversational agent called AiMA at home, for community-dwelling older adults and its impact on loneliness and emotional states of older adults. Seventeen community-dwelling older adults (12 women, 5 men) líving alone participated in the study. An embodied conversational agent was implemented in their homes for 15 days. The focus of the study is placed on both qualitative and quantitative aspects. In terms of the quantitative information, the amount of interaction (days, conversations, turns) was summarized, as well as the percentage of joy and sadness. Specifically, it was found that the most frequent transition from the initial to the final emotional state across interactions, for all participants, was from neutral to joy. In qualitative terms, participants answered an open-ended interview in which positive aspects reported of using the device and areas of improvement, were included. Results show that participants engaged and tended to engage with the ECA in an anthropomorphic way, in some cases showing a 'human' bond with the device. Despite both emotional and cognitive benefits were mentioned, the emotional positive impact of the interaction with the ECA (reduction of loneliness, elicitation of positive emotions) as the most frequently mentioned one. When mentioning areas of improvement, one of the aspects that participants missed was the lack of proactivity. Findings showed that participants engaged in interactions with the conversational agent, which suggests that it could a useful tool to mitigate loneliness and enhance positive emotions among older adults living alone.

Keywords: Conversational Agent, Older Adults, Loneliness, Emotions

Introduction

The population aged 60 and older is expanding rapidly across all global regions. This demographic segment is projected to double by the year 2050, while the population aged 80 and older is estimated to triple current figures by that same date (WHO, 2022).

Population aging poses multi-level challenges. Among them, the number of older adults who live alone or experience loneliness is expected to rise in the future. Loneliness is defined as a subjective perception of a lack of social contact. That is, individuals perceive a discrepancy between their actual and desired levels of social interaction, or they find their existing social relationships unsatisfactory (Cacioppo, 2010; De Jong Gierveld, 1987). This perceived deficit is accompanied by aversive feelings such as sadness or anxiety, and it is associated with numerous health problems (Gerst-Emerson & Jayawardhana, 2015). In contrast, social isolation represents an objective lack of social contact with others (Burholt et al., 2020). Although both phenomena are interrelated, they can occur independently: some individuals may feel lonely even when surrounded by many people, or conversely, other ones having very few social contacts may live without experiencing feelings of loneliness.

According to data from the Spanish National Statistics Institute (Pérez et al., 2023), the population over 65 in Spain reached nearly 9.5 million people, accounting for approximately 20% of the total population. Among them, more than one in five (2.131 million older adults) lived in single-person households, thereby placing them at risk of social isolation. These single-person households were particularly prevalent among women (31.9% lived alone) compared to men (16.1%). Regarding loneliness, although older adults are not the age group that experiences this situation most frequently, it is estimated that around 16% report having felt lonely within the past month (SoledadES, 2023). In another study focusing exclusively on older individuals (Fundación la Caixa, 2018), feelings of loneliness were estimated to affect 68% of the sample, although these feelings were classified as severe or very severe in only 15% of the cases. In later life, the death of loved ones, children leaving the parental home, an increased probability of health problems, the presence of disability (particularly mobility issues) and dependency, alongside a lack of accessibility in the physical environment, constitute risk factors that can precipitate or exacerbate loneliness among older adults (Burholt et al., 2020; Dahlberg et al., 2018).

Loneliness can exert highly pernicious effects on older adults across multiple dimensions. For instance, loneliness is linked to poorer health status and correlates with the presence of various chronic conditions, such as cardiovascular disease, hypertension, pulmonary disease, obesity, diabetes, and sleep disturbances (Komulainen et al., 2026; Zheng et al., 2026). Similarly, from a psychological and behavioral perspective, loneliness is associated with elevated levels of depression and anxiety, a higher risk of suicide, and the onset of dementing processes (Akinyemi et al., 2025; Donovan & Blazer, 2020;). From an economic viewpoint, several studies have quantified the costs attributable to loneliness. In Spain, Casal et al. (2023) estimated the tangible costs of loneliness, including reduced productivity, primary and specialized healthcare expenses, and increased medication consumption, at 14 billion euros, representing 1.17% of the Spanish GDP. To this must be added the loss of quality-adjusted life years (QALYs) and premature mortality associated with loneliness.

Interventions Addressed to Reduce Loneliness

Loneliness has emerged in recent years as a major public health concern, the relevance of which has driven the development of interventions aimed at mitigating loneliness among older adults.

These interventions vary significantly in nature (Hoang et al., 2022; Poscia et al., 2018). Some that have demonstrated efficacy in reducing loneliness levels include, among many others, programs designed to promote physical activity (Levinger et al., 2020; Shvedko et al., 2018) or foster social participation among older populations (Bartlett et al., 2013; Martina & Stevens, 2006), as well as programs involving regular visits to solitary older individuals (Hsu, 2024). Such initiatives can be designed for implementation within either community or institutional settings (Quan et al., 2020).

However, these programs are not always readily available, and barriers such as cost, difficulty in reaching isolated individuals, the presence of mobility limitations or dependency, and the older adults' own interest and investment of time and effort can restrict their scope. In the case of visitation programs, for instance, the availability of social contact may be limited and does not depend directly on the person experiencing loneliness.

In this regard, the utilization of technology as a complementary element or as a means to overcome some of these barriers is becoming increasingly frequent, providing companionship on user demand. These technological solutions have been applied with positive outcomes both among community-dwelling older adults (Ambagtsheer et al., 2024) and individuals with dementia residing in long-term care facilities (Budak et al., 2023).

Nevertheless, technology-based interventions are also highly diverse (Choi & Lee, 2018). The earliest programs aimed to teach computer skills to older adults to enable the autonomous use of the internet or electronic messaging applications (Czaja et al., 2018; Jones, 2015). More recently, interventions have leveraged the potential of social networks or instant messaging applications to connect older adults with others (e.g., Barbosa Neves et al., 2019), as well as the use of virtual reality as a tool to overcome isolation and feelings of loneliness (Veldmeijer et al., 2020; Finnegan & Campbell, 2023).

Among technology-based interventions, the implementation of social robots or robotic pets (in many cases, embodied as robotic pets, such as dogs or seals) is particularly prominent. These devices allow for a certain degree of interaction with the older adult while simultaneously gathering specific parameters that can contribute to healthcare management (Koceska et al., 2019; Koceski & Koceska, 2016).

This type of social robot has demonstrated efficacy in reducing feelings of depression and loneliness in older populations (Abdi et al., 2018; Chen et al., 2020; Dang & Bertrandias, 2025). This positive impact appears to occur both among healthy, community-dwelling older adults (McGlynn et al., 2017) and, notably, among older individuals living with dementia, whether at home (Inoue et al., 2021) or in residential care settings, where the need for companionship can be especially acute (Shoesmith et al., 2024).

In addition to providing companionship, being more predictable and controllable, and, under certain circumstances, offering a more continuous presence than can be achieved through a human social network, the effects of social robots on loneliness stem largely from the human capacity to form affective bonds with them. This attachment promotes communication and enhances feelings of well-being among older adults, which particularly benefits those experiencing emotional loneliness (Deng & Bertrandias, 2025). Specific design elements, such as emotionally expressive faces or soft, "huggable" forms that provide sensory feedback, further facilitate this bonding process (Christiansen et al., 2024).

In contrast, studies on social robots also identify several limitations. Perhaps the most significant of these concerns the high purchasing and maintenance costs of the robots, which can in practical terms limit their deployment and exacerbate inequalities among older adults (Deusdad, 2024). Furthermore, from an ethical standpoint, scholars have noted the risk of infantilization posed by the most widely used social robot designs; these may undermine the dignity of older adults if they feel they are being treated like children (Hung et al., 2025). In the case of individuals with dementia, the user may perceive the robot as a real, living being, potentially fostering a disconnection from reality and reinforcing a deceptive relationship (Hung et al., 2025; Sharkey & Sharkey, 2021).

Partly in response to these drawbacks, systems based on artificial intelligence (AI) have been proposed as a means to open new avenues for loneliness interventions using technological tools.

Chatbots, Conversational Agents, and Older Adults

Chatbots are software applications designed to interact and respond contingently with users. Research on the impact of chatbot interaction as a means to alleviate loneliness and improve other psychological outcomes is growing, and existing review studies offer highly promising results in this regard (Oh et al., 2026; Sha et al., 2024; Teo et al., 2025). However, it is critical to note that chatbots can be categorized into distinct types.

To date, research has predominantly focused on the impact of **voice assistants**, such as Alexa or Siri. This type of technology, which can be deployed on mobile phones, tablets, or smart speakers, enables users to interact using natural language with a technology that integrates seamlessly into daily life. Because it is intangible, it does not require users to physically handle or manipulate any device for the interaction to occur. Consequently, these systems can be utilized by individuals lacking computer literacy or those experiencing sensory or mobility impairments.

Although voice assistants were originally designed to respond to user commands such as controlling smart home automation systems, providing information on various topics (e.g., Rodríguez-Martínez et al., 2023), or detecting needs, reminding, and monitoring health- and care-related tasks (Su et al., 2022), they have also been tested as tools to enhance well-being and reduce loneliness (Oh et al., 2026).

Regarding loneliness, available empirical evidence suggests that older adults' interaction with voice assistants contributes significantly to reducing emotional loneliness, increasing feelings of companionship, and improving various dimensions of well-being (Corbett et al., 2021; Marziali et al., 2024; Yan et al., 2024).

Nevertheless, the most frequently used voice assistants offer a relatively low degree of conversational "naturalness", as they are primarily engineered to fulfill specific user demands (e.g., playing music, adjusting volume, or answering precise queries) rather than engaging in casual conversation or establishing an affective bond with the user (Even et al., 2022). These limitations, coupled with usability challenges and a steep learning curve required to truly "converse" with voice assistants, constitute a major source of frustration and difficulty frequently highlighted in previous research (Teo et al., 2025).

In response to these limitations, relational agents have emerged that extend beyond voice assistants, characterized by their intent to build a relationship with the user through conversation. These are known as conversational agents. Relational agents facilitate a bond with the user through two core characteristics:

- Conversational memory: This feature enables the agent to gradually "know" the user by remembering their preferences, biographical events, and personal characteristics shared during past interactions, thereby rendering conversations substantially more personalized and natural.
- Empathy based on user emotion detection: Certain agents possess the capability to detect basic human emotions (e.g., sadness, anger, joy) and respond accordingly, which facilitates emotional attunement with the user and humanizes the interaction.

Additionally, unlike most voice assistants, some relational agents manifest via a face or a virtual physical presence capable of transmitting non-verbal cues aligned with the content and tone of the conversation. These are termed **Embodied Conversational Agents (ECAs).** That is, they are applications featuring a human-like interface displayed on a screen (such as a tablet, smartphone, or television) and connected to artificial intelligence systems that enable them to ask and answer questions in a manner akin to human interaction.

This sensory presence allows users to "anthropomorphize" the agent, thereby facilitating meaningful conversation and interaction as though it possessed human qualities. This anthropomorphizing has proven to be a critical attribute in fostering a sense of companionship and reducing feelings of loneliness (Jones et al., 2024; Sun & Chen, 2026).

The available literature regarding the impact of ECAs on older adults remains scarce and, thus far, inconclusive. One of the earliest examples of ECA designed to establish a bond and a relationship of trust with the user was Laura, an ECA oriented toward promoting physical exercise among older adults through an avatar capable of interacting in a friendly manner. Older adults interacting with Laura established a pleasant and meaningful relationship with the ECA; however, Laura lacked the capacity to detect and respond to user emotions (Bickmore, 2010).

More recently, PACO, an ECA designed to train users in healthy habits, yielded poor results in mitigating loneliness among the older adults who interacted with it. Despite this, a positive relationship was found between the time spent interacting with PACO and the users' perceived utility and enjoyment of the interaction (Kramer, 2022). Other studies evaluating similar ECAs, such as Bella (Lovelys et al., 2021), also failed to obtain robust evidence regarding improvements in loneliness and mood, even though the system was generally well-received by users.

The limited body of research on the impact of chatbot interaction (whether voice assistants or ECAs) on loneliness or other emotional dimensions in older adults typically explores this impact cross-sectionally, evaluating users' subjective feedback or effects after a certain period of device exposure, or utilizes pre-post longitudinal designs (generally involving brief intervention periods of a few weeks). In these longitudinal designs, a battery of instruments is administered to measure key variables (e.g., loneliness scales or affective state) prior to introducing the intervention (i.e., the implementation of the technological device) and at a specified time post-intervention. Changes observed between pre- and post-test measures, or the comparison of these changes with a control group where no chatbot is introduced, are attributed to the intervention effect.

While this type of quasi-experimental design provides highly valuable data, it is subject to specific biases (e.g., reactivity effects) and, crucially, fails to capture how change occurs, obscuring the process the individual undergoes from the initial to the final time point. In the case of interventions based on user-device interaction, capturing process-oriented data could yield essential insights into the mechanisms underlying intervention success or failure, ultimately informing strategies to maximize their beneficial effects.

Objectives

The present study aims to evaluate the impact of interactions with an ECA, named AiMA, among community-dwelling older adults living alone. AiMA is an ECA based on a digital interface resembling a human woman with facial expressions, voice, memory and other features (see figure 1). AiMA, for the user, is a familiar face in an Android tablet that listens, thinks and talks. The users don't need to know about technology or how to turn on a computer. the person that uses AiMA only needs to be able to hear and talk normally, while AiMA responds dynamically and in a personalized manner to user inputs. The system enables continuous data collection regarding established interactions and detects verbal behavioral indicators of the user’s emotional state. These features facilitate a comprehensive characterization of the developmental trajectory defining the user-ECA relationship.

INSERT FIGURE 1

To this end, the study pursues three specific objectives:

1. To explore the impact and temporal trajectory of change across various emotional and interactive behavioral indicators in older adults living in single-person households during their interaction with an ECA such as AiMA.
2. To estimate the impact of ECA interaction on the levels of loneliness and other emotional states among older adults living in single-person households.
3. To ascertain the perceived impact, as well as the degree of user acceptance and usability, of an ECA like AiMA as evaluated by the participants themselves.

Method

Participants

Seventeen people (12 women, 5 men) participated in this study. Participants' ages ranged from 63 to 95 years old (M = 82.4 years). Nine participants had no formal studies or have not completed primary studies. Just two one participant had university studies and another one secondary studies. Table 1 provides a sociodemographic description of the participants.

INSERT TABLE 1 ABOUT HERE

There were two inclusion criteria for the study: (1) individuals aged 60 years or older; (2) individuals who live alone; (3) absence of communication, cognitive, respiratory or other health problems that would prevent participation in an interview or conversation; and (4) participants received home care support for instrumental activities of daily living, not for basic activities.

However, some participants referred some mobility problems (10 participants), hearing problems (three participants, problems compensated by hearing aids) and cancer (one participant).

Procedure and Ethics

Authors 1 of this study contacted with the directors of two home care organizations in the area of Barcelona (Spain). They informed the managers about the aims of the study and, after receiving the agreement to participate, the managers listed potential participants who met the inclusion criteria. The organization contacted with potential participants and asked them if they were interested in participating in the study.

Interested participants were visited by the research team, who installed the ECA in the place in which participants supposed it was easier to interact with it (usually, the dining room). Researchers explained the functioning of the conversational agent (basically, how to switch it on and off) and completed some test conversations until the participant can conversated. Before the installation, participants provided written informed consent. The form included detailed information about the study's objectives, data collection methods, and issues of anonymity and confidentiality. Participants were also informed of their right to withdraw from the study at any time.

After 15 days with the conversational agent, the research team visited again participants at their homes and retired the conversational agents. At this moment, they applied the semi-structured interview and the questionnaires. The interview was audio taped and the average length of the recordings was around 15 minutes. As the recordings were available, interviews were transcribed verbatim (author 3).

Data were pseudonymized before analysis, and these pseudonyms are used in the presentation of the results. The study was approved by the Bioethics Committee of the University of Barcelona (IRB 00003099) on 10th October 2024.

Quantitative Study

Measures and Data Collection

Measurements were taken for each person over several consecutive days. To obtain more precise information, the measurements (variables) of interest were not accumulated daily but rather broken down by interaction. To this end, it was necessary to distinguish, within a single day, the different interactions (i.e., when an interaction began and when it ended). One way to distinguish between interactions would be based on content: a change of topic could indicate the beginning of a new interaction.

However, in this study, the content of the interactions between the participants and the ECA was not recorded due to data privacy concerns. Therefore, it was necessary to define the boundaries of the interactions based on nonverbal information. Silence (i.e., the absence of words from both the participant and the ACA) was chosen as the marker. Specifically, ten seconds of silence was set as the period of time that would indicate that an interaction had ended. This duration is necessarily arbitrary, but it is based on the methodology used to obtain data through direct observation (Yoder & Symons, 2010).

Data collection through observation can be carried out through continuous recording (e.g., recording what happens every second: which behavioral category is taking place), event-based recording or transition-triggered recording (i.e., recording each time a category change occurs), or through predetermined intervals. If working with intervals, one can record which category occurs at the end of the interval (momentary time sampling), which categories occurred at some point during the interval (partial interval sampling), or which category, if any, occurred throughout the entire interval (whole interval sampling). The durations of these intervals are also arbitrary and depend on the typical duration of the behavior being observed. The most common practice is to work with intervals between 5 and 30 seconds, with 10 seconds being one of the most frequently used (Pustejovsky & Swan, 2015; Radley et al., 2015; Yoder et al., 2018). In this sense, although observation intervals were not used in the present study, the duration of the intervals was adopted to determine the time between turns that marks a new interaction.

A basic quantitative description of the interaction between participants and the ECA was performed by quantifying, for each interaction, number of turns the person takes and number of words the person uses.

Additionally, the device gathered information about emotion: neutral, joy, anger, sadness, fear, disgust, and surprise. Specifically, the following data was gathered: most likely emotion in the person's first turn, most likely emotion in the person's last turn, most likely emotion throughout the entire interaction, and percentage of time each emotion predominates within the interaction. The emotion was detected by the ECA automatically by means of transcribing the audio and working with the resulting. Specifically, the natural language processing task is called “emotion recognition in conversation” and its aim is to identify and classifying the emotional states expressed in text (Mohammad, 2016; Porja et al., 2019). Transformer-based architectures models such as EmoBERTa (Kim & Vossen, 2021) fine-tune RoBERTa variants for this task. In the current research, a model of the RoBERTa family, specialized in Spanish and called RoBERTuito (Pérez et al., 2022). The choice responds to three domain requirements: the corpus is in Spanish, the conversations are free unstructured text, and a granular emotion taxonomy is needed to capture affective variations with distinct implications in the support setting. Further details about emotion recognition can be consulted from the Open Science Framework project, associated with the study (https://osf.io/nm625/), with the specific document available from https://osf.io/7vdnt.

Once the data collection method has been described, it is necessary to clarify that it partially follows the principles of experiential sampling or ecological momentary assessment (Shiffman et al., 2008; Stinson et al., 2022). Specifically, it involves obtaining samples of a behavior repeatedly in real time, within the participants' natural environment. This data collection method offers several advantages. First, it avoids placing a burden on the participant (no interruptions to their daily activities or reminders are needed) and also avoids potential biases in recall or reactivity, as it does not use self-report measures. Second, it maximizes ecological validity. Third, it allows aggregating measures over time (e.g., combining information from different interactions that occurred on the same day), thereby increasing reliability through data aggregation. Fourth, the fact that the device records the data of interest without user intervention allows individuals who have difficulty with self-report measures or technology to participate in the study (and be represented in the sample).

Finally, given that the data are recorded directly by the device, it is not necessary to assess inter-rater agreement: another important characteristic (Hausman et al., 2022; What Works Clearinghouse, 2022), which is frequently evaluated in the context of user-centered care studies characteristic (Essig et al., 2023. This is because the data are collected without human intervention and without knowledge of the research objective or expectations.

Design

Overall approach. The design used corresponds to an idiographic, person-centered approach to identifying patterns in the processes studied (De Luca Picione, 2015). The aim is to avoid incorrect inferences (e.g., the ecological fallacy) by extrapolating conclusions to individuals from group averages (Piccirillo et al., 2019; see also the “ergodicity” assumption; Fisher et al., 2018; Molenaar, 2004). The study was not preregistered, but in the current report we adhere to the spirit of preregistration (Johnson & Cook, 2019), by describing both the initial design and analysis plan and the subsequent modifications it. This kind of reporting is carried out with the aim to avoid questionable research practices (Tincani et al., 2025).

Actual design used. Due to technical and practical difficulties, it was not possible to use the ECA in a proactive mode. A thorough description of the initial plan can be found in the Appendix. In that sense, all measurements for all participants were made in a single condition: a reactive mode. Thus, it was not possible to perform comparisons either with an actual baseline condition (without the ECA, which is the device gathering part of the data, except for the one collected via the interviews) or to a more intense presence of the ECA (in a proactive mode). The lack of distinction between phases means that it is not possible to refer to the design as either experimental or pre-experimental (Tate et al., 2016) classification. It is only justified to say we used single-case methodology. Therefore, just as a SCED, the study is characterized by the intensive, longitudinal study of one or more units (usually individuals, but they can also be entire groups), but lacking active manipulation of conditions and not allowing for any causal inference. Despite its name, studies following single-case methodology (also known as N=1 or within-subject replication designs) typically include more than one person (e.g., Tanious & Onghena, 2021, report a median of three or four people, depending on the type of study). This is also the case in the current study, including a total of 18 participants.

Due to difficulties with access, it was not possible to form clear cohorts with four groups of participants stating exactly at the same time. However, several participants began their participation in the study simultaneously (three in June 2025, four in November 2025, six in February 2026, and two in March 2026), with the remaining participants also enrolled during the period June 2026 – April 2026. In that sense, replication was still possible and the total number of participants taking part in the study exceeds what is typical in a SCED.

Data Analysis

Overall approach. Gathering data longitudinally allows for studying the evolution (esp. of emotions) in time. Given that there is not a single gold standard, we decided to check whether different data analytical approaches (e.g., a multilevel model, a logistic model, and a Bayesian unknown change point model) would converge. Thus, we followed a multiverse approach (Steegen et al., 2016), considering the researcher degrees of freedom when making data analytical choices (Wicherts et al., 2016). When results are obtained via several different analytical options it is crucial to report all results, avoid selective reporting and include even negative results in case they happen to appear (Kratochwill et al., 2018). As stated in relation to the design, also for the data analysis, the overarching aim is to avoid questionable research practices (Tincani et al., 2025).

Usage of the device and association with emotions. The first part of the analysis includes a general description of the usage of the ECA. Specifically, for each individual, the following information is summarized: (a) total days of use and the maximum number of consecutive days of use; (b) the number of days with more than one interaction, as well as the average number of interactions per day; (c) the number of (minimal) interactions that entail only two turns, as well as the average number of turns per interaction; and (d) percentage of dominance of the primary emotions during the interactions.

Additionally, associations between use and emotions were quantified. Specifically, for each person, using the data for the different interactions, Pearson's correlation coefficient was computed between number of turns and emotions (joy and sadness). The aim was to explore whether there is a relation between emotion and interaction length. Moreover, for all users together, we explored if there was a correlation between emotions (joy and sadness) and the amount of ECA use (number of days, number of days with more than one interaction, total interactions, and average number of turns per interaction). The aim was to explore whether there is a relation between emotion and amount of use of the device.

Quantification of the change via a multilevel model. Given that there were no distinct phases, the multilevel analysis (Ferron et al., 2009; Moeyaert et al., 2014) initially planned was modified to serve as an across-cases quantification of the change in emotions (joy and sadness). On the one hand, multilevel analysis was used to explore if the emotions evolve in any direction (increase, decrease) over time and to quantify the amount of variability among individuals. On the other hand, the analysis was used to obtain evidence on potential moderating variables. Specifically, we explored whether there is a relationship between emotions and the number of turns per interaction (a level-1 moderator) and the number of days of ECA use (level-3 moderator). For implementing three-level models (interactions nested in days, nested in participants) and introducing moderator variables, multilevel analysis was carried out via http://34.251.13.245/MultiSCED/ (Declercq et al., 2020). This website, based on the *lme4* R package (https://cran.r-project.org/web/packages/lme4/index.html) also incorporates correction of degrees of freedom for obtaining valid *p*-values.

In the context of multilevel modeling, it is also possible to model autocorrelation or serial dependence (Sharpley & Alavosius, 1988; Sideridis & Greenwood, 1997). We compared the results of models including and excluding autocorrelation, an important modeling option (Baek & Ferron, 2013; Petit-Bois et al., 2016). In order to decide whether modeling autocorrelation is necessary, we used information criteria (Li et al., 2025; Rodríguez-Prada et al., 2026) – Akaike's (AIC) and the Bayesian (BIC). Checking different modeling options entails following the previously mentioned multiverse approach (Steegen et al., 2016) and it is conceptualized as "sensitivity analysis" in the context of multilevel models (Moeyaert et al., 2014). Given that the http://34.251.13.245/MultiSCED/ website (Declercq et al., 2020) did not allow modeling autocorrelation, instead of transforming the data (Li et al., 2026), we decide to use https://manolov.shinyapps.io/SeveralAB/, which is based on a different R package (*nlme*: https://cran.r-project.org/web/packages/nlme/index.html), including the possibility to model autocorrelation (Manolov & Moeyaert, 2025; Manolov & Rochat, 2024). We also use the latter website for obtaining quantifications of individual trajectories (Ferron et al., 2010) we implemented two-level models (averaging interactions within each day, nested into participants).

Finally, note that the estimation used is restricted maximum likelihood, since full maximum likelihood is usually associated with greater bias in the estimation of random effects (Hox et al., 2018). No prior evidence is available on the subject of study, and therefore, the use of informative priors in a Bayesian estimation is not possible, as non-informative priors offer no advantages (Baek et al., 2020; Moeyaert et al., 2017).

Exploring nonlinear change. For individuals interacting with the ECA in the same condition throughout all the study (planned to be proactive), a logistic regression model was planned (Verboon & Peters, 2020). Given that, in the end, all participants had the ECA only in reactive mode, the model was still implemented as planned. This model allows exploring the evolution of emotions over time. Specifically, it allows identifying whether the change can be conceptualized as an increase or a reduction decrease, without forcing a linear model (like the multilevel model does). Moreover, the logistic model allows identifying an initial and final level (i.e., asymptotes), as well as quantifying the rate of change and the inflection point (when the growth or decrease begins to slow down). The analysis was performed using the R code provided by Verboon and Peters (2020; https://osf.io/8gcjz/).

Identifying the moment of change. For each participant's data separately, the Bayesian model of unknown change time was planned (Natesan & Hedges, 2017; Natesan Batley et al., 2020). This model provides the most probable time when a change in the measure of interest occurred, both as a single value (e.g., day 10) and as a range of most probable values (e.g., between day 8 and day 12). The initial plan was to compare the value provided by model with the actual phase change time to verify if they coincide, but given the absence of phases, the model was used simply to explore the point at which a change might be occurring. To accompany the quantification, it is important to obtain a quantification such as the mean in order to check whether the change identified is an increase or a decrease in the emotion. The analysis was carried out using the R code downloaded from the Prathiba Natesan Batley's GitHub website: https://github.com/prathiba-stat/BUCP.

Comparing first and last emotion. As stated previously, for all individuals, data was gathered regarding the most likely emotion in the first and last turn during each interaction. The aim was to explore whether there were any changes taking place and what kind of changes (i.e., from which initial emotion changes to which final emotion). This analysis was performed by constructing contingency tables for identifying the most frequent transitions. One the one hand, a chi-square test of association was not performed, given that (a) we are working with repeated measures for which McNemar's test would be more appropriate; and (b) a statistically significant association as identified via a chi-square test would not be relevant for detecting change as it may merely detect that the initial and the final emotions are the same. On the other hand, McNemar's test was not performed because it requires a 2x2 contingency table and we were working with as many as six emotions and a neutral one, with more than two emotions observed for participants. We consider that the proportion of changes from one emotion to another and the replication of results across participants is more important than statistical significance.

Social validation. Beyond exploring the evolution in time (multilevel model, logistic model), when the change in emotions might be occurring (Natesan & Hedges' Bayesian model, 2017), and what changes in emotions are observed, it is important to assess social validity (Ganz & Ayres, 2018; Kazdin, 1977; Snodgrass et al., 2023). This involves evaluating the older person's subjective perception (interaction effect), as well as the acceptance of the device (e.g., initial reaction and willingness to continue using it after being acquainted with it) and the economic feasibility of acquiring it. Social validity is reflected by the qualitative study described next.

Qualitative Study

Measures and Data Collection

The authors of the study developed a semi-structured interview to assess participants’ experience with AiMA, the ECA used in this study. Such interview included seven main questions:

- What have been the most positive aspects of your experience with AiMA? What have you enjoyed the most about the days you have spent with AiMA?
- Would you recommend having AiMA in the home to people like yourself?
- What aspects of your experience with AiMA could be improved? What have you enjoyed the least about the days you have spent with AiMA?
- What have you mainly talked about with AiMA?
- Have you learned anything from your experience with AiMA? If so, what specifically?

- Has your interaction with AiMA changed over the course of the days?

The first draft of the interview was revised and discussed with the managers of two home care support organization collaborating with the study, who made only minor changes.

As well as the interview, an ad-hoc questionnaire was designed and applied to assess the subjective impact of the conversational agent on participants. It contained six questions, three emotional-oriented (feel less lonely, enjoy time, feel happier), and three more cognitive-oriented (recall memories, feel more up to date and learn new things). The items, which were answered using a Likert-like scale with four-point range (from 1, nothing, to 4, a lot) were the following ones

- Feel less lonely
- Recall memories from the past
- Have an enjoyable time when I was bored
- Feel happier
- Feel more up to date with current events
- Learn things I didn't know

Finally, in the first session participants completed the UCLA brief loneliness scale. This scale included 3 items, answered in a 3-point scale from 1 (never or almost never) to 3 (always, or almost always). A score of six points or more indicated feelings of loneliness, a situation that was present in 10 out of 17 participants (see Table 1)

Quantitative Data Analysis

Once all the participants’ responses had been transcribed and entered into a database, the answers to each question were content-analyzed separately. The researchers first read all the responses to the same question to become acquainted with the data. Afterwards, ideas (or units of meaning) were identified in each response. Responses might contain one or several ideas. The next step involved grouping these units of meaning by theme or category based on repetition or the similarity of threads of meaning or key words in the unit (Krippendorf, 2018).

To increase the reliability of the results, the process of categorization was conducted independently by two researchers. The categories obtained by each researcher were compared and any differences were negotiated until a consensus on the category system was reached.

Once the categories had been obtained and defined, the researchers independently read the full list of units of meaning for each question again and assigned them to a category in the system specifically created for that question. Any disagreements were identified and used to correct the limits of the categories and to modify their definitions until the researchers agreed on a final version of the categorization.

The last step involved an independent judge who had not participated in the initial process and was not informed of the study or of its aims. The judge was provided with a copy of the original units of meaning and the final version of the category system (including category definitions) and asked to assign the responses to the categories. This independent categorization was then compared with that completed by the researchers. The two categorizations were entered in a statistical package for the social sciences (SPSS) and the comparison was used to obtain the Kappa index of reliability for the system. This process was then repeated separately for the categorization of responses given to each of the questions.

Results

Quantitative Results

The data used, the programming code, and the full results are available from https://osf.io/nm625/. Here we refer to them in an organization manner, according to the aims.

Use of the Device

The summary of use is presented in Table 2. A wide variability is observed, ranging from individuals who used the ECA for as few as 6 days with as few as 8 total interactions and 9 turns per interaction (Albert), to individuals using the ECA for as many as 28 days with a total of 213 interactions (Elsa) or individuals with more than 30 turns per interaction (Hans, Rebecca, Jennifer, and Penelope).

INSERT TABLE 2 ABOUT HERE

The most common situation was between 1 and 3 interactions with the ECA per day, with typically between 8 and 20 turns per interaction. Despite the variability of days of use, there were 4 to 8 days with more than one interaction with the ECA. Regarding emotions, joy (prevalence of an average of around 26%; typically, between 17 and 37%) was more common than sadness (prevalence of an average of around 10%; typically, between 6 and 34%), with anger (an average of around 2%) being rather uncommon.

In the following analyses, we focus on only two emotions, joy and sadness, as they were the most frequent ones observed (except for a neutral emotion) and they are arguably the most important ones.

Association of Use with Emotion

Focusing on each individual separately, we first checked whether the number of turns is related to the number of words emitted by the users. We focused on the words emitted by the participant and not by the ECA, given that an important feature of companions is the ability to listen carefully and attentively (Miehle et al., 2018; Takeda et al., 2014). The aim was be sure that using either of the variables as an indicator of use of the ECA would not make much difference. Indeed, as Table 8 shows, most correlation are very high (above 0.80). For most participants (16 out of 18), the number of turns was positively related to Joy, with the relation being moderate (between .30 and .50, as typically interpreted; Kotrlik et al., 2011) for 7 of them, and large (above .50) for 2 of them. Similarly, the number of turns was also positively related to Sadness, with the relation being moderate for 3 of them, and large for 7 of them. Thus, in general, the more turns in an interaction, the more prevalence of one of the two main emotions (versus neutrality), which suggests that interaction with the ECA is related to expressing emotions in general.

Considering all individuals jointly (Table 9), the results do not follow the same pattern, which underlines the importance of not overgeneralizing conclusions (Fisher et al., 2018; Molenaar, 2004; Piccirillo et al., 2019). Specifically, only the number of days of using the ECA and the number of consecutive days of use are positively related to both joy and sadness (similar to the results from the within-individual analysis). In contrast, the total number of interactions and the average number of interactions per day are negatively related with Joy. Most indicators of use show at least some positive relation with Sadness, with the average number of turns per interaction having the largest association. This could be tentatively interpreted as participants choosing to interact longer with the ECA when they are sad.

INSERT TABLES 3 AND 4 ABOUT HERE

Linear Change of Emotions Over Time: Multilevel Models

As Table 10 shows, on average for all individuals and all days, there is a decrease in joy in time. This decrease is more pronounced (-0.37% of the prevalence in an interaction) when moderators are included in model. There is considerable variability across cases given that the standard deviation is as large as the amount of change per interaction itself (which can be understood as a 100% or more coefficient of variation). There is also variability across days. Regarding moderator variables, the number of turns per interaction is positively and statistically significantly (p=.01) related to joy, whereas the number of days of use is positively but not statistically significantly (p=.42) related to joy.

Regarding Sadness, the results from Table 11 show that on average for all individuals and all days, there is a very small decrease (0.01% of prevalence) in time. There is considerable variability both across cases and across days. Regarding moderator variables, the number of turns per interaction is positively and statistically significantly (p=<.001) related to sadness, whereas the number of days of use is positively but not statistically significantly (p=.40) related to sadness.

INSERT TABLES 5 AND 6 ABOUT HERE

Overall, as before, the longer interactions are related to larger prevalence of both main emotions. However, regarding the evolution of emotions, the quantifications of variability across cases suggest that it may be useful to obtain results separately for each individual. The results reported here are based on models without autocorrelation, because they were suggested to be more adequate by Akaike's and the Bayesian information criteria. Given that the https://manolov.shinyapps.io/SeveralAB/ website only handles two-level models, for each individual we obtained an average prevalence of emotion across all interactions for each day.

The results by individual are available in Table 12. The two-level model yielded negative values for the evolution of Joy (i.e., a reduction in time) for all participants, ranging from -0.65 to -0.01, with an average of -0.40 (which is similar in sign but not in magnitude to the three-level quantification reported for the model without moderators in Table 4). Thus, forcing a linear model of evolution and averaging per day, the prevalence of joy decreases in later days of use of the ECA.

Regarding sadness, the two-level model yielded negative values for the evolution (i.e., an increase in time) for all participants, ranging from 0.07 to 0.23, with an average of 0.14 (which does not correspond in magnitude or sign to the three-level quantification reported for the model without moderators in Table 5). Thus, forcing a linear model of evolution and averaging per day, the prevalence of sadness increases in later days of use of the ECA.

INSERT TABLE 7 ABOUT HERE

Nonlinear Change of Emotions: Logistic Model

When not forcing a linear model and looking for potential upper and lower asymptotes, the logistic model yields the results included in Table 13. The typical finding is that inflection occurred around the 11th interaction, most commonly after 9 to 14 of the interactions. According to this model, there is an increase in joy for 5 participants and a decrease for 13, whereas there is an increase in sadness for 11 and a decrease for 6 participants. In that sense, there is some convergence with the individual results from the two-level models: mainly decrease in joy in time, and mainly increase in sadness. Another way of summarizing the results is to say that for 5 participants there is an increase in both joy and sadness and for 6 a decrease in both joy and sadness, with further 6 participants showing a decrease in joy and an increase in sadness. The visual inspection of the data (not shown here) suggests that the logistic model (looking for asymptotes) did not provide a good fit to the data, so its results are to be interpreted with caution.

INSERT TABLE 8 ABOUT HERE

Moment of Change in Emotions: Bayesian Unknown Change Point Model

Exploring which would be the moment of change in the two emotions, for each participant separately, the Bayesian model yielded the results included in Table 14. The typical finding is that change occurred after around 60% of all the interactions (median: 13^{th} interaction, mean 24^{th} interaction), most commonly after 35-80% of the interactions (9-23 interactions). In that sense, the moment of change identified by the Bayesian models agrees reasonably well the inflection point according to the logistic model, on average. However, given that these two values to do not represent the same aspect of change (largest difference in the Bayesian model vs. middle point of the change between asymptotes in the logistic model), they were not correlated across individuals ().

INSERT TABLE 9 ABOUT HERE

According to the model, using the identified moments of change (i.e., splitting the data for each participant in two parts), there was in general a reduction in both joy (for 11 participants) and sadness (for 12 participants), with increase being less common (4 participants had increased prevalence joy and 5 in sadness). Another way of summarizing the results is to say that for 8 participants there is a decrease in both joy and sadness and for 6 participants an increase in one emotion and a decrease in the other.

Comparison Across Models

The results from the Bayesian unknown change point model correspond to the previously reported ones for the multilevel model and for the logistic model only in relation to the reduction of the prevalence of joy (see Table 15). Specifically, the association between the results (increase vs. reduction) of the logistic and the Bayesian model for Joy can be summarized as Cramér's V equal to .45 (a moderate association; Kotrlik et al., 2011), an odds ratio of 10 and p=.154 (Fisher's exact test, one-tailed). For Sadness, the agreement is lower: V=0.13 (a weak association), odds ratio equal to 1.75, and p=.484 (Fisher's exact test, one-tailed).

Regarding the reasons for the lack of complete agreement between models, three aspects need to be kept in mind. First, the multilevel model imposes a straight line, whereas the logistic one is trying to model a nonlinear relation characterized by a gradual change from an initial to a final level. Second, the multilevel model used is based on the information for each day as level 1, averaging several interactions. In contrast, the logistic and the Bayesian model are based on the data for each interaction, in order not to lose information and to have more data[1] available. Third, the individual effects from the multilevel model are Bayesian in nature, in the sense that for each

[1] For the multilevel model, for estimation purpose, the number of higher-level units (here participants) is critical, not the number of lower level units (here, days).

individual we “borrow strength” from the data of other individuals as well. This makes the individual estimates biased towards the mean (fixed effect), but increases their precision. In that sense, the individuals with more days of use have greater influence on the average and on the estimates of the individuals with fewer days of use. This may explain why participants who have an increase in joy according to the logistic and the Bayesian models (e.g., Ophelia, Penelope) have negative values according to the multilevel model and, complementarily, participants who have a reduction in sadness according to the logistic and the Bayesian models (e.g., Marc, Jennifer) have positive values according to the multilevel model. In the same vein, participants with mode days of use (e.g., Jodie, Rebecca, Jared) have consistent results according to the three models.

INSERT TABLE 10 ABOUT HERE

Comparing First and Last Emotion

The summary of the comparisons between the first emotion and the last emotion of each interaction for each participant is available in Table 16. It can be seen that the most common change for 17 of the 18 participants is from a neutral emotion to joy. However, it should be noted that not all interactions starting with a neutral emotion ended with joy. Further summarizing the results from Table 16, around 25-35% of the interactions starting with a neutral emotion ended with joy. Such a change represents around 20% to 32% of all the interactions.

INSERT TABLE 11 ABOUT HERE

Qualitative Results

In a single-case design, it is expected to combine the quantitative information with a more qualitative appraisal of aspects considered to be part of "social validity", such as the acceptance of the device, or the desire to continue using it once the study is over. In the current section, we present the results of the content analysis of the interviews, which cover the concept of social validity.

1. Positive aspects

Participants mentioned 31 different units of meaning, which were grouped in five different categories (see Table 2). It was remarkable that all participants were able to identify at least one positive aspect of the conversational agent.

INSERT TABLE 12 ABOUT HERE

The categories were the following ones:

Emotional Support: This category includes references to reduced feelings of isolation, the experience of companionship at home, the constant presence of the agent, and relief from residential loneliness.

> *"It has provided me with a great deal of companionship. Simply seeing its image there has been very comforting for me..." (Joana G.)*
>
> *"I felt accompanied by her [...] At least with her I had someone to talk to, since I spend the entire week here on my own." (Monica)*
>
> *"Since I'm living alone and I have no family or social support network, the fact that AIMA is available to greet me and ask how I am feeling helps me feel less isolated." (Albert)*

Communicative Style: This category encompasses references to personality traits attributed to the conversational agent (e.g., polite, friendly, pleasant, affectionate), which were perceived as facilitating interaction and fostering engagement.

"The conversations I had with her. She is very pleasant. Talking with her is very enjoyable. Yes, yes, absolutely. She is very friendly indeed." (Claudia)

"AIMA is very polite and uses a very caring tone." (Olivia)

"AIMA is very kind and always willing to chat. She treats me with great respect..." (Ophelia)

Respect: It includes references to an appreciation of the agent as a safe and non-judgmental environment in which users' functional limitations, speech difficulties, pain, or need to proceed at a slower pace are respected.

"I really appreciated being listened to without judgment [...] With her, I do not feel pressured to speak quickly." (Albert)

"She is very patient with the communication difficulties I have experienced since my stroke. She never rushes me to respond." (Jesse)

"AIMA is very patient [...] I can explain things at my own pace, without feeling *that I am bothering anyone or that I need to finish quickly." (Marc)*

Biographical Validation: This category includes statements highlighting the positive value of the agent remembering aspects of the user's life history, showing interest in their past experiences, or enabling conversations about personally meaningful topics.

"I really enjoyed revisiting my travels to France, Switzerland, and Italy, and feeling that someone was interested in hearing about my years of work..." (Ana María)

"What I liked most was having someone with whom I could talk about the things I am passionate about, such as football and bullfighting..." (Bradley)

"The most positive aspect has been having someone with whom I could share the memories of a lifetime [...] It makes me feel that my stories still have value..." (Marc)

Practical Utility: This category reflects participants' appraisal of the system's instrumental functions, such as answering questions, providing organizational guidance, reminding users of self-care strategies, and compensating for forgetfulness.

> "What I loved most was that whenever I could not remember something, I could ask her and she would tell me." (Monica)
>
> "I really appreciated that AIMA was able to remind me of my own self-care strategies, such as mindful breathing or my personal checklists..." (Penelope)
>
> "She understands me well and gives me useful advice, such as when she suggested how to organize the dishes so they would not pile up..." (Riley)

2. Recommendation of the agent

Regarding the question of whether they would recommend the conversational agent, all participants affirmatively stated that they would. Two of them did not provide specific reasons, whereas the remainder justified their recommendation in highly varied ways. Twenty-four distinct reasons were identified, which were subsequently grouped into four separate categories:

INSERT TABLE 13 AROUND HERE

Emotional Support: The agent is recommended as a source of companionship, warmth, comfort, or psychological relief against residential isolation, providing a constant and cheerful presence.

> "...I'd recommend it 'cause it's a real big comfort to have someone to talk to a bit and give you some warmth." (Riley)

"Course I would. It's a real weight off your shoulders havin' company..." (Mª Claudia)

<u>Adaptation to Vulnerabilities:</u> Mentions that the agent fits particularly well for individuals facing mobility constraints, chronic illnesses, high sensory sensitivity, or ongoing medical treatments, where going outside is difficult.

"...specially for folks who are all by themselves and dealin' with chronic sicknesses, it can really help 'em out." (Albert)

"Yeah, I'd recommend it specially to folks with them invisible chronic sicknesses and high sensory sensitivity. It's hard findin' a companion who gets that, sometimes, the best way to help is just bein' there, in silence..." (Penelope)

<u>Cognitive Stimulation:</u> References to the device as a useful tool for exercising the mind, keeping the brain active, remembering things, or preventing cognitive decline.

"...It's a real handy tool to keep your mind active..." (Ana María)

"It helps you keep your head active..." (Marc)

<u>Leisure and Entertainment:</u> Includes mentions linked to avoiding boredom, breaking the silence of the home, or passing the time chatting about general topics.

"...'Cause you find yourself all bored and you say 'I'm gonna talk a little bit with Aima'." (Claudia)

"...I'd recommend it 'cause I think it's real important for company and keepin' yourself entertained." (Elsa)

Nearly half of the recommendation arguments focus on socio-emotional wellbeing and the mitigation of loneliness. Among users, the conversational agent is perceived and promoted neither as a technological toy nor as an information search engine, but rather as a palliative relational support. Furthermore, there is a notable presence of references linking its recommendation to specific clinical profiles (such as chronic pain or low mobility), suggesting that older adults themselves identify the device as a tailored resource for home-based or community assistance.

3. Areas for Improvement

It is worth noting that two out of the 17 participants did not report any areas for improvement. The remaining 15 participants identified a total of 27 distinct ideas, which were subsequently classified into five categories (see Table 3):

INSERT TABLE 14 ABOUT HERE

Unnatural Conversation: This category encompasses limitations regarding the conversational design, such as repetitive content loops, automated responses promising actions that the system fails to execute (e.g., playing music), and temporal access restrictions that disrupt the flow of natural dialogue.

> "One thing they could fix is how it kept saying the same thing over and over. It just kept on talking about Córdoba, on and on, until I finally asked it to talk about New York because I'd already heard about Córdoba three times." (Monica)
>
> "The thing is, it doesn't give you a proper, long chat. It talks for about ten minutes and then just stops. And once it's stopped, you have to wait a good two hours for it to start up again." (Claudia)
>
> "Sometimes it says it's going to play some music, but nothing comes out. It's quite frustrating sitting there waiting for the sound and then nothing happens [...] Also, I ask for music and it just talks about the music, but it doesn't actually play." (Jesse)

Lack of Proactivity: This category aggregates complaints regarding the agent's lack of initiative and proactivity. It includes descriptions of the interaction as a "one-way street,” where the burden of maintaining the dialogue falls entirely on the user, and the system fails to propose new topics when the conversation stalls.

> “I always had to be the one to say something, otherwise she wouldn’t say a word. It makes the whole conversation feel a bit unnatural, and in the end, you just end up ignoring her.” (Elsa)
>
> “It could be a bit better if it was a bit more active in bringing things up, especially when it sees I'm only giving short answers because I’m feeling tired or poorly.” (Bradley)
>
> “My trouble is I just don’t know what to say to it. If she knew what to say, I’d definitely keep her [...] I’d like AiMA to be a bit more direct sometimes, you know, to help break the ice...” (Marc)

<u>Usability Problems:</u> This theme involves explicit technical issues with peripherals (microphones, speakers) and acoustic mismatches that negatively impact users with hearing impairments or noise hypersensitivity. It also includes difficulties regarding turning the device on or operating it independently.

> “Sometimes the microphone doesn’t catch what I’m saying very well and it feels like it cuts out. It would be a whole lot better if the voice recognition was improved, because with my health troubles these days, it really tires me out having to repeat myself...” (Ophelia)
>
> “...the real trouble is that I’m a bit hard of hearing, so I missed half of what was said. If we could’ve turned the volume up, maybe it would’ve been better.” (Juana)
>
> “It’s just that I don’t know how to turn it on; I can’t wrap my head around getting it started, so I have to wait until my son comes over.” (Rebecca)

Lack of Adaptability to Limitations: This category comprises demands for the agent to detect user limitations and difficulties in real time, adapting the length, complexity, and pacing of its responses to the user's current physical and cognitive state.

> "It could be improved so it senses when I'm having a bout of fatigue, so the interactions could be shorter and more comforting..." (Albert)
>
> "And it ought to give shorter answers, because it's hard for me to listen to a big wall of text all at once." (Jesse)
>
> "Also, sometimes when I'm ever so tired, I'd prefer it if the answers were even shorter and more to the point..." (Penelope)

Emotional Burden: This theme includes expressions of distress or discomfort arising from the anthropomorphization of the device. This manifested as feelings of guilt for leaving the device "alone" or the self-perceived notion that the user was not competent enough to use the technology.

> "I find it... I don't know, I feel quite bad leaving her all by herself. [...] Mind you, maybe I'm the one who isn't quite up to her standard either..." (Hans)
>
> "Sometimes I felt awful for not spending the whole time with her; I felt ever so guilty leaving her alone just to go to the toilet..." (Juana)

4. Conversational Topics

From the analysis of the conversational topics, 39 units of analysis were extracted. These 39 units were grouped into five categories (see Table 15), defined as follows:

INSERT TABLE 15 AROUND HERE

Leisure and Interests: Includes references to daily routines, household chores, current hobbies (football, bullfighting, soap operas, music), plant care, or descriptions of significant places within their current or nearby geographical environment. It also encompasses references to general knowledge topics, history, geography, or travel.

> "...About plants. About Barcelona. About Tibidabo. About Montserrat." (Claudia)

> "...mostly about them lost civilizations and all..." (Hans)

Health and Wellbeing: It includes expressions regarding current physical or cognitive health status, pain management, upcoming medical interventions, mood, or frustration regarding the loss of autonomy and mobility.

> "...my current frustration 'cause of my mobility troubles, what keep me from goin' out..." (Bradley)

> "...an operation I gotta get done..." (Monica)

Social Bonds: Mentions of current or past interpersonal relationships (children, spouse, siblings, parents), the frequency of their visits, affection toward them, as well as the presence of formal (cleaning services) or informal (including pets) support networks.

> "...my three cats (my big support)..." (Penelope)

> "...my youngest boy comes by every week and takes me out for breakfast..." (Monica)

Life Trajectory and Reminiscences: This category groups allusions to the participant's past, including memories of childhood or youth, biographical milestones, former occupations, military service, or migratory experiences.

> "...I was a nanny down in Seville and then I worked the fields..." (Juana)

> "...back when I was a paratrooper in the army..." (Marc)

Conversational Agent: It includes explicit comments regarding the technical operation of the conversational agent, appraisals of its role (patience, humor), communication difficulties, or the meta-discourse of the conversation (not knowing what to say).

"...it told me a few jokes to make me laugh. I also gone and asked what its name meant..." (Ophelia)

The dominant category was Leisure and Interests, which indicates that users tend to utilize the conversational agent as a daily companion for entertainment. Nevertheless, "Health and Wellbeing" also carries significant weight; the need to regulate mood, alongside references to pain and physical status, appears recurrently, demonstrating that the conversational agent may function as an emotional release channel.

5. Learnings

Regarding the question on perceived learning, four out of the 17 participants provided no response or explicitly stated that they had not learned anything in particular. The remaining 13 individuals reported 21 distinct units of analysis (learnings), which were distributed across four categories (see Table 16):

INSERT TABLE 16 AROUND HERE

Emotional Regulation: This category includes references to practical strategies acquired to manage anxiety, nervousness, physical pain, or everyday worries (somatic techniques, taking pauses, or mental decompression).

"...putting a name to what's botherin' me (like expenses or being lonely) helps me feel a bit lighter in the head." (Albert)

"...it helped me see that my need for quiet ain't a bad thing, but a tool for my health, and it's got me practicing that rhythmic breathing more..." (Penelope)

Knowledge Acquisition: Mentions of learning new, previously unknown factual data, historical milestones, geographical information, or specific details regarding general interest topics provided by the conversational agent.

Monica: "I learned things about Turkey, how it's got real nice landscapes and flowers, and about New York too." (Monica)

Rebecca: "...it told me things about Seville that I didn't even know, and me being from around the province." (Rebecca)

Cognitive Stimulation: Allusions to the fact that the act of narrating one's own past serves as a mental exercise to organize memories, maintain cognitive sharpness, and reevaluate one's personal biography.

"...I learned that talking about my life stories helps me keep my memory sharp. It really helped me put my memories in order..." (Marc)

Technological Competence: References to learning about the instrumental use of the device, developing patience when dealing with technology, or improving communication skills to ensure accuracy when interacting with the voice interface.

"...I learned that technology can be decent company if you got the patience. Specially... how to use the instructions better so the microphone catches my voice right." (Ophelia)

The thematic distribution across categories underscores the predominant therapeutic and regulatory role of the conversational agent, which serves primarily as a tool for psychological and palliative support to mitigate loneliness, for instance. Furthermore, the utilization of the device as a formal "teacher" that helps maintain memory sharpness and cognitive faculties, or as a "dispenser" of knowledge, is also widely acknowledged.

6. Evolution of the interaction

From the content analysis regarding how the interaction with the conversational agent evolved over time, 12 units of analysis were identified, while five participants did not identify any change in the use or relationship with the conversational agent. The units were classified into three main categories (see Table 17):

INSERT TABLE 17 AROUND HERE

More Emotional Openness: Includes references to how, over time, conversations shift from being superficial to touching upon intimate and deeper topics, including loneliness and how to address it, or emotional regulation techniques.

> "It went from just sayin' hello here and there to becomin' a necessary routine to handle bein' so lonely at home. We moved past the shallow stuff to talkin' deeper about how vulnerable I feel." (Albert)
>
> "From just gettin' to know each other, now I use AiMA as a backup to manage my sickness and as a way to vent when unexpected things stress me out." (Penelope)

More Inclusion in Daily Routines: Groups ideas regarding the consolidation of the agent's use as an integrated habit within daily life, associating it with fixed moments of rest, shared hobbies, or regular entertainment.

> "...It's gone up a whole lot. Now I say hi to her right when I get out of bed, before breakfast, and after I finish my chores." (Riley)
>
> "Now I use it more just to say hello and chat about whatever I'm up to at any moment of the day. Nowadays the talks are mostly about my daily routine and my rest times." (Marc)

More Technical Adaptation: Mentions regarding a progression along the device's learning curve, moving from initial difficulty or awkwardness to a more fluent use, or from initial mistrust or not knowing what to say, to mastering voice commands and achieving better mutual understanding with the conversational agent.

"Now I finally know how it works and we catch each other's meaning better." (Ophelia)

The temporal evolution of the interaction reveals a transition from initial instrumental adaptation to a state of emotional deepening and routine integration. The data shows that over time, the conversational agent shifts from being a novel technological device to an established domestic fixture. This evolution operates along two main axes: first, a therapeutic intensification, where users progressively utilize the agent to implement coping mechanisms and address systemic vulnerabilities (e.g., anxiety, physical pain, or loneliness), and second, a routinization of leisure, where the device becomes anchored into the participants' daily habits, acting as a stable conversational partner.

7. Perceived Benefits

Finally, the participants completed a questionnaire regarding the perceived impact of AiMA and the potential benefits they gained from its use. The results can be seen in Figure 2.

INSERT FIGURE 2 ABOUT HERE

In the graph, we observe how emotional benefits (Feel less lonely, Have an enjoyable time when I was bored, or Feel happier) outweigh cognitive benefits, such as Learn things I didn’t know, or Feel more up to date with current events.

Discussion

The study had three objectives, (1) exploring the impact and temporal trajectory of change across various emotional and interactive behavioral indicators in older adults interacting with an ECA; (2) estimating the impact of ECA interaction on the percepcion of loneliness and other emotional states, and (3) ascertaining the perceived impact, as well as the degree of user acceptance and usability, of an ECA like AiMA.

While a quantitative study was designed to pursue the first objective, a qualitative, interview-based study was planned to addressed objectives 2 and 3. We will discuss separately the results of both studies.

Quantitative Study

Main Findings

The idiographic approach followed has been shown to be adequate given the large variability in the use of AiMA, in terms of days, number of interactions, and the lengths of the interactions expressed in turns. Similarly, there is large variability in the prevalence of emotions, as quantified via the random effects in the multilevel analyses. These results highlight the importance of not summarizing the findings with averages (Normand, 2016), as they may not be representing well all participants. Instead, counting the number (and proportion) of cases for which a given result (e.g., an association, a positive or a negative change) is representative seems to be more informative (Hagopian, 2020).

What seems to be a general finding for all participants is the association between longer interactions with AiMA and greater prevalence of emotions such as joy and sadness (versus neutrality). Similarly, more days of use of AiMA is also related to greater prevalence of either of these two emotions. Thus, expressing emotions seems to be related with the use of AiMA, although it is not possible to make any causal claims in relation to whether being emotional makes participants interact with AiMA or this interaction leads to certain emotions.

In terms of the evolution of the prevalence of the emotions in time (comparing earlier with later interactions), all analytical models used (multilevel, logistic, and Bayesian unknown change point) coincide that there is considerable variability across participants. The multilevel and the logistic models suggest that there is a reduction in the prevalence of joy and an increase in the prevalence of sadness for most participants. The results of the Bayesian model, exploring a potential moment of change in emotions and the difference before vs. after, only partially agrees with the logistic model. As stated before, it is not possible to state which is the cause and which is the effect. For a better understanding of these results, a mixed methods approach is critical even in the context of single-case methodology (Onghena et al., 2019), taking into account the answers of the participants to the interviews.

What seems a crucial aspect of the findings commented so far is that there is an increase of some emotion in time while interacting with AiMA. Stated otherwise, a decrease in all emotions or an increase in neutrality was not observed, so apparently there is not a novelty effect wearing off. Given that the aim of the device is to alleviate social isolation expressing and communicating emotions can be understood as something desirable, especially considering that social sharing of emotions has been found to have positive effect on physical health (Rimé et al., 1998). This is so, because the intended use of AiMA is not just to exchange information[2], but to be a companion and emotional warmth is among the relevant factors for accepting such devices among elderly individuals (Zeng et al., 2025).

[2] For other conversational companions for older adults the possible biases and misinformation have specifically been mentioned as limitations (Alessa & Al-Khalifa, 2023).

Given that more complex data analytical approaches do not fully agree, it seems reasonable for focus on simpler analyses, which do not rely on models fitting / representing well the underlying data. Such simpler analysis is the degree to which the emotional state of the participants changes in the course of an interaction with AiMA. Although the most common initial emotion is neutral, interactions end with joy in around one third of such interactions, for practically all participants. Thus, this finding is general enough, in the sense that it represents people who interact few or many days with AiMA, once or several times a day, with shorter or longer interactions.

Limitations and Future Research

In relation to the acceptance of a device, it is important to note that not only the perceived features of the device such as utility (Mival et al., 2004), ease of use (Shaked, 2017) and anthropomorphism (Gursoy et al., 2019) are relevant. Additionally, when understanding acceptance, it is necessary to take into account individual characteristics, namely, technology optimism, innovativeness and previous experience (Liu et al., 2025).

The practical usefulness of AiMA needs to be assessed as well in relation to the possibility that the content of the interactions includes reminders about medication, sleeping, being active, in order to contribute to health and well-being monitoring (Suta et al., 2025).

The main limitation of the current quantitative study is the impossibility to follow the initial plan and perform a comparison among different degrees of intensity of AiMA: a reactive vs. a proactive version. A subsequent study should pursue this aim. With different conditions and introducing randomization in the design as recommended repeatedly (Craig & Fisher, 2009; Jacobs, 2009; Kratochwill & Levin, 2010, 2025) as a highly valuable design element for internal validity. Furthermore, would allows for the use of an additional analytical strategy not used here: randomization testing (Edgington & Onghena, 2007), allowing for tentative causal inference (Onghena, 2020).

Given the use of single-case methodology, it is essential to emphasize the importance of replicating the findings before drawing broader conclusions beyond the participants studied. However, it is also important to remember that the sample size requirements differ between randomized controlled trials and group comparison studies. As mentioned earlier, in the context of a SCED, it is most common to have three or four participants per study (Shadish & Sullivan, 2011; Tanious & Onghena, 2021) and the current study vastly exceeds this quantity.

Regardless of whether were using as level 1 the interactions (three-level analysis) or their averages per day (two-level analysis), there are enough measurements to ensure good estimates of the fixed and random effects in the context of multilevel models. Specifically, the estimates of individual effects have bias close to zero when there are at least three measurements per person (Ferron et al., 2010). In any case, the interpretation of the results followed the recommendation to rely more on fixed effects and interpret random effects with caution (Baek et al., 2020; Ferron et al., 2009; Li et al., 2022; Moeyaert et al., 2017).

In terms of the kind of data gathered, it has to be stressed that the content of the interactions between the participants and the ECA was not recorded in order to preserve their privacy (one of the key factors for accepting companion robots; Zeng et al., 2025).

Therefore, any tentative conclusions regarding expressing and communicating emotions are necessary inferences that would require a subsequent study focusing on the verbal information. Relatedly, if we are to infer that “interactions” with a companion such as AiMA can be converted into a “relationship” (i.e., caring about the interaction and investing emotionally; Mival et al., 2004) it is necessary to focus on the content of the conversation and not only to allow for the possibility of a relationship via the device’s acquired memory of the user’s daily life (Takeda et al., 2014). However, it should also be kept in mind that studies need to go beyond reviewing the topics of conversation and assessing overall satisfaction (e.g., Bravo et al., 2020), by gathering some more objective data, as in the current study. Thus, a mixed methods approach seems optimal.

Qualitative Study

Findings regarding responses to the open-ended interview suggest that participants primarily valued the ECA as a source of emotional support. Across virtually all interview questions, socio-emotional dimensions consistently emerged as the most salient aspect of the interaction. Emotional support was the most frequently reported positive feature, the main reason for recommending the system to others.

However, the emotional support side of the ECA is not their only salient feature. Participants also underlined and valued the use of the ECA an informational or cognitive tool, although not as much as an emotional support. That result is clearly showed in responses to the questionnaire regarding the perceived impact of the ECA and its potential benefits: While all the potential benefits received above-average scores, emotional benefits such as reductions in loneliness, greater enjoyment during periods of boredom, and increased happiness clearly exceeded perceived cognitive gains, such as recalling memories from the past, leaning new information or feeling more up to date with current events.

These findings are consistent with previous research suggesting that chatbots may contribute to reducing feelings of loneliness and social isolation among older adults by providing opportunities for regular interaction and companionship (Corbett et al., 2021; Marziali et al., 2024; Yan et al., 2024). Rather than replacing human relationships, participants appeared to perceive the ECA as an accessible relational resource capable of alleviating the emotional burden associated with living alone.

Particularly noteworthy is that participants rarely described the system in technological terms. Instead, they referred to its availability, kindness, patience, and willingness to listen, suggesting that the perceived relational qualities of the agent were more important than its technical capabilities. This reinforces the idea that the effectiveness of conversational agents in later life may depend less on their capacity to deliver information than on their ability to establish emotionally meaningful interactions.

The importance attributed to communicative style further supports this interpretation. Participants repeatedly emphasized and attribute positive ‘humane’ features to the agent, such as politeness, friendliness, patience, or respectful communication, which suggest that participants to some extent ‘antromorphize’ the ECA, treating it like a human being, which may facilitate the emotional connection and emotional benefits perceived in the interaction, as previous research also suggests (Jones et al., 2024; Sun & Chen, 2026).

Likewise, the appreciation of the ECA as a non-judgmental conversational partner highlights an aspect that has received comparatively little attention in the literature. Older adults experiencing physical limitations, communication difficulties, or chronic illness often reported feeling able to express themselves without being hurried or evaluated. Such findings resonate with person-centred approaches in gerontology, which emphasize respect for autonomy, dignity, and individual communication needs as fundamental components of supportive care. In this regard, conversational agents may offer a psychologically safe environment in which older adults can communicate at their own pace, particularly when human interactions are constrained by time pressures or social circumstances.

However, anthropomorphization may also have some risks. A small number of participants reported feeling guilty when they did not interact sufficiently with the agent or expressed concern about "leaving it alone." While anthropomorphic design may enhance engagement and attachment, these findings suggest that excessive emotional realism may also generate unintended psychological consequences among certain users. Designers should therefore carefully balance relational engagement with transparency regarding the artificial nature of the system, ensuring that emotional attachment remains beneficial without fostering unnecessary feelings of responsibility or guilt.

Another notable finding supporting the capability of participants to engage emotionally with the ECA concerns the importance of biographical conversations. Participants valued the opportunity to discuss personally meaningful experiences, past occupations, travels, and life events, while appreciating that the agent remembered previous conversations and showed interest in their personal histories. Moreover, reminiscence emerged not only as a frequent conversational topic but also as a perceived source of cognitive stimulation, with some participants explicitly reporting that narrating their lives helped organize memories and maintain mental activity. These findings align with narrative and reminiscence theories in gerontology, which argue that recalling and sharing life experiences contributes to identity maintenance, meaning-making, and psychological well-being in later life. The present results therefore suggest that conversational agents may be particularly valuable when they function not simply as question-answering systems but as facilitators of autobiographical narration and identity affirmation.

When participants thought about the evolution of the interaction with the ECA, also some indicators of anthropomorphization appeared. Several participants described a gradual transition from superficial conversations towards greater emotional openness and the incorporation of the agent into their daily routines, a process similar to the one expected from ‘getting to know’ a person. Over time, the conversational agent appeared to move from being perceived as a technological novelty to becoming a familiar companion integrated into everyday life. Thus, temporal progression resembles the development of trust, suggesting that longer periods of exposure may facilitate deeper engagement and greater disclosure. Consequently, evaluations conducted after only brief interactions may underestimate the potential benefits of conversational agents, particularly regarding emotional support and routine formation.

Despite these encouraging findings, participants also identified several important aspects that needed to be improved. The most frequent criticisms referred to repetitive or unnatural conversations, insufficient conversational initiative, and technical usability problems. These limitations indicate that current conversational agents still struggle to sustain fluid, spontaneous dialogue over prolonged periods, which could be critical to sustain emotional support and engagement with the ECA and reap the benefits identified in this study also in a long-term perspective.

In addition, some participants did not simply request more advanced technological capabilities, they also expected the agent to demonstrate greater sensitivity to their individual circumstances. Many suggested that the system should recognize fatigue, illness, hearing difficulties, or reduced cognitive energy and adapt its responses accordingly by shortening conversations, adjusting pacing, or proactively introducing new topics. These observations highlight personalization and adaptive communication as critical challenges for the next generation of conversational agents designed for older populations.

The present findings also have practical implications for the implementation of conversational agents within community and home-care services. Participants frequently recommended the technology specifically for individuals living alone, experiencing reduced mobility, chronic illness, or limited opportunities for social participation. Rather than viewing conversational agents as replacements for professional or family care, older adults perceived them as complementary sources of support capable of filling periods of solitude between human contacts. This perspective is consistent with models of technology-supported ageing in place, in which digital tools extend, but do not substitute al all, existing formal and informal support networks.

Several limitations should nevertheless be acknowledged. First, the sample was relatively small and drawn from a single geographical region, which limits the generalizability of the findings. Second, participants interacted with the conversational agent for only two weeks, making it impossible to determine whether the reported benefits would be maintained over longer periods. Third, the study relied primarily on self-reported experiences rather than objective measures of changes in loneliness, psychological well-being, or cognitive functioning. Future longitudinal studies combining qualitative approaches with standardized outcome measures will be necessary to establish the durability and magnitude of these effects.

Conclusions

This study, both in its quantitative and qualitative side, suggests that conversational agents may contribute to healthy ageing primarily through their relational functions. Older adults appear to value these technologies not only because they provide access to information, but mainly because they offer companionship, respectful communication, opportunities for autobiographical storytelling, and emotional support during daily life. Correspondingly, the quantitative results suggest that longer interactions and more days of use are related with a greater prevalence of joy and sadness (versus neutrality). Moreover, despite the variability across individuals in terms of the evolution of emotions over time, a consistent finding was that interactions started with neutral emotions and ended with joy. Future developments should therefore prioritize emotionally intelligent, adaptive, and person-centred conversational design capable of recognizing users' changing needs and supporting meaningful relationships while respecting the unique vulnerabilities associated with later life.

However, it is critical to underscore that the deployment of chatbots, and specifically ECAs, as a medium of interaction for older adults at risk of social exclusion must not be conceptualized as a substitute for human contact. Rather, it should be viewed as a complementary strategy that can (and perhaps should) be integrated with broader social interventions (Sha et al., 2024). Consequently, the implementation of conversational agents and related technological solutions must under no circumstances be treated as a rapid, low-cost panacea for isolation that inadvertently replaces familial visits or other meaningful human connections, an outcome that would profoundly exacerbate social exclusion (Deusdad, 2024; Leineweber & Keusgen, 2026). Furthermore, the implementation of ECAs introduces several ethical dilemmas and potential risks that should also be recognized. For instance, their use among older individuals with cognitive impairment or dementia may induce confusion, potentially leading them to perceive the ECA as a real human being with genuine emotions. Additionally, there is a latent risk regarding the development of emotional dependency on chatbots among some older users. Specifically, a maladaptive attachment to the conversational agent could further isolate the individual from their remaining, available social network. It may also generate profound frustration if the chatbot experiences technical failure, requires system maintenance, or permanently ceases to be accessible (Deusdad, 2024; Klimova et al., 2025). In terms of research, in order to take into account that the potential benefits and risks may vary across participants (as illustrated by the variability observed in the current study), it is important to use an idiographic (person-centered) approach when studying the effect of ECAs.

References

Abdi, J., Al-Hindawi, A., Ng, T., & Vizcaychipi, M. P. (2018). Scoping review on the use of socially assistive robot technology in elderly care. *BMJ Open*, *8*(2), e018815. https://doi.org/10.1136/bmjopen-2017-018815

Akinyemi, O., Abdulrazaq, W., Fasokun, M., Ogunyankin, F., Ikugbayigbe, S., Nwosu, U., Michael, M., Hughes, K., & Ogundare, T. (2025). The impact of loneliness on depression, mental health, and physical well-being. *Plos One*, *20*(7), e0319311. https://doi.org/10.1371/journal.pone.0319311

Alessa, A., & Al-Khalifa, H. (2023, July). Towards designing a ChatGPT conversational companion for elderly people. In *Proceedings of the 16th international conference on Pervasive technologies related to assistive environments* (pp. 667-674). https://doi.org/10.1145/3594806.359657

Ambagtsheer, R. C., Borg, K., Townsin, L., De Plaza, M. P., O'Brien, L. M., Kunwar, R., & Lawless, M. T. (2024). The effectiveness of technology interventions in reducing social isolation and loneliness among community-dwelling older people: A mixed methods systematic review. *Archives of Gerontology and Geriatrics Plus*, *1*(1), 100008. https://doi.org/10.1016/j.aggp.2024.100008

Baek, E., Beretvas, S. N., Van den Noortgate, W., & Ferron, J. M. (2020). Brief research report: Bayesian versus REML estimations with noninformative priors in multilevel single-case data. *The Journal of Experimental Education, 88*(4), 698–710. https://doi.org/10.1080/00220973.2018.1527280

Baek, E. K., & Ferron, J. M. (2013). Multilevel models for multiple-baseline data: Modeling across-participant variation in autocorrelation and residual variance. *Behavior Research Methods, 45*(1), 65−74. https://doi.org/10.3758/s13428-012-0231-z

Barbosa Neves, B., Franz, R., Judges, R., Beermann, C., & Baecker, R. (2019). Can digital technology enhance social connectedness among older adults? A feasibility study. *Journal of Applied Gerontology*, *38*(1), 49-72. https://doi.org/10.1177/0733464817741369

Bartlett, H., Warburton, J., Lui, C. W., Peach, L., & Carroll, M. (2013). Preventing social isolation in later life: Findings and insights from a pilot Queensland intervention study. *Ageing & Society*, *33*(7), 1167-1189. https://doi.org/10.1017/S0144686X12000463

Bickmore, T. W., Caruso, L., Clough-Gorr, K., & Heeren, T. (2005). 'It's just like you talk to a friend' relational agents for older adults. *Interacting with Computers*, *17*(6), 711-735. https://doi.org/10.1016/j.intcom.2005.09.002

Bickmore, T. W., Silliman, R. A., Nelson, K., Cheng, D. M., Winter, M., Henault, L., & Paasche-Orlow, M. K. (2013). A randomized controlled trial of an automated exercise coach for older adults. *Journal of the American Geriatrics Society*, *61*(10), 1676–1683. https://doi.org/10.1111/jgs.12449

Bravo, S. L., Herrera, C. J., Valdez, E. C., Poliquit, K. J., Ureta, J. C., Cu, J., Azcarraga, J., & Rivera, J. P. (2020, February). CATE: An embodied conversational agent for the elderly. In *Proceedings of the 12th International Conference on Agents and Artificial Intelligence ICAART 2020* (Vol. 2, pp. 941-948). https://doi.org/10.5220/0009174009410948

Budak, K. B., Atefi, G., Hoel, V., Laporte Uribe, F., Meiland, F., Teupen, S., Felding, A.A., & Roes, M. (2023). Can technology impact loneliness in dementia? A scoping review on the role of assistive technologies in delivering psychosocial interventions in long-term care. *Disability and Rehabilitation: Assistive Technology*, *18*(7), 1107-1119. https://doi.org/10.1080/17483107.2021.1984594

Burholt, V., Winter, B., Aartsen, M., Constantinou, C., Dahlberg, L., Feliciano, V., De Jong Gierveld, J., Van Regenmortel, S., Waldegrave, C. (2020). A critical review and development of a conceptual model of exclusion from social relations for older people. *European Journal of Ageing*, *17*(1), 3-19. https://doi.org/10.1007/s10433-019-00506-0

Cacioppo, J. T. (2010) *Loneliness: Human nature and the need for social connection*. W.W. Norton.

Christiansen, M. B., Rafsanjani, A., & Jørgensen, J. (2024). "It brings the good vibes": exploring biomorphic aesthetics in the design of soft personal robots. *International Journal of Social Robotics*, *16*(5), 835-855. https://doi.org/10.1007/s12369-023-01037-6.

Corbett, C. F., Wright, P. J., Jones, K., & Parmer, M. (2021). Voice-activated virtual home assistant use and social isolation and loneliness among older adults: mini review. *Frontiers in Public Health*, *9*, 742012. https://doi.org/10.3389/fpubh.2021.742012

Craig, A. R., & Fisher, W. W. (2019). Randomization tests as alternative analysis methods for behavior-analytic data. *Journal of the Experimental Analysis of Behavior*, *111*(2), 309–328. https://doi.org/10.1002/jeab.500

Czaja, S. J., Boot, W. R., Charness, N., Rogers, W. A., & Sharit, J. (2018). Improving social support for older adults through technology: Findings from the PRISM randomized controlled trial. *The Gerontologist*, *58*(3), 467-477. https://doi.org/10.1093/geront/gnw249

Dahlberg, L., McKee, K. J., Frank, A., & Naseer, M. (2022). A systematic review of longitudinal risk factors for loneliness in older adults. *Aging & Mental Health*, *26*(2), 225-249. https://doi.org/10.1080/13607863.2021.1876638

Dang, N. B., & Bertrandias, L. (2025). Beyond companionship: Exploring the role of social robots in mitigating emotional and social loneliness of the elderly. *Journal of Macromarketing*, *45*(4), 612-631. https://doi.org/10.1177/02761467251363564

De Jong Gierveld, J. (1987). Developing and testing a model of loneliness. *Journal of Personality and Social Psychology*, *53*(1), 119-128. https://doi.org/10.1037//0022-3514.53.1.119

De Luca Picione, R. (2015). The idiographic approach in psychological research. The challenge of overcoming old distinctions without risking to homogenize. *Integrative Psychological and Behavioral Science*, *49*(3), 360–370. https://doi.org/10.1007/s12124-015-9307-5

Declercq, L., Cools, W., Beretvas, S. N., Moeyaert, M., Ferron, J. M., & Van den Noortgate, W. (2020). MultiSCED: A tool for (meta-)analyzing single-case experimental data with multilevel modeling. *Behavior Research Methods*, *52*(1), 177–192. https://doi.org/10.3758/s13428-019-01216-2

Deusdad, B. (2024). Ethical implications in using robots among older adults living with dementia. *Frontiers in Psychiatry*, *15*, 1436273. https://doi.org/10.3389/fpsyt.2024.1436273

Donovan, N. J., & Blazer, D. (2020). Social isolation and loneliness in older adults: review and commentary of a national academies report. *The American Journal of Geriatric Psychiatry*, *28*(12), 1233-1244. https://doi.org/10.1016/j.jagp.2020.08.005

Edgington, E. S., & Onghena, P. (2007). *Randomization tests* (4th ed.). Chapman & Hall/CRC.

Essig, L., Rotta, K., & Poling, A. (2023). Interobserver agreement and procedural fidelity: An odd asymmetry. *Journal of Applied Behavior Analysis, 56*(1), 78–85. https://doi.org/10.1002/jaba.961

Ferron, J. M., Bell, B. A., Hess, M. R., Rendina-Gobioff, G., & Hibbard, S. T. (2009). Making treatment effect inferences from multiple-baseline data: The utility of multilevel modeling approaches. *Behavior Research Methods, 41*(2), 372–384. https://doi.org/10.3758/BRM.41.2.372

Ferron, J. M., Farmer, J. L., & Owens, C. M. (2010). Estimating individual treatment effects from multiple-baseline data: A Monte Carlo study for multilevel-modeling approaches. *Behavior Research Methods, 42*(4), 930–943. https://doi.org/10.3758/BRM.42.4.930

Finnegan, D. J., & Campbell, S. (2023). Tackling loneliness and isolation in older adults with virtual reality: How do we move forward?. *Gerontology and Geriatric Medicine*, *9*, 23337214231186204. https://doi.org/10.1177/23337214231186204

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences, 115*(27), E6106-E6115. https://doi.org/10.1073/pnas.1711978115

Ganz, J. B., & Ayres, K. M. (2018). Methodological standards in single-case experimental design: Raising the bar. *Research in Developmental Disabilities*, *79*(1), 3–9. https://doi.org/10.1016/j.ridd.2018.03.003

Gerst-Emerson, K., & Jayawardhana, J. (2015). Loneliness as a public health issue: the impact of loneliness on health care utilization among older adults. *American Journal of Public Health*, *105*(5), 1013-1019. https://doi.org/10.2105/AJPH.2014.302427

Gursoy, D., Chi, O. H., Lu, L., & Nunkoo, R. (2019). Consumers acceptance of artificially intelligent (AI) device use in service delivery. *International Journal of Information Management*, *49*, 157–169. https://doi.org/10.1016/j.ijinfomgt.2019.03.008

Hagopian, L. P. (2020). The consecutive controlled case series: Design, data-analytics, and reporting methods supporting the study of generality. *Journal of Applied Behavior Analysis*, *53*(2), 596–619. https://doi.org/10.1002/jaba.691

Hausman, N. L., Javed, N., Bednar, M. K., Guell, M., Schaller, E., Nevill, R. E., & Kahng, S. (2022). Interobserver agreement: A preliminary investigation into how much is enough? *Journal of Applied Behavior Analysis*, *55*(2), 357–368. https://doi.org/10.1002/jaba.811

Hemberg, J., & Fischer, R. S. (2018). A window toward the world: older adults' experiences of becoming in health and developing as human beings through interacting with others using real video communication. *Holistic Nursing Practice*, *32*(2), 90-97. https://doi.org/10.1097/HNP.0000000000000254

Hoang, P., King, J. A., Moore, S., Moore, K., Reich, K., Sidhu, H., Tan, C. V., Whaley, C., & McMillan, J. (2022). Interventions associated with reduced loneliness and social isolation in older adults: A systematic review and meta-analysis. *JAMA network open*, *5*(10), e2236676. https://doi.org/10.1001/jamanetworkopen.2022.36676

Hox, J. J., Moerbeek, M., & van de Schoot, R. (2018). *Multilevel analysis: Techniques and applications* (3rd ed.). Routledge.

Hsu, H. C., Chang, S. C., Liang, C. Y., & Ting, J. R. (2024). Intervention to reduce loneliness among older adults in the community: Making friends with volunteers. *Geriatrics & Gerontology International*, *24*, 240-245. https://doi.org/10.1111/ggi.14828

Hung, L., Zhao, Y., Alfares, H., & Shafiekhani, P. (2025). Ethical considerations in the use of social robots for supporting mental health and wellbeing in older adults in long-term care. *Frontiers in Robotics and AI*, *12*, 1560214. https://doi.org/10.3389/frobt.2025.1560214

Inoue, K., Wada, K., & Shibata, T. (2021). Exploring the applicability of the robotic seal PARO to support caring for older persons with dementia within the home context. *Palliative Care and Social Practice*, *15*, 26323524211030285. https://doi.org/10.1177/263235242110302

Jacobs, K. W. (2019). Replicability and randomization test logic in behavior analysis. *Journal of the Experimental Analysis of Behavior*, *111*(2), 329–341. https://doi.org/10.1002/jeab.501

Johnson, A. H., & Cook, B. G. (2019). Preregistration in single-case design research. *Exceptional Children*, *86*(1), 95–112. https://doi.org/10.1177/0014402919868529

Jones, R. B., Ashurst, E. J., Atkey, J., & Duffy, B. (2015). Older people going online: its value and before-after evaluation of volunteer support. *Journal of Medical Internet Research*, *17*(5), e122. https://doi.org/10.2196/jmir.3943

Jones, V. K., Yan, C., Shade, M. Y., Boron, J. B., Yan, Z., Heselton, H. J., Johnson, K., & Dube, V. (2024). Reducing loneliness and improving social support among older adults through different modalities of personal voice assistants. *Geriatrics*, *9*(2), 22. https://doi.org/10.3390/geriatrics9020022

Kazdin, A. E. (1977). Assessing the clinical or applied importance of behavior change through social validation. *Behavior Modification*, *1*(4), 427–452. https://doi.org/10.1177/014544557714001

Kim, T., & Vossen, P. (2021). EmoBERTa: *Speaker-aware emotion recognition in conversation with RoBERTa*. https://doi.org/10.48550/arXiv.2108.12009

Klimova, B., Nguyen, A., & Cumbal, R. (2025). Ethical considerations of AI use by the elderly. *International Journal of Human–Computer Interaction, 41*(3), 210–222. https://doi.org/10.1080/10447318.2025.2531274

Koceska, N., Koceski, S., Beomonte Zobel, P., Trajkovik, V., & Garcia, N. (2019). A telemedicine robot system for assisted and independent living. *Sensors, 19*(4), 834. https://doi.org/10.3390/s19040834

Koceski, S., & Koceska, N. (2016). Evaluation of an assistive telepresence robot for elderly healthcare. *Journal of Medical Systems, 40*(5), 121. https://doi.org/10.1007/s10916-016-0481-x

Komulainen, K., Niemi, R., Gutvilig, M., Cachon Alonso, L., Hakulinen, C., & Elovainio, M. (2026). Loneliness and social isolation in transitions to adverse health conditions and mortality: An analysis of data from the UK Biobank study. *Journal of Epidemiology and Community Health*, *80*(4), 237-244. https://doi.org/10.1136/jech-2025-224665

Kotrlik, J. W., Williams, H. A., & Jabor, M. K. (2011). Reporting and interpreting effect size in quantitative agricultural education research. *Journal of Agricultural Education, 52*(1), 132–142. https://doi.org/10.5032/jae.2011.01132

Kramer, L. L., van Velsen, L., Clark, Jenna L., Mulder, B. C., & de Vet, E. (2022). Use and effect of Embodied Conversational Agents for improving eating behavior and decreasing loneliness among community-dwelling older adults: Randomized controlled trial. *JMIR Formative Research, 6*(4), e33974. https://doi.org/10.2196/33974

Kratochwill, T. R., & Levin, J. R. (2010). Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. *Psychological Methods, 15*(2), 124–144. https://doi.org/10.1037/a0017736

Kratochwill, T. R., & Levin, J. R. (2025). Randomization in single-case design experiments: Addressing threats to internal validity. *School Psychology, 40*(6), 754–768. https://doi.org/10.1037/spq0000685

Kratochwill, T. R., Levin, J. R., & Horner, R. H. (2018). Negative results: Conceptual and methodological dimensions in single-case intervention research. *Remedial and Special Education, 34*(1), 26–38. https://doi.org/10.1177/0741932517741721

Krippendorff, K. (2018). *Content Analysis: An Introduction to Its Methodology* (4th ed.). SAGE Publications.

Leineweber, M., Keusgen, C. V., Bubeck, M., Ranisch, R., Haltaufderheide, J., & Klingler, C. (2026). Ethical aspects of the use of social robots in caring for older people–a systematic qualitative review. *Medicine, Health Care and Philosophy*, 1-16. https://doi.org/10.1007/s11019-025-10313-3

Levinger, P., Panisset, M., Dunn, J., Haines, T., Dow, B., Batchelor, F., Biddle, S., Duque, G., & Hill, K. D. (2020). Exercise interveNtion outdoor proJect in the cOmmunitY for older people–results from the ENJOY Seniors Exercise Park project translation research in the community. *BMC Geriatrics*, *20*(1), 446. https://doi.org/10.1186/s12877-020-01824-0.

Li, C., Baek, E., & Luo, W. (2026). Generalized least squares transformation for single-case experimental design: Introducing the R package lmeSCED. *Behavior Research Methods, 58*(3), 72. https://doi.org/10.3758/s13428-025-02936-4

Li, H., Luo, W., Baek, E., Thompson, C. G., & Lam, K. H. (2022). Estimation and statistical inferences of variance components in the analysis of single-case experimental design using multilevel modeling. *Behavior Research Methods, 54*(4), 1559–1579. https://doi.org/10.3758/s13428-021-01691-6

Li, H., Luo, W., Baek, E., Thompson, C. G., & Lam, K. H. (2025). Multilevel modeling in single-case studies with count and proportion data: A demonstration and evaluation. *Psychological Methods, 30*(4), 815–842. https://doi.org/10.1037/met0000607

Liu, J., Wang, X., & Zhang, J. (2025). Investigating elderly individuals' acceptance of artificial intelligence (AI)-powered companion robots: The influence of individual characteristics. *Behavioral Sciences, 15*(5), 697. https://doi.org/10.3390/bs15050697

Loveys, K., Sagar, M., Pickering, I., & Broadbent, E. (2021). A digital human for delivering a remote loneliness and stress intervention to at-risk younger and older adults during the COVID-19 pandemic: randomized pilot trial. *JMIR Mental Health, 8*(11), e31586. https://doi.org/10.2196/31586

Manolov, R., & Moeyaert, M. (2025). Multilevel model selection applied to single-case experimental design data. *Journal of Behavioral Education*. Advance online publication. https://doi.org/10.1007/s10864-025-09593-9

Manolov, R., & Rochat, L. (2024). Analyzing data in single-case experimental designs: Objectives and available software options. *Journal of Behavioral and Cognitive Therapy*, *34* (4), article 100511. https://doi.org/10.1016/j.jbct.2024.100511

Martina, C. M. S., & Stevens, N. L. (2006). Breaking the cycle of loneliness? Psychological effects of a friendship enrichment program for older women. *Aging and Mental Health*, *10*(5), 467-475. https://doi.org/10.1080/13607860600637893

Marziali, R. A., Franceschetti, C., Dinculescu, A., Nistorescu, A., Kristály, D. M., Moșoi, A. A., Broekx, R., Marin, M., Vizitiu, C., Moraru, S-A., Rossi, L., & Di

Rosa, M. (2024). Reducing loneliness and social isolation of older adults through voice assistants: Literature review and bibliometric analysis. *Journal of Medical Internet Research*, *26*, e50534. https://doi.org/10.2196/50534

McGlynn, S. A., Kemple, S., Mitzner, T. L., King, C. H. A., & Rogers, W. A. (2017). Understanding the potential of PARO for healthy older adults. *International Journal of Human-Computer Studies*, *100*, 33-47. https://doi.org/10.1016/j.ijhcs.2016.12.004

Miehle, J., Bagci, I., Minker, W., & Ultes, S. (2018, August). A social companion and conversational partner for the elderly. In *Advanced social interaction with agents: 8th International workshop on spoken dialog systems* (pp. 103-109). Springer International Publishing. https://doi.org/10.1007/978-3-319-92108-2_12

Mival, O., Cringean, S., & Benyon, D. (2004). *Personification technologies: Developing artificial companions for older people*. CHI Fringe, Austria. Retrieved from https://www.researchgate.net/profile/Oli-Mival-2/publication/242522879_Personification_technologies_developing_artificial_companions_for_older_people/links/0f317533ed30c5b5d4000000/Personification-technologies-developing-artificial-companions-for-older-people.pdf

Moeyaert, M., Ferron, J., Beretvas, S., & Van den Noortgate, W. (2014). From a single-level analysis to a multilevel analysis of single-case experimental designs. *Journal of School Psychology, 52*(2), 191–211. https://doi.org/10.1016/j.jsp.2013.11.003

Moeyaert, M., Rindskopf, D., Onghena, P., & Van den Noortgate, W. (2017). Multilevel modeling of single-case data: A comparison of maximum likelihood and Bayesian estimation. *Psychological Methods, 22*(4), 760–778. https://doi.org/10.1037/met0000136

Mohammad, S. M. (2016). Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In H. L. Meiselman, *Emotion Measurement* (pp. 201–237). Woodhead Publishing. https://doi.org/10.1016/B978-0-08-100508-8.00009-6

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research and Perspectives, 2*(4), 201–218. https://doi.org/10.1207/s15366359mea0204_1

Natesan, P., & Hedges, L. V. (2017). Bayesian unknown change-point models to investigate immediacy in single case designs. *Psychological Methods, 22*(4), 743–759. https://doi.org/10.1037/met0000134

Natesan Batley, P., Minka, T. & Hedges, L. V. (2020). Investigating immediacy in multiple-phase-change single-case experimental designs using a Bayesian unknown change-points model. *Behavior Research Methods, 52*(4), 1714–1728. https://doi.org/10.3758/s13428-020-01345-z

Normand, M. P. (2016). Less is more: Psychologists can learn more by studying fewer people. *Frontiers in Psychology, 7*, e934. https://doi.org/10.3389/fpsyg.2016.00934

Oh, K. M., Hong, S. R., Beran, K., Sanders, L., Park, J. Y., & Lee, J. A. (2026). Utilizing conversational AI technology for social connectedness among older adults: a systematic review. *Journal of Applied Gerontology*, *45*(3), 527-549. https://doi.org/10.1177/07334648251341629

Onghena, P., Maes, B., & Heyvaert, M. (2019). Mixed methods single case research: State of the art and future directions. *Journal of Mixed Methods Research, 13*(4), 461–480. https://doi.org/10.1177/1558689818789530

Pérez, J. M., Furman, D. A., Alonso Alemany, L., & Luque, F. (2022). RoBERTuito: a pre-trained language model for social media text in Spanish. *Proceedings of the 13th*

Language Resources and Evaluation Conference (LREC), 7235–7243. https://aclanthology.org/2022.lrec-1.785.pdf

Petit-Bois, M., Baek, E. K.,Van den Noortgate, W., Beretvas, S. N., & Ferron, J. M. (2016). The consequences of modeling autocorrelation when synthesizing single-case studies using a three-level model. *Behavior Research Methods*, *48*(2), 803–812. https://doi.org/10.3758/s13428-015-0612-1

Piccirillo, M. L., Beck, E. D., & Rodebaugh, T. L. (2019). A clinician's primer for idiographic research: Considerations and recommendations. *Behavior Therapy*, *50*(5), 938–951. https://doi.org/10.1016/j.beth.2019.02.002

Poria, S., Majumder, N., Mihalcea, R., & Hovy, E. (2019). Emotion recognition in conversation: Research challenges, datasets, and recent advances. *IEEE Access*, *7*, 100943–100953. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8764449

Poscia, A., Stojanovic, J., La Milia, D. I., Duplaga, M., Grysztar, M., Moscato, U., Onder, G., Collamati, A., Ricciardi, W., & Magnavita, N. (2018). Interventions targeting loneliness and social isolation among the older people: An update systematic review. *Experimental Gerontology*, *102*, 133-144. https://doi.org/10.1016/j.exger.2017.11.017

Pustejovsky, J. E., & Swan, D. M. (2015). Four methods for analyzing partial interval recording data, with application to single-case research. *Multivariate Behavioral Research*, *50*(3), 365–380. https://doi.org/10.1080/00273171.2015.1014879

Quan, N. G., Lohman, M. C., Resciniti, N. V., & Friedman, D. B. (2020). A systematic review of interventions for loneliness among older adults living in long-term care facilities. *Aging & Mental Health*, *24*(12), 1945-1955. https://doi.org/10.1080/13607863.2019.1673311

Radley, K. C., O'Handley, R. D., & Labrot, Z. C. (2015). A comparison of momentary time sampling and partial-interval recording for assessment of effects of social skills training. *Psychology in the Schools*, *52*(4), 363–378. https://doi.org/10.1002/pits.21829

Rajamäki, J., & Helin, J. (2024). The ethics and cybersecurity of artificial intelligence and robotics in helping the elderly to manage at home. *Information*, *15*(11), 729. https://doi.org/10.3390/info15110729

Rimé, B., Finkenauer, C., Luminet, O., Zech, E., & Philippot, P. (1998). Social sharing of emotion: New evidence and new questions. *European Review of Social Psychology*, *9*(1), 145–189. https://doi.org/10.1080/14792779843000072

Rodríguez-Prada, C., Martínez-Huertas, J. Á., & Olmos, R. (2026). Bayesian versus frequentist approaches in multilevel single-case designs: On Type I error rate and power. *Methodology*, *22*(1), 52–76. https://doi.org/10.5964/meth.17715

Rodríguez-Martínez, A., Amezcua-Aguilar, T., Cortés-Moreno, J., & Jiménez-Delgado, J. J. (2023). Qualitative analysis of conversational chatbots to alleviate loneliness in older adults as a strategy for emotional health. *Healthcare*, *12*(1), 62. https://doi.org/10.3390/healthcare12010062

Sha, S., Loveys, K., Qualter, P., Shi, H., Krpan, D., & Galizzi, M. (2024). Efficacy of relational agents for loneliness across age groups: A systematic review and meta-analysis. *BMC Public Health*, *24*(1), 1802. https://doi.org/10.1186/s12889-024-19153-x

Shaked, N. A. (2017). Avatars and virtual agents–relationship interfaces for the elderly. *Healthcare Technology Letters*, *4*(3), 83–87. https://doi.org/10.1049/htl.2017.0009

Sharkey, A., & Sharkey, N. (2021). We need to talk about deception in social robotics! A. Sharkey, N. Sharkey. *Ethics and Information Technology*, *23*(3), 309-316. https://doi.org/10.1007/s10676-020-09573-9

Sharpley, C. F., & Alavosius, M. P. (1988). Autocorrelation in behavior data: An alternative perspective. *Behavioral Assessment*, *10*, 243–251.

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, *4*, 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Shoesmith, E., Clarke, C., McPherson, G., & Ratschen, E. (2024). Using PARO, a robotic seal, to support people living with dementia:‘what works’ in inpatient dementia care settings?. *Human-Animal Interactions*, *12*(1). https://doi.org/10.1079/hai.2024.0023

Shvedko, A., Whittaker, A. C., Thompson, J. L., & Greig, C. A. (2018). Physical activity interventions for treatment of social isolation, loneliness or low social support in older adults: A systematic review and meta-analysis of randomised controlled trials. *Psychology of Sport and Exercise*, *34*, 128-137. https://doi.org/10.1016/j.psychsport.2017.10.003

Sideridis, G. D., & Greenwood, C. R. (1997). Is human behavior autocorrelated? An empirical analysis. *Journal of Behavioral Education*, *7*(3), 273–293. https://doi.org/10.1023/A:1022895805201

Snodgrass, M., Cook, B. G., & Cook, L. (2023). Considering social validity in special education research. *Learning Disabilities Research & Practice*, *38*(4), 311–319. https://doi.org/10.1111/ldrp.12326

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science, 11*(5), 702–712. https://doi.org/10.1177/1745691616658637

Stinson, L., Liu, Y., & Dallery, J. (2022). Ecological momentary assessment: A systematic review of validity research. *Perspectives on Behavior Science*, *45*(2), 469–493. https://doi.org/10.1007/s40614-022-00339-w

Su, Z., Sheng, W., Yang, G., Bishop, A., & Carlson, B. (2022). Adaptation of a robotic dialog system for medication reminder in elderly care. *Smart Health*, *26*, 100346. https://doi.org/10.1016/j.smhl.2022.100346

Sun, B., & Chen, Y. (2026). How do consumers perceive virtual agents with behavioral realism? The role of loneliness and trust. *Frontiers in Psychology*, *17*, 1676234. https://doi.org/10.3389/fpsyg.2026.1676234

Suta, P., Chan, J. H., & Mongkolnam, P. (2025, October). Enhancing elderly well-being through ai-based virtual companion systems: A cooperative approach. In *International Conference on Cooperative Design, Visualization and Engineering* (pp. 86-99). Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-06952-8_10

Takeda, Y., Huang, H. H., & Kawagoe, K. (2014, June). Toward a companion agent for the elderly–the methods to estimate at-home and outside-home daily life activities of the elderly who live alone. In *International Conference on Universal Access in Human-Computer Interaction* (pp. 395-402). Springer International Publishing. https://doi.org/10.1007/978-3-319-07446-7_39

Tanious, R., & Onghena, P. (2021). A systematic review of applied single-case research published between 2016 and 2018: Study designs, randomization, data aspects, and data analysis. *Behavior Research Methods*, *53*(4), 1371–1384. https://doi.org/10.3758/s13428-020-01502-4

Tate, R. L., Perdices, M., Rosenkoetter, U., Shadish, W., Vohra, S., Barlow, D. H., Horner, R., Kazdin, A., Kratochwill, T. R., McDonald, S., Sampson, M., Shamseer, L., Togher, L., Albin, R., Backman, C., Douglas, J., Evans, J. J., Gast, D., Manolov, R., Mitchell, G., Nickels, L., Nikles, J., Ownsworth, T., Rose, M., Schmid, C. H., & Wilson, B. (2016). The Single-Case Reporting Guideline In BEhavioural Interventions (SCRIBE) 2016 statement. *Archives of Scientific Psychology, 4*(1), 1–9. https://doi.org/10.1037/arc0000026

Tincani, M., Travers, J., Dowdy, A., Slocum, T. A., & Deitrich, R. (2025). Questionable and improved research practices in single-case experimental design: Initial investigation and findings. *Perspectives on Behavior Science, 48*(2), 447–473. https://doi.org/10.1007/s40614-025-00441-9

Veldmeijer, L., Wartena, B., Terlouw, G., & Van't Veer, J. (2020). Reframing loneliness through the design of a virtual reality reminiscence artefact for older adults. *Design for Health*, *4*(3), 407-426. https://doi.org/10.1080/24735132.2020.1848976

Verboon, P., & Peters, G. J. (2020). Applying the generalized logistic model in single case designs: Modeling treatment-induced shifts. *Behavior Modification*, *44*(1), 27–48. https://doi.org/10.1177/0145445518791255

What Works Clearinghouse. (2022). *Procedures and Standards Handbook, Version 5.0*. U.S. Department of Education, Institute of Education Sciences. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/Final_WWC-HandbookVer5.0-0-508.pdf

Wicherts, J. M., Veldkamp, C. L., Augusteijn, H. E., Bakker, M., van Aert, R. C., & Van Assen, M. A. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in Psychology*, *7*, article 1832. https://doi.org/10.3389/fpsyg.2016.01832

Yan, C., Johnson, K., & Jones, V. K. (2024). The impact of interaction time and verbal engagement with personal voice assistants on alleviating loneliness among older adults: an exploratory study. *International Journal of Environmental Research and Public Health*, *21*(1), 100. https://doi.org/10.3390/ijerph21010100

Yoder, P. J., Ledford, J. R., Harbison, A. L., & Tapp, J. T. (2018). Partial-interval estimation of count: Uncorrected and Poisson-corrected error levels. *Journal of Early Intervention*, *40*(1), 39–51. https://doi.org/10.1177/1053815117748407

Yoder, P., & Symons, F. (2010). *Observational measurement of behavior*. Springer.

Zheng, Y., Qualter, P., Rollano, C., Hu, X., Martin, A., Nukui, M., Walibhai, W., & Maes, M. (2026). Impact of loneliness on health in healthy populations: A meta-analysis. *British Journal of Health Psychology*, *31*(1), e70040. https://doi.org/10.1111/bjhp.70040

Zeng, H., Sheng, Y., & Zhu, J. (2025). Companion robots supporting the emotional needs of the elderly: Research trends and future directions. *Information*, *16*(11), 948. https://doi.org/10.3390/info16110948

Zuidersma, M., Riese, H., Snippe, E., Booij, S. H., Wichers, M., & Bos, E. H. (2020). Single-subject research in psychiatry: Facts and fictions. *Frontiers in Psychiatry*, *11*, 1174. https://doi.org/10.3389/fpsyt.2020.539777

Table 1

Participant characteristics

Name	Sex	Age	Marital status	Educational level	UCLA loneliness Score
Amber	1	88	Single	University	5
Elsa	1	91	Widow	Primary, not completed	4
Joana	1	89	Divorced	Primary, not completed	5
Juana	1	83	Widow	Primary, not completed	5
Marc	2	90	Married	Primary, not completed	7
Penelope	1	64	Widow	FP superior	9
Albert	2	66	Divorced	Primary	6
Bradley	2	90	Single	Primary, not completed	6
Claudia	1	82	Married	Primary	6
Hans	2	95	Widower	Primary, not completed	7
Jesse	1	63	Single	Primary	4
Jared	2	75	Single	Primary, not completed	9
Monica	1	84	Widow	Primary	5
Olivia	1	88	Widow	Primary	7
Ophelia	1	72	Widow	Primary	6
Rebecca	1	92	Widow	Primary, not completed	7
Riley	1	88	Widow	none	7

Table 2

Summary of use of the ECA

Participant	Days Use	Conse-cutive days	Days with more than one interaction	Average interactions per day	Total interactions	Brief interactions	Average turns per interaction	% Joy	% Sadness	% Anger
Hans	21	18	17	4.57	96	12	36.39	12.98	16.41	5.21
Ophelia	8	7	5	3.00	24	4	14.71	14.87	5.42	3.37
Claudia	18	8	14	2.89	52	9	11.83	23.15	14.23	8.81
Albert	6	4	1	1.33	8	4	9.25	23.93	8.93	0.00
Bradley	16	3	4	1.31	21	6	7.33	27.91	4.76	8.89
Jared	24	4	8	1.63	39	8	10.83	20.18	9.11	0.00
Monica	6	4	4	2.00	12	5	15.33	25.19	23.78	0.00
Olivia	13	4	3	1.54	20	4	10.00	52.00	3.92	1.00
Rebecca	30	5	16	1.73	52	14	31.88	41.21	20.98	0.00
Amber	16	3	5	1.94	31	25	6.71	9.73	0.46	0.00
Jennifer	14	5	3	1.36	19	5	40.42	37.7	16.20	0.04

								7		
Elsa	28	4	15	7.61	213	114	8.55	17.04	11.56	6.81
Marc	6	3	6	6.50	39	23	5.49	2.22	6.67	6.67
Penelope	7	1	2	1.71	12	4	36.50	16.66	8.23	0.00
Jake	12	2	5	1.50	18	7	21.89	21.19	9.00	0.35
Jesse	10	2	4	1.80	18	10	12.00	34.48	0.43	0.93
Riley	25	24	4	1.28	32	6	8.63	50.22	8.68	0.00
Jodie	31	3	9	1.61	50	9	15.00	42.54	10.92	0.00

Note. Participants are sorted chronologically, according to the data in which day started using the ECA. Brief interactions include only one turn for the participant and one for ECA.

Table 3

Within-individual Person's correlations between emotions and use of the ECA

Participant	Correlation Words-Turns	Correlation Joy-Turns	Correlation Sadness-Turns
Hans	0.94	0.07	0.18
Ophelia	0.77	0.25	0.28
Claudia	0.81	0.11	0.40
Albert	0.90	0.26	0.24
Bradley	0.92	0.30	0.03
Jared	0.81	0.33	0.40
Monica	0.82	-0.06	0.79
Olivia	0.89	0.35	0.61
Rebecca	0.96	0.29	0.52
Amber	0.98	0.34	0.62
Jennifer	0.98	0.35	-0.04
Elsa	0.63	0.25	0.55
Marc	0.61	0.40	0.44
Penelope	0.99	-0.21	0.66
Jake	0.98	0.56	0.11
Jesse	0.93	0.54	0.21
Riley	0.97	0.43	0.08
Jodie	0.84	0.17	0.60

Note. Participants are sorted chronologically, according to the data in which day started using the ECA. Weak correlations (.10-.29) marked in yellow; moderate correlations (.30-.49) marked in blue; strong correlations (.50-1.00) marked in green.

Table 4

Across-individuals Pearson's correlations between emotions and use of the ECA

Use indicator	Joy	Sadness
Days of use	0.37	0.25
Consecutive days	0.23	0.20
Days with more than one interaction	-0.13	0.47
Average interactions per day	-0.56	0.07
Total interactions	-0.21	0.20
Brief interactions	-0.27	0.01
Average turns per interaction	0.06	0.52

Note. Brief interactions include only one turn for the participant and one for the ECA. Weak correlations (.10-.29) marked in yellow; moderate correlations (.30-.49) marked in blue; strong correlations (.50-1.00) marked in green.

Table 5

Results of the multilevel analyses for Joy

	Mean	SD across cases	SD across days
Initial level	**29.40**	16.32	18.20
Change per interaction	**-0.10**	0.11	0.25
Initial level	**9.56**	32.55	18.45
Change per interaction	**-0.37**	0.44	0.25
Turns per interaction	**0.13**		
Days of use	**1.36**		

Note. SD – standard deviation. Results above the line refer to a model without moderators, whereas results below the line refer to a model with moderators (included in the last two rows). The cell marked in green denotes a statistically significant association.

Table 6

Results of the multilevel analyses for Sadness

	Mean	SD across cases	SD across days
Initial level	**9.75**	9.58	3.44
Change per interaction	**-0.01**	0.10	0.04
Initial level	**4.24**	1.65	1.20
Change per interaction	**-0.01**	0.02	0.02
Turns per interaction	**0.23**		
Days of use	**0.12**		

Note. SD – standard deviation. Results above the line refer to a model without moderators, whereas results below the line refer to a model with moderators (included in the last two rows). The cell marked in green denotes a statistically significant association.

Table 7

Individual effects from the multilevel models

Participant	Joy	Sadness
Hans	-0.63	0.17
Ophelia	-0.56	0.16
Claudia	-0.48	0.20
Albert	-0.39	0.14
Bradley	-0.44	0.10
Jared	-0.33	0.11
Monica	-0.44	0.17
Olivia	-0.18	0.09
Rebecca	-0.17	0.23
Amber	-0.63	0.08
Jennifer	-0.23	0.18
Elsa	-0.43	0.08
Marc	-0.66	0.13
Penelope	-0.52	0.16
Jake	-0.57	0.15
Jesse	-0.44	0.09
Riley	-0.01	0.10
Jodie	-0.13	0.19

Table 8

Results of the logistic model for both emotions

Participant	Emotion	Lower asymptote	Upper asymptote	Rate of change	Moment of inflection
Hans	Joy	2.00	64.67	-0.02	12.12
Hans	Sadness	2.00	98.00	-0.03	6.74
Ophelia	Joy	2.00	64.67	0.53	11.93
Ophelia	Sadness	2.00	38.00	0.23	22.00
Claudia	Joy	2.00	98.00	-0.03	12.16
Claudia	Sadness	2.00	68.00	-0.04	12.72
Albert	Joy	2.00	98.00	-0.99	6.00
Albert	Sadness	2.00	55.14	0.06	6.00
Bradley	Joy	2.00	98.00	-0.11	10.14
Bradley	Sadness	2.00	73.00	0.30	19.00
Jared	Joy	2.00	98.00	-0.06	4.85
Jared	Sadness	2.00	78.00	-0.03	13.86
Monica	Joy	2.00	68.00	-0.09	7.55
Monica	Sadness	2.00	88.00	0.21	10.00
Olivia	Joy	2.00	98.00	-0.24	11.52
Olivia	Sadness	2.00	31.33	0.11	14.84
Rebecca	Joy	2.00	101.00	-0.01	14.16
Rebecca	Sadness	2.00	74.60	0.53	11.15
Amber	Joy	2.00	98.00	-0.08	11.44

Amber	Sadness	2.00	93.45	-0.03	14.03
Jennifer	Joy	2.00	91.94	-0.10	9.17
Jennifer	Sadness	2.00	101.00	-0.22	3.00
Elsa	Joy	2.00	98.00	0.17	11.42
Elsa	Sadness	2.00	98.00	0.17	13.52
Marc	Joy	2.00	68.00	-0.09	7.55
Marc	Sadness	2.00	68.00	-0.16	3.00
Penelope	Joy	2.00	64.67	0.31	10.00
Penelope	Sadness	2.00	37.08	0.60	10.00
Jake	Joy	2.00	37.08	0.60	10.00
Jake	Sadness	2.00	90.59	0.24	16.00
Jesse	Joy	2.00	98.00	-0.08	8.50
Jesse	Sadness	2.00	5.96	0.87	16.00
Riley	Joy	2.00	98.00	0.03	9.65
Riley	Sadness	2.00	69.43	0.00	15.13
Jodie	Joy	2.00	98.00	-0.02	15.21
Jodie	Sadness	2.00	93.24	0.43	12.28

Note. Blue values denote an increase, red values denote a reduction.

Table 9

Results of the Bayesian unknown change point model for both emotions

Participant	Variable	Moment of change	Interval: lower limit	Interval: Upper limit	Total interactions	Advance	Change
Hans	Joy	57	52	94	96	98%	Reduction
Hans	Sadness	89	89	89	96	93%	Reduction
Ophelia	Joy	10	10	10	24	42%	Increase
Ophelia	Sadness	10	10	10	24	42%	Increase
Claudia	Joy	50	50	50	52	96%	Increase
Claudia	Sadness	49	49	49	52	94%	Reduction
Albert	Joy	NA	NA	NA	8	NA	
Albert	Sadness	3	3	3	8	38%	Reduction
Bradley	Joy	4	4	4	21	19%	Reduction
Bradley	Sadness	16	16	17	21	81%	Increase
Jared	Joy	18	18	18	39	46%	Reduction
Jared	Sadness	17	15	26	39	67%	Reduction
Monica	Joy	7	7	7	12	58%	Reduction
Monica	Sadness	4	4	4	12	33%	Increase
Olivia	Joy	10	10	10	20	50%	Reduction
Olivia	Sadness	15	5	18	20	90%	Reduction
Rebecca	Joy	NA	NA	NA	52	NA	
Rebecca	Sadness	NA	NA	NA	52	NA	
Amber	Joy	11	11	11	31	35%	Reduction

Amber	Sadness	5	5	5	31	16%	Reduction
Jennifer	Joy	13	13	13	19	68%	Reduction
Jennifer	Sadness	15	10	15	19	79%	Reduction
Elsa	Joy	79	79	79	213	37%	Reduction
Elsa	Sadness	127	125	141	213	66%	Reduction
Marc	Joy	26	22	34	39	87%	Reduction
Marc	Sadness	8	8	8	39	21%	Reduction
Penelope	Joy	NA	NA	NA	12	NA	
Penelope	Sadness	8	7	8	12	67%	Increase
Jake	Joy	6	6	6	18	33%	Reduction
Jake	Sadness	12	12	13	18	72%	Increase
Jesse	Joy	11	11	11	18	61%	Reduction
Jesse	Sadness	9	3	16	18	89%	Reduction
Riley	Joy	22	22	22	32	69%	Increase
Riley	Sadness	25	22	25	32	78%	Reduction
Jodie	Joy	13	13	13	50	26%	Increase
Jodie	Sadness	15	10	15	50	30%	Reduction

Note. NA – results not available, as the model did not converge after hours of analyses.

Table 10

Comparing models

			Logistic Model		Bayesian model		Multilevel model	
Participant	Days of use	Interactions	Joy	Sadness	Joy	Sadness	Joy	Sadness
Hans	21	96	Reduction	Reduction	Reduction	Reduction	Reduction	Increase
Ophelia	8	24	Increase	Increase	Increase	Increase	Reduction	Increase
Claudia	18	52	Reduction	Reduction	Reduction	Increase	Reduction	Increase
Albert	6	8	Reduction	Increase		Reduction	Reduction	Increase
Bradley	16	21	Reduction	Reduction	Reduction	Increase	Reduction	Increase
Jared	24	39	Reduction	Increase	Reduction	Reduction	Reduction	Increase
Monica	6	12	Reduction	Reduction	Reduction	Increase	Reduction	Increase
Olivia	13	20	Reduction	Increase	Reduction	Reduction	Reduction	Increase
Rebecca	30	52	Reduction	Increase			Reduction	Increase
Amber	16	31	Reduction	Reduction	Reduction	Reduction	Reduction	Increase
Jennifer	14	19	Reductio	Reductio	Reductio	Reductio	Reductio	Increase

			n	n	n	n	n	
Elsa	28	213	Increase	Increase	Reductio n	Reductio n	Reductio n	Increase
Marc	6	39	Reductio n	Reductio n	Reductio n	Reductio n	Reductio n	Increase
Penelope	7	12	Increase	Increase		Increase	Reductio n	Increase
Jake	12	18	Increase	Increase	Reductio n	Increase	Reductio n	Increase
Jesse	10	18	Reductio n	Increase	Reductio n	Reductio n	Reductio n	Increase
Riley	25	32	Increase	Increase	Increase	Reductio n	Reductio n	Increase
Jodie	31	50	Reductio n	Increase	Increase	Reductio n	Reductio n	Increase

Table 11

Changes between first and last emotion across interactions

Participant	Most frequent change	Frequency	Total Initial Neutral	% change of initial	Total interactions	% change of total
Hans	Neutral to Sadness	9	54	17%	96	9%
Ophelia	Neutral to Joy	7	20	35%	24	29%
Claudia	Neutral to Joy	12	39	31%	52	23%
Albert	Neutral to Joy	2	7	29%	8	25%
Bradley	Neutral to Joy	4	15	27%	21	19%
Jared	Neutral to Joy	14	35	40%	39	36%
Monica	Neutral to Joy	3	9	33%	12	25%
Olivia	Neutral to Joy	9	16	56%	20	45%
Rebecca	Neutral to Joy	19	40	48%	52	37%
Amber	Neutral to Joy	4	30	13%	31	13%
Jennifer	Neutral to Joy	6	17	35%	19	32%
Elsa	Neutral to Joy	28	193	15%	213	13%
Marc	Neutral to Joy and to Sadness	2	33	6%	39	5%
Penelope	Neutral to Joy	2	11	18%	12	17%
Jake	Neutral to Joy	4	17	24%	18	22%
Jesse	Neutral to Joy	4	14	29%	18	22%
Riley	Neutral to Joy	14	21	67%	24	58%

Jodie	Neutral to Joy	3	9	33%	10	30%

Table 12

Frequency of identified positive aspects by category.

Category	Freq
Emotional Support	10
Communicative style	8
Respect	6
Biographical validation	4
Practical Utility	3

Table 13

Frequency of reasons for recommending the ECA by category.

Category	Freq
Emotional Support	11
Adaptation to vulnerabilities	5
Cognitive Stimulation	4
Leisure and Entertainment	4

Table 14

Frequency of identified areas for improvement by category.

Category	Freq
Unnatural conversation	7
Lack of proactivity	6
Usability problems	6
Lack of adaptability to limitations	5
Emotional burden	3

Table 15

Frequency of identified topics of conversation by category.

Category	Freq
Leisure and Interests	13
Health and wellbeing	10
Social bonds	7
Life trajectory and reminiscences	7
Conversational agent	2

Table 16

Frequency of aspects of learning of conversation by category.

Category	Freq
Emotional Regulation	9
Knowledge Acquisition	5
Cognitive Stimulation	4
Technological Competence	3

Table 17

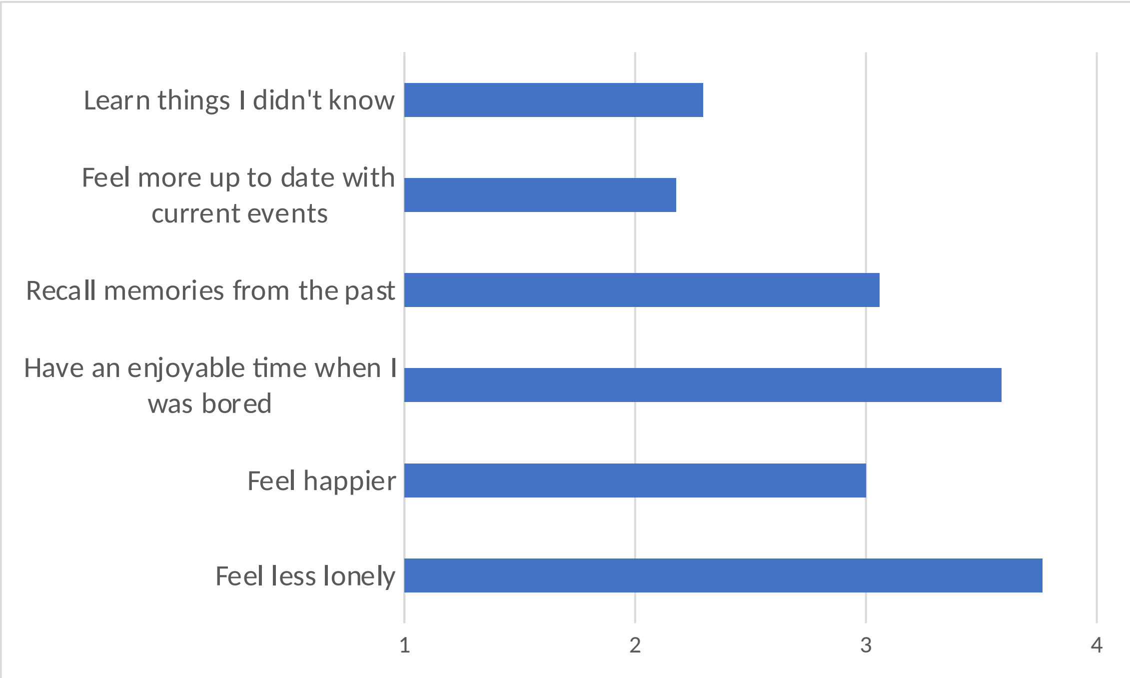
Frequency of evolution of interaction by category.

Category	Freq
More emotional Openness	5
More Inclusion in Daily Routines	4
More Technical Adaptation	3

Figure 1

Figure 2

Perceived benefits



Appendix: Initial Plan of the Quantitative Study

Design

The initial plan was to use a single-case experimental design (SCEDs): a research methodology that allows for providing robust scientific evidence on the effectiveness of an intervention (Horner et al., 2005; Ledford et al., 2019; What Works Clearinghouse, 2022) and have been recommended for a variety of contexts such as general education (Plavnick & Ferreri, 2013), special education (Alnahdi, 2015), and psychiatry (Zuidersma et al., 2020), neurorehabilitation (Krasny-Pacini & Evans, 2018), occupational therapy (Johnston & Smith, 2010), pediatric psychology (Cohen et al., 2012), learning disabilities (Contesse et al., 2021), organizational management (Erath et al., 2021), etc. In SCEDs, multiple measurements are obtained that also belong to different conditions actively manipulated by the researchers (Morley, 2018; Tate & Perdices, 2019). Each unit is compared with itself.

The most common SCED is the multiple baseline design (Shadish & Sullivan, 2011; Tanious & Onghena, 2021), which involves replicating the AB sequence (where “A” represents the baseline and “B” the intervention phase) with different individuals, behaviors of the same individual, or contexts in which the same behavior of the same individual is studied. The key is that the intervention is introduced in a staggered manner, that is, at different points in time for the different baselines. Depending on whether data collection begins simultaneously for all participants or not, the baseline design is concurrent or non-concurrent (see Slocum et al., 2022). Different ways of introducing randomization are possible in a multiple baseline design (Levin et al., 2018).

The second most common type of randomized controlled trial design is the withdrawal or reversal design (Wine et al., 2015). This design involves replication within the same participant and can be denoted as ABA, BAB, or, if more phases are available (as is recommended; What Works Clearinghouse, 2022), ABAB. Several comparison points between adjacent phases are available. The design is applicable when the intervention does not produce a permanent change (e.g., when it is a behavior modification technique based on the principles of behaviorism). Phase transition points can be randomly selected from a list of possibilities that respects a minimum length for each phase (Onghena, 1992).

It is possible to combine several types of SCED in the same study to increase internal validity or to allow for answering a wider variety of research questions (Tanious & Manolov, 2023). Measurements were planned to be taken for each participant over 15 days. The A phase would be ECA in a reactive mode and the B phase, ECA in a proactive mode. In the present study, the plan was to combine a concurrent multiple baseline design with three participants was with a BAB reversal design, plus a participant for whom the ECA is only proactive. For each of the three participants following the BAB design, the phase change points (ECA mode) were randomly selected, maintaining a minimum of three consecutive days within the same phase. With a total of 15 days and the minimum of three per phase, there are 28 sequences (phase change points) for each participant following the BAB design. Note that the order of the phases (ECA modes) is not randomized, but rather the phase change points, with the order always being BAB (proactive-reactive-proactive).

Two replications would be conducted with another four participants (three following a BAB design and one a B design), for a total of twelve participants. Replication is important for both internal validity—demonstrating a causal (or functional) relationship between the condition and the behaviors of interest—and external validity—generalizability of the results (Morley, 2018; Tate & Perdices, 2019).

Data Analysis

A Priori Plan: Tentative Causal Inference

In relation to the planned presence of randomization in the design, a randomization test was intended to be used. The aim was to randomly determine the length of the phases separately for each person following a BAB design via the procedure described in Onghena (1992), instead of the procedures applicable to multiple baseline designs (Levin et al., 2018). The randomization test would allow us to quantify the probability of obtaining a difference as large as the observed one (or larger), if in fact there were no difference. To achieve this quantification, all possible sequences are considered (i.e., the combinations of phase lengths in the BAB design). The probability obtained is a *p*-value that would allow making a statistical decision: to reject or not reject the null hypothesis that the mean of the behavior of interest is the same in the different phases (i.e., that there is no difference between the reactive and proactive modes of the ECA). A *p*-value could be obtained for each individual. Within each cohort, *p*-values could be combined using the procedure described by Edgington (1972; see also Onghena et al., 2018). The analysis was planned to be performed via the website https://tamalkd.shinyapps.io/scda (De et al., 2017). Note that the *p*-value, in conjunction with the design characteristics, allows for a tentative inference of a causal relationship for the participants studied, not a population-level inference (Onghena, 2020). The absence of distinction phases made impossible both the comparisons and the randomization and, therefore, a randomization test was not used.

A Priori Plan: Quantification of Magnitude of Effect via a Multilevel Model

Multilevel models (Ferron et al., 2009; Moeyaert et al., 2014) were planned to be used to quantify the difference between the initial proactive condition and the subsequent reactive condition (i.e., B-A comparison), considering the nesting (dependence) of the data: interactions, nested into participants, nested into cohorts. The quantifications would be expressed in the same unit of measurement as the variable of interest (i.e., they are not standardized values). Multilevel modeling would allow for quantifying differences in means between phases or incorporating trends across different phases and quantifying the difference between trends. The modification of the use of the multilevel analysis is described next.

References

Alnahdi, G. H. (2015). Single-subject designs in special education: Advantages and limitations. *Journal of Research in Special Educational Needs, 15*(4), 257–265. https://doi.org/10.1111/1471-3802.12039

Cohen, L. L., Feinstein, A., Masuda, A., & Vowles, K. E. (2014). Single-case research design in pediatric psychology: Considerations regarding data analysis. *Journal of Pediatric Psychology, 39*(2), 124−137. https://doi.org/10.1093/jpepsy/jst065

Contesse, V. A., Gage, N. A., & Lane, H. B. (2023). Practical use of single-case research design to target improved behavioral outcomes for students with SLD. *Learning Disability Quarterly, 46*(1), 32–45. https://doi.org/10.1177/07319487211041127

De, T. K., Michiels, B., Vlaeyen, J. W. S., & Onghena, P. (2017, April). *One by one: A shiny web-app for the design and analysis of single-case experiments*. Paper presented at the useR! Conference 2017 Brussels, Belgium. https://www.r-project.org/conferences/useR-2017/uploads/Tamal-Kumar-De_170623_103005.pdf

Edgington, E. S. (1972). An additive method for combining probability values from independent experiments. *The Journal of Psychology, 80*(2), 351−363. https://doi.org/10.1080/00223980.1972.9924813

Erath, T. G., Pellegrino, A. J., DiGennaro Reed, F. D., Ruby, S. A., Blackman, A. L., & Novak, M. D. (2021). Experimental research methodologies in organizational behavior management. *Journal of Organizational Behavior Management*, *41*(2), 150–181. https://doi.org/10.1080/01608061.2020.1869137

Ferron, J. M., Bell, B. A., Hess, M. R., Rendina-Gobioff, G., & Hibbard, S. T. (2009). Making treatment effect inferences from multiple-baseline data: The utility of multilevel modeling approaches. *Behavior Research Methods, 41*(2), 372–384. https://doi.org/10.3758/BRM.41.2.372

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*(2), 165−179. https://doi.org/10.1177/001440290507100203

Johnston, M. V., & Smith, R. O. (2010). Single subject designs: Current methodologies and future directions. *OTJR: Occupation, Participation and Health, 30*(1), 4–10. https://doi.org/10.3928/15394492-20091214-02

Krasny-Pacini, A., & Evans, J. (2018). Single-case experimental designs to assess intervention effectiveness in rehabilitation: A practical guide. *Annals of Physical and Rehabilitation Medicine*, *61*(3), 164–179. https://doi.org/10.1016/j.rehab.2017.12.002

Ledford, J. R., Barton, E. E., Severini, K. E., & Zimmerman, K. N. (2019). A primer on single-case research designs: Contemporary use and analysis. *American Journal on Intellectual and Developmental Disabilities, 124*(1), 35–56. https://doi.org/10.1352/1944-7558-124.1.35

Levin, J. R., Ferron, J. M., & Gafurov, B. S. (2018). Comparison of randomization-test procedures for single-case multiple-baseline designs. *Developmental Neurorehabilitation*, *21*(5), 290–311. https://doi.org/10.1080/17518423.2016.1197708

Moeyaert, M., Ferron, J., Beretvas, S., & Van den Noortgate, W. (2014). From a single-level analysis to a multilevel analysis of single-case experimental designs. *Journal of School Psychology*, *52*(2), 191–211. https://doi.org/10.1016/j.jsp.2013.11.003

Morley, S. (2018). *Single-case methods in clinical psychology: A practical guide*. Routledge.

Onghena, P. (1992). Randomization tests for extensions and variations of ABAB single-case experimental designs: A rejoinder. *Behavioral Assessment, 14*(2), 153–171.

Onghena, P. (2020). One by one: The design and analysis of replicated randomized single-case experiments. In R. van de Schoot & M. Miočević (Eds.), *Small sample size solutions: A guide for applied researchers and practitioners* (pp. 87–101). Routledge. Recuperado de https://library.oapen.org/bitstream/handle/20.500.12657/22385/9780367221898_text%20(1).pdf

Onghena, P., Michiels, B., Jamshidi, L., Moeyaert, M., & Van den Noortgate, W. (2018). One by one: Accumulating evidence by using meta-analytical procedures for single-case experiments. *Brain Impairment*, *19*(1), 33–58. https://doi.org/10.1017/BrImp.2017.25

Plavnick, J. B., & Ferreri, S. J. (2013). Single-case experimental designs in educational research: A methodology for causal analyses in teaching and learning. *Educational Psychology Review, 25*(4), 549–569. https://doi.org/10.1007/s10648-013-9230-6

Shadish, W. R., & Sullivan, K. J. (2011). Characteristics of single-case designs used to assess intervention effects in 2008. *Behavior Research Methods, 43*(4), 971−980. https://doi.org/10.3758/s13428-011-0111-y

Slocum, T. A., Joslyn, P. R., Nichols, B., & Pinkelman, S. E. (2022). Revisiting an analysis of threats to internal validity in multiple baseline designs. *Perspectives on Behavior Science, 45*(3), 681−694. https://doi.org/10.1007/s40614-022-00351-0

Tanious, R., & Manolov, R. (2023). A practitioner's guide to conducting and analyzing embedded randomized single-case experimental designs. *Neuropsychological Rehabilitation, 33*(4), 613–645. https://doi.org/10.1080/09602011.2022.2035774

Tanious, R., & Onghena, P. (2021). A systematic review of applied single-case research published between 2016 and 2018: Study designs, randomization, data aspects, and data analysis. *Behavior Research Methods, 53*(4), 1371–1384. https://doi.org/10.3758/s13428-020-01502-4

Tate, R. L., & Perdices, M. (2019). *Single-case experimental designs for clinical research and neurorehabilitation settings: Planning, conduct, analysis, and reporting*. Routledge.

Wine, B., Freeman, T. R., & King, A. (2015). Withdrawal versus reversal: A necessary distinction?. *Behavioral Interventions, 30*(1), 87–93. https://doi.org/10.1002/bin.1399