Studies in Educational Evaluation 86 (2025) 101493

Contents lists available at ScienceDirect

Studies in Educational Evaluation

journal homepage: www.elsevier.com/locate/stueduc

Evaluating implementation fidelity in traditional vs. computerized formats of Super Skills for Life program

Teresa Galán-Luque [*,1], Mireia Orgilés [2]

Child and Adolescence Research Center, Miguel Hernández University of Elche, Alicante, Spain

ARTICLE INFO

Keywords:
Implementation fidelity
Super Skills for Life
Transdiagnostic intervention
Children
Emotional problems

ABSTRACT

This study evaluated the implementation fidelity (IF) of Super Skills for Life (SSL) program delivered in individual traditional and computerized formats for Spanish children aged 8–12 years with anxiety or depressive symptoms. Participants ($N = 97$) were randomly assigned using simple randomization to either the traditional format ($n = 51$) or the computerized format ($n = 46$). Outcomes were assessed using the Spence Children's Anxiety Scale and the Mood and Feelings Questionnaire at pretest and posttest. Results demonstrated high IF levels across both formats, with significantly higher satisfaction in the traditional format compared to the computerized format ($p = .043$, $r_b = 0.236$). Strong correlations were found between session satisfaction, perceived usefulness, and willingness to recommend the program, with therapist kindness significantly predicting reductions in anxiety ($\beta = –0.17$, $p = .039$) and depression ($\beta = –0.24$, $p = .005$) at posttest. These findings suggest that while the computerized format offers a structured and cost-effective delivery option, the traditional format provides enhanced satisfaction due to the personalized involvement of a therapist. This study underscores the importance of therapist training and fidelity monitoring in implementing SSL program, contributing to the optimization of socio-emotional interventions for children.

1. Introduction

In Spain, a third of children and adolescents feel worried, sad or unhappy, and in a study involving almost 3000 participants, they reported that this aspect had the greatest negative impact on their lives (Gasol Foundation, 2022). Although anxiousness and sadness are adaptive emotional responses (Leventhal, 2008; Rosen & Schulkin, 1998), when they become recurrent and disrupting, they can result in psychopathology. These problems are prevalent and often comorbid in childhood (Chu, 2012; Essau et al., 2018; Goldberg et al., 2009). If left unaddressed, they are associated with a chronic course (Long et al., 2018) and in Spain, they account for up to 50 % of annual mental health costs (Diego-Castaño et al., 2023). Therefore, addressing anxiety and depression early through effective interventions is crucial to ensure children lead healthy, productive, and fulfilled lives, and to reduce the economic burden on mental health systems.

Transdiagnostic interventions, which target common underlying factors of various emotional disorders, have gained prominence for their efficiency and cost-effectiveness in treating multiple problems within single protocols (Melero, Orgilés, Espada, et al., 2021). Research indicates that these interventions may be more effective than specific-disorder cognitive behavioral therapy (CBT) interventions because they address concurrent conditions and ensure that all symptoms are treated (Orgilés et al., 2023). Transdiagnostic protocols for emotional disorders aim to decrease regulation problems (Barlow et al., 2016; Melero, Orgilés, Espada, et al., 2021) and promote the adoption of more adaptive coping mechanisms (Carthy et al., 2010). However, there is a need to explore how these interventions can be implemented effectively across different formats and cultures to maximize their impact. In this context, randomized controlled trials (RCTs) are widely recognized as the gold standard for generating high-quality evidence in educational and psychological research. Leading scientific organizations, such as the U.S. National Research Council, have emphasized the importance of basing educational policies on rigorously tested interventions supported by such designs (Brignardello-Petersen, Ioannidis, Tomlinson, & Guyatt, 2015; National Research Council, 2002). Similarly, U.S. agencies like the Institute of Education Sciences and the Office of Special Education and Rehabilitative Services identify RCTs as the

* Correspondence to: Child Research Centre, Miguel Hernández University of Elche, Ave. de la Universidad, s/n, Alicante, Elche 03202, Spain.
E-mail address: tgalan@umh.es (T. Galán-Luque).
[1] https://orcid.org/0000-0001-6166-663X
[2] https://orcid.org/0000-0003-1064-9199

https://doi.org/10.1016/j.stueduc.2025.101493
Received 6 August 2024; Received in revised form 30 May 2025; Accepted 17 July 2025
Available online 23 July 2025

preferred design for evaluating well-developed programs, particularly in large-scale efficacy, replication, and scale-up studies conducted across multiple settings (Christ, 2013).

Super Skills for Life (SSL) program (Essau & Ollendick, 2013), is an evidence-based CBT transdiagnostic intervention designed to improve the mental health of young people experiencing emotional difficulties. SSL program has been shown to be effective in reducing emotional symptoms, anxiety-related interference, and peer and conduct problems in diverse populations, as demonstrated through various randomized controlled trials (e.g., Orgilés et al., 2023). Using a "train-the-trainer" approach, SSL has been implemented in 17 countries, equipping over 25,000 teachers and mental health professionals with the tools to deliver the program effectively. SSL program has positively impacted the mental health of approximately 600,000 children and adolescents worldwide. Further details about its global reach can be found on the official website: www.superskillsforlife.com. In Spain, SSL has become a primary line of research within child and adolescent applied psychology, as it is the only validated transdiagnostic protocol for preventing and treating emotional disorders in Spanish children. It has proven efficacy in group settings (Fernández-Martínez et al., 2019, 2020), individual applications (Galán-Luque, Cabello, & Orgilés, 2024a; Melero, Orgilés, Espada, & Morales, 2021), and an online self-applied format (Orgilés et al., 2023). Additionally, it has shown benefits for children diagnosed with anxiety and related conditions, with over 85 % of these achieving remission and not requiring further reinforcement visits until the 3-month follow-up (Diego-Castaño et al., 2023). Furthermore, the program has shown an impact on various factors associated with children's well-being, such as emotional self-concept, social competence, and emotion regulation strategies (Galán-Luque, Cabello, & Orgilés, 2024a).

The program was designed incorporating innovative components, such as behavioral activation (not common in anxiety-specific CBT protocols) and video-feedback with cognitive preparation, and traditional evidence-based techniques like relaxation and cognitive restructuring (Essau & Ollendick, 2013). After each session, children are assigned homework (*Supertask*) to practice the skills learned. Recently, it was adapted into a computerized format that takes advantage of technological advancements and was applied at home with parents as cotherapists (Orgilés et al., 2023). Despite its promising results on symptom outcomes, lower attrition rates and the need for additional support were noted due to the remote nature of the program (Orgilés et al., 2023). To address these limitations, the program was further evaluated in a clinical setting under the supervision of a trained psychotherapist, comparing it to the traditional individual version through a randomized comparative effectiveness trial (Galán-Luque, Cabello, & Orgilés, 2024a).

An important area of research on the SSL program has focused on the impact of implementation fidelity (IF) on its effectiveness (Fernández-Martínez, Orgilés, Espada, Essau, & Morales, 2021; Melero, Orgilés, Espada, & Morales, 2021). IF is crucial for ensuring the integrity of evidence-based practice, as its systematic evaluation enhances study validity, provides a good estimate of implementation quality, and helps prevent Type-III errors (Breitenstein et al., 2010; Carroll et al., 2007). Generally, higher fidelity is associated with better outcomes (Oosthuizen & Louw, 2013). However, some authors argue that the impact of IF on intervention results is often overlooked and rarely reported (Nelson et al., 2020). Furthermore, the scarcity of studies in Spain evaluating IF in programs addressing childhood emotional problems underscores the need for further investigation (Melero, Orgilés, Espada, et al., 2021).

Several models have been proposed to conceptualize and measure IF. According to Dusenbury (2003), fidelity encompasses five dimensions: *adherence* (referring to how closely the implementation follows the prescribed protocol), *dose* (the amount of the program delivered in terms of number and duration of sessions), *quality of delivery* (skills and competence of the facilitator), *differentiation* (core components are uniquely identifiable), and *responsiveness* (participant's engagement and satisfaction). The first four dimensions are related to the implementer and the protocol's execution, while *responsiveness* pertains to the participants' reactions (Berkel et al., 2011). Research have highlighted the interrelated nature of the five components and suggested that not all need to be present for effective implementation (Gázquez Pertusa et al., 2011). In the review by Dusenbury and colleagues (2003), illustrative examples are specifically provided regarding how each dimension might be measured.

1.1. Evaluating IF in the context of Super Skills for Life program research

IF plays a critical role in determining the effectiveness of intervention programs such as SSL. Surprisingly, despite the global dissemination of SSL through the 'train-the-trainer' approach, fidelity has not yet been systematically measured outside of Spain. To date, only two Spanish studies have examined the impact of IF on SSL outcomes, both highlighting the significant influence of high fidelity on the effectiveness of the intervention. To better contextualize the development of SSL research and its evaluation across different settings and formats, a visual timeline is presented in Fig. 1.

Melero, Orgilés, Fernández-Martínez, et al. (2021) conducted a study with 119 schoolchildren aged 8–12 years who exhibited internalizing problems. Their study aimed to evaluate: (1) the Spanish SSL group program's IF level; (2) the relationship between IF dimensions; and (3) outcomes at post-test and follow-up depending on IF. Participants were categorized into high-fidelity (HFG) and low-fidelity groups (LFG) based on the fidelity level of program delivery. At post-test and 12-month follow-up, the HFG showed significantly better outcomes than the LFG, including lower levels of depression, overall behavioral and emotional difficulties, and higher family self-concept. The study concluded that while the SSL program is effective even with low implementation fidelity, it is more effective when delivered with high fidelity. Similarly, Fernández-Martínez et al. (2021) examined the short- and long-term effectiveness of the SSL program depending on implementation fidelity in a sample of 123 Spanish-speaking children aged 6–8 years with internalizing symptoms. Participants were divided into an HFG, an LFG, and a control group (CG). The HFG achieved better outcomes than the LFG at both short and long term, including reductions in internalizing and externalizing problems. These findings highlight the importance of implementing the SSL program with high fidelity to achieve optimal effectiveness, aligning with existing research on prevention programs that demonstrates that higher IF is associated with greater reductions in symptoms and improved skill acquisition (Botvin & Griffin, 2004; Botvin, Griffin, Botvin, Murphy, & Acevedo, 2018; Durlak & DuPre, 2008; Dusenbury, 2003)

Despite these promising findings, there are limitations in the existing studies. First, no studies have explored IF in the individual format, which is increasingly relevant for personalized interventions. Second, no studies have tracked changes in fidelity measures over the sessions, limiting understanding of how IF evolves during the intervention. Third, previous studies have categorized children into low- and high-IF groups by combining all dimensions into a single categorical measure. This approach may not have allowed for a detailed understanding of which specific IF components have the most significant impact on the outcomes. Finally, both studies were conducted in the school context, whereas the present study focuses on the clinical context, offering new insights into how IF operates in a different setting.

1.2. Novel contributions of the current study

This study addresses several key limitations of previous research and makes novel contributions to the literature on IF in intervention programs. Specifically, it provides a comprehensive analysis of IF in two individual formats of the SSL program: the traditional face-to-face format and a novel computerized intervention. This dual-format comparison is the first of its kind for the SSL program, expanding its potential

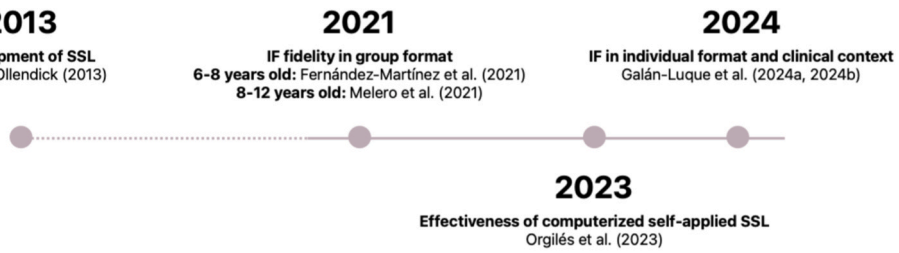

Fig. 1. Timeline of Key Milestones in the Development, Adaptation, and Evaluation of Super Skills For Life Program.

applicability across diverse delivery settings while applying a randomized allocation of participants to enhance the credibility and causal interpretability of the findings.

Unlike prior studies, which evaluated IF in group formats (Fernández-Martínez, Orgilés, Espada, Essau, & Morales, 2021; Melero, Orgilés, Espada, & Morales, 2021), this study focuses exclusively on individual formats, offering insights into the role of IF in one-on-one interventions. Additionally, while earlier research grouped children into high- and low-fidelity categories based on combined dimensions, this study evaluates individual IF dimensions separately to provide a nuanced understanding of their specific impacts on outcomes.

Several methodological innovations further distinguish this study. First, all sessions were attended by children (*dose*), and therapists adhered strictly to the protocol (*adherence*), ensuring consistent delivery. While this consistency limited variability in these measures, it allowed for an in-depth focus on other critical IF dimensions. For the first time, the study evaluated the *quality of delivery* through therapist kindness, following recommendations from previous research (Hansen et al., 1991). Therapist kindness was selected as an indicator of *quality of delivery* because facilitator warmth, friendliness, and a supportive attitude can be identified as components of facilitator competence. These characteristics contribute directly to higher-quality interactions between therapist and participants, fostering increased participant engagement, motivation, and responsiveness, which ultimately enhance the effectiveness of interventions (Hansen et al., 1991). In this study, therapist kindness was operationalized through child-reported ratings collected after program completion, assessing the degree of kindness the children perceived from their therapist. Additionally, it assessed several measures of *responsiveness*, including self-report variables (session satisfaction and perceived usefulness, and the degree of recommendation to other children) and an objective measure: the number of home-tasks completed.

The objectives of the study were fourfold:

(1) to evaluate the level of IF of the Spanish version of the individual SSL program and compare it between formats.
(2) to track how IF measures change over the sessions and compare these changes between formats.
(3) to analyze the relationships between IF measures.
(4) to examine the impact of fidelity variables on intervention's outcomes.

2. Materials and methods

2.1. Study design

The sample comes from a randomized trial (clinicaltrial.gov ID: NCT05574491) aimed at comparing the effectiveness of traditional and computerized individual SSL versions on internalizing symptoms in children between 8 and 12 years of age from the southeast region of Spain (see Galán-Luque, Cabello, & Orgilés, 2024a,b for more details). This specific age range was chosen for several reasons. First, children in this age group are at a critical developmental stage for socio-emotional skill acquisition, making early interventions particularly effective in preventing the progression of emotional disorders (Ecclestone & Hayes, 2009). Second, previous research with SSL program has consistently separated participants into age groups of 6–8 and 8–12 years to account for developmental differences and has found that children aged 8–12 benefit significantly from the intervention (Galán-Luque, Cabello, & Orgilés, 2024a; Essau and Ollendick, 2013). Thus, focusing on this age group allowed for consistency with prior studies and facilitated comparison of results across different implementations of the program (Fernández-Martínez, Orgilés, Espada, Essau, & Morales, 2021; Melero, Orgilés, Espada, & Morales, 2021).

Participants were selected based on the presence of emotional symptoms (anxiety and/or depression), identified by parents using standardized screening tools. Inclusion criteria were: (a) aged 8–12 years, (b) exhibiting emotional symptoms as reported by parents, (c) not currently receiving pharmacological or psychological treatment for emotional/behavioral problems, and (d) no diagnosis of a neurodevelopmental disorder. Fig. 2 shows the flow of participants in the study. Of the initial 200 parents assessed, 109 met the inclusion criteria and were randomly assigned to two groups: traditional (n = 54) and computerized format (n = 55). Four children discontinued the computerized intervention. Of the remaining 105, 97 children (53.61 % girls) completed the fidelity measures at post-test, constituting the final sample for this study.

2.2. Measures

2.2.1. Sociodemographic variables

Parents provided sociodemographic information through the initial online questionnaire, including items about child's age, sex, number of siblings, parents' marital status, age, educational level, employment status, employment sector, and monthly income.

2.2.2. Parental screening

To assess the presence of emotional problems (anxiety and/or depressive symptoms), parents completed the following screening tools:

Spence Children's Anxiety Scale – Parent Version (SCAS-P; Spence, 1998*)*. It consists of 38 items rated on a 4-point Likert scale (0 = never, 1 = sometimes, 2 = often, 3 = always). The total score ranges from 0 to 114, with higher scores indicating more severe symptoms. For this study, children had to score ≥ 25, indicating elevated anxiety levels (Spence, 2021). The Spanish version of SCAS-P has shown strong reliability ($\alpha = 0.91$) and satisfactory validity (Orgilés et al., 2019). In our sample, the ordinal α was 0.91.

Mood and Feelings Questionnaire – Parent Version (MFQ-P; Angold et al., 1995). It includes 34 items rated on a 3-point Likert scale (0 = not true, 1 = somewhat true, 2 = true) to assess depressive symptoms in children. The total score ranges from 0 to 68, with higher scores indicating more severe symptoms. Children needed a score of ≥ 20 to meet the inclusion criteria, reflecting subclinical levels of depression (Daviss et al., 2006; Kent, Vostanis, & Feehan, 1997; Wood, Kroll, Moore, & Harrington, 1995). Previous research has shown good psychometric properties of MFQ-P (Fernández-Martínez et al., 2020). In our sample, the ordinal α was 0.93.

CONSORT
TRANSPARENT REPORTING of TRIALS

CONSORT 2010 Flow Diagram

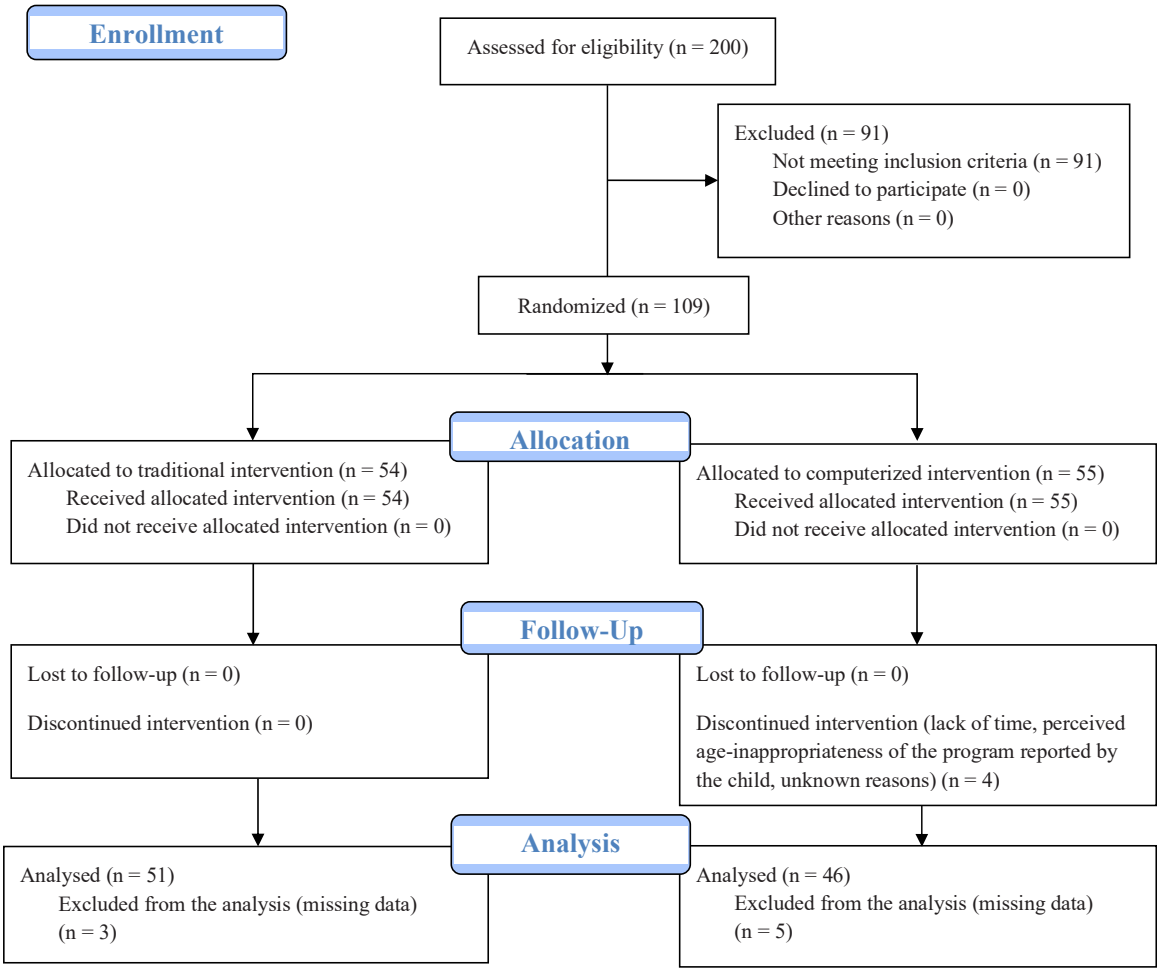

Fig. 2. Flow Diagram of Children Participating in this Study.

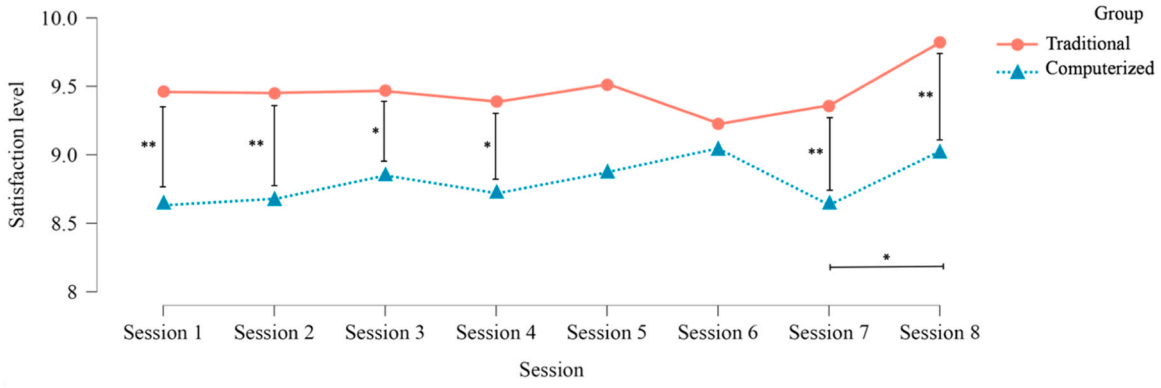

Fig. 3. *Session Satisfaction Levels Across the Sessions by Group. Note.* *p < .05, **p < .01, ***p < .001.

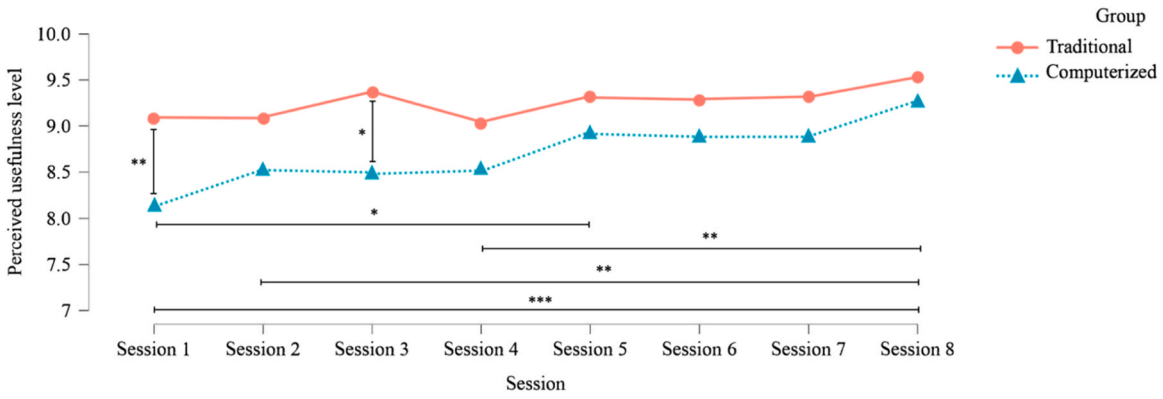

Fig. 4. *Session Perceived Usefulness Levels Across the Sessions by Group. Note.* *p < .05, **p < .01, ***p < .001.

2.2.3. Children's primary outcome measures

Spence Children's Anxiety Scale – Child Version (SCAS-C; Spence, 1998*).* It consists of 38 items that children rate on a 4-point Likert scale (0 = never, 1 = sometimes, 2 = often, 3 = always). The total score ranges from 0 to 114, with higher scores indicating more severe anxiety symptoms. The Spanish version of the SCAS-C was used, which has shown good psychometric properties (Orgilés et al., 2019). In our sample, the ordinal α was 0.91. Children completed the SCAS-C at pretest and posttest.

Short Mood and Feelings Questionnaire – Child Version (SMFQ-C; Angold et al., 1995*).* It includes 13 items that children rate on a 3-point Likert scale (0 = not true, 1 = sometimes true, 2 = true). The total score ranges from 0 to 26, with higher scores indicating more severe depressive symptoms. The SMFQ-C has shown good reliability and validity in various populations (e.g., Daviss et al., 2006). In this study, the Spanish version of the SMFQ-C was used, which has been validated among Spanish children (Espada, Gonzálvez, Fernández-Martínez, Orgilés, & Morales, 2022). In our sample, the ordinal α was 0.92. Children completed the SMFQ-C at pretest and posttest.

2.2.4. Implementation fidelity

IF was assessed through measures specifically developed for this study, consistent with the dimensions proposed by Dusenbury et al. (2003) and previous literature on intervention fidelity (Durlak & DuPre, 2008; Hansen, Graham, Wolkenstein, & Rohrbach, 1991). Given that the SSL program was delivered individually, we developed tailored self-report and observational measures suitable for this study's objectives.

Responsiveness was assessed through multiple self-report and observational measures:

Session satisfaction and perceived usefulness. After each session, children rated their satisfaction with the session (*"How much did you like this session?"*) and the perceived usefulness of the session (*"How useful do you think this session was for…?"*), using a Likert scale from 1 (not at all) to 10 (very much). The items evaluating usefulness varied slightly depending on session content (e.g., relaxation skills, cognitive restructuring, improving self-esteem).

Recommendation to others. At post-test, children rated their likelihood of recommending the program to peers by answering the question *"How much would you recommend the program to other children?"*, rated from 1 (not at all) to 10 (very much).

Number of home tasks completed. Therapists tracked the number of homework assignments completed by each child after each session, providing an objective measure of children's responsiveness to the intervention.

Quality of delivery was evaluated through children's perceptions of therapist kindness. At post-test, children rated the therapist's kindness on a Likert scale ranging from 1 (not at all kind) to 10 (very kind) by answering the question *"How kind did you find your therapist?"*.

2.3. Characteristics of the sample

Table 1 shows the sociodemographic characteristics according to intervention condition. Children had a mean age of 9.57 years (SD = 1.44) and a mean of 1.01 siblings (SD = 0.82). All children were Spanish-speaking. Most parents were aged between 35 and 54 years (92.78 %), were married (78.35 %), had at least secondary education (91.75 %), and were actively working (84.53 %). Most families had a monthly income of 1000–1999 euros (29.90 %), 2000–2999 euros (27.84 %), or 3000–4999 euros (21.65 %).

2.4. Baseline equivalence and symptom levels

There were no significant differences at pretest between groups regarding sociodemographic variables or primary outcome measures of anxiety and depression (all p values >.05; see Table 1). Specifically, children's self-reported anxiety scores, measured using the SCAS-C (Spence, 1998), indicated moderate levels of anxiety (mean scores around 34), consistent with previous studies conducted with community and at-risk samples (Essau et al., 2011). Depressive symptoms, measured through the SMFQ-C (Angold et al., 1995), reflected mild to moderate levels of depressive symptomatology (mean scores around 8), comparable to normative data reported in similar age groups. Thus, the baseline measures confirm that both groups had comparable emotional difficulties at the start of the intervention.

2.5. Assessment of attrition bias at postest

Internal validity analyses found no significant differences (p = 0.071) in retention between the traditional (5.56 %; n = 3) and computerized (16.36 %; n = 9) groups. External validity at postest showed no significant differences between children who dropped out at postest and those who completed the study in sociodemographic and outcome variables: children's age (p = 0.980), number of siblings (p = 917), sex (p = 0.391), parents' marital status (p = 420), age (p = .911), educational level (p = 272), employment status (p = 0.460), employment sector (p = 268), monthly income (p = 617), anxiety (SCAS total score) (p = 0.969), and depression (SMFQ total score) (p = 0.988).

2.6. Procedure

Information about the study was disseminated through multiple channels, including schools, social media, and personal contacts of the research team. Parents could voluntarily participate by completing an online form to assess inclusion criteria. For eligible children, contact was made via phone to provide additional information, address queries, and obtain consent. After obtaining consent, children were assigned a numerical ID, and a simple randomization procedure was applied using

Table 1
Sociodemographic Characteristics and Main Outcomes of Participants by Group.

	Traditional (n = 51)		Computerized (n = 46)		Total (N = 97)			
	n	%	*n*	%	*n*	%	χ^2	*p*-Value
Sociodemographics								
Gender								
Male	23	45.10	22	47.83	45	46.39	.072	.788
Female	28	54.90	24	52.17	52	53.61		
Parental marital status								
Married/partnered	38	74.51	38	82.61	76	78.35	1.582	.664
Separated/divorced	8	15.69	5	10.87	13	13.40		
Single	4	7.84	3	6.52	7	7.22		
Widowed	1	1.96	0	0.00	1	1.03		
Parent age								
25–34 years	3	5.88	2	4.35	5	5.16	2.795	.424
35–44 years	25	49.02	28	60.87	53	54.64		
45–54 years	21	41.18	16	34.78	37	38.14		
55–65 years	2	3.92	0	0.00	2	2.06		
Educational level								
Primary education	4	7.84	4	8.70	8	8.25	2.371	.499
Secondary education	23	45.10	14	30.44	37	38.14		
Higher education	19	37.26	21	45.65	40	41.24		
Masters' or PhD	6	9.80	7	15.22	12	12.37		
Employment status								
Full time	31	60.78	26	56.52	57	58.76	7.876	.163
Part-time	2	3.92	9	19.57	11	11.34		
Self-employed	7	13.73	6	13.04	13	13.40		
Student	1	1.96	0	0.00	1	1.03		
Public employee	1	1.96	0	0.00	1	1.03		
Unemployed	9	17.65	5	10.87	14	14.43		
Employment sector								
Education/professional services	13	25.49	13	28.26	26	26.80	5.269	.384
Financial/administrative/real estate services	14	27.45	12	26.09	26	26.80		
Health/social services	6	11.77	12	26.09	18	18.56		
Trade/transport/storage	8	15.69	5	10.87	13	13.40		
Manufacturing/energy/building	4	7.84	2	4.35	6	6.19		
Information/communication/hospitality	6	11.77	2	4.35	8	8.25		
Monthly income								
Up to 499 euros	1	1.96	0	0.00	1	1.03	5.083	.533
500–999 euros	6	11.77	2	4.35	8	8.25		
1000–1999 euros	12	23.53	9	19.57	21	21.65		
2000–2999 euros	12	23.53	17	36.96	29	29.90		
3000–4999 euros	13	25.49	14	30.44	27	27.84		
5000 euros or more	1	1.96	1	2.17	2	2.06		
Prefer not to answer	6	11.77	3	6.52	9	9.28		
Dependent variables (children reported)	*M*	*SD*	*M*	*SD*	*M*	*SD*	*U*-test	*p*-Value
Anxiety (SCAS-C total score)	35.28	15.41	33.78	13.62	34.57	14.53	1222.500	0.723
Depression (SMFQ-C total score)	8.28	5.91	8.24	4.98	8.26	5.46	1137.500	0.797

Note. SCAS-C = Spence Children's Anxiety Scale – Child version; SMFQ-C = Short Mood and Feelings Questionnaire – Child version. Higher scores indicate higher anxiety or depressive symptoms.

Microsoft Excel's = RAND() function to generate a random number for each participant. The resulting list was then sorted in ascending order, and participants were alternately allocated to either the traditional or computerized intervention condition. This method ensured random assignment without prior knowledge of the final group allocation. After randomization, we verified that the groups were homogeneous in key characteristics such as age, gender, and baseline symptom severity. Recruitment, randomization, and intervention delivery were all conducted by the same team of researchers and trained therapists. While this ensured continuity and consistency in implementation, it also represents a potential source of bias, as allocation concealment was not fully independent. This limitation is acknowledged in the discussion section. Nonetheless, adherence to the randomization protocol and training procedures helped reduce the potential impact of this issue on internal validity.

The intervention was administered in a clinical setting at the authors' institution. Children participated in two sessions each week, completing all eight sessions over a four-week period. Each child participated in all eight sessions. The program facilitators were selected from the author's institution. All facilitators received comprehensive training prior to program delivery and attended weekly supervision meetings, which emphasized strict adherence to the SSL protocol and ensured consistency across sessions. After each session, children were assigned the *Supertask*, were provided with a summary of the session content (*Remember!* handout) and received a stamp as a reward for their effort, which was added to the *Good job!* handout. Parents were given printed materials detailing the session objectives and practical advice for reinforcing the learned concepts at home.

2.7. Interventions

SSL program aims to equip children with skills in emotional regulation, cognitive restructuring, behavioral activation, social skills, relaxation, and social problem-solving. It comprises eight structured sessions lasting between 45 and 60 min, depending on the format, and is facilitated by trained therapists using an intervention manual (Orgilés et al., 2023). Both modalities consisted of individual sessions sharing the same content and objectives but differing in methodology (Galán-Luque, Cabello, & Orgilés, 2024a).

2.7.1. Traditional format

The intervention was based on the face-to-face individual SSL version described by Melero, Orgilés, Fernández-Martínez, et al. (2021), but in the current study, video content was included to support the program activities. Each session lasts 60 min and incorporates a variety of activities, including writing exercises, games, readings, and role-playings with the therapist. Participants also use a workbook containing explanations and activities.

2.7.2. Computerized format

The computerized modality uses the platform from the self-applied SSL (Orgilés et al., 2023) but is administered in a clinical setting under therapist supervision. Each session lasts 45 min, with children independently completing all activities guided by the digital platform. The therapist is present throughout the session to offer support and clarification if needed, ensuring participant engagement, adherence, and fidelity to the intervention protocol.

2.8. Statistical analyses

All statistical analyses were conducted using RStudio Version 1.1.453 (R Core Team, 2018) and JASP version 0.18.3 software (JASP Team, 2024). Only participants who completed all measures were included in the analyses. Equivalence in sociodemographic and main outcomes between groups and attrition analyses were conducted. Chi-square and Mann-Whitney *U*-tests were used. The internal consistency of the outcome measures was calculated using the ordinal alpha coefficient.

For Aim 1, descriptive analyses and Mann-Whitney *U*-tests (two-tailed) were used, and the matched rank biserial correlation (r_b) for the effect size. Interpretation criteria were as follows: 0.10–0.29 small, 0.30–0.49 medium, and $\geq$ 0.50 large (López-Martín & Ardura, 2023). For Aim 2, mixed-factor ANOVAs were performed to assess session satisfaction and perceived usefulness measures, with session number as a within-subjects factor and intervention format as a between-subjects factor. Partial eta squared (η_p^{22}) was reported for the effect size, with the following interpretation criteria: < 0.01 very small, 0.01–0.05 small, 0.06–0.13 moderate, and $\geq$ 0.14 large (López-Martín & Ardura, 2023). A mixed–effects logistic regression model was conducted to analyze the pattern of *Supertask* completion (completed vs. not-completed) over sessions and differences between intervention groups.

For Aim 3, Spearman's correlations were calculated for both intervention groups. For Aim 4, independent multiple linear regression analyses for each outcome were performed. Differences in the total postest scores of the main outcome measures (SCAS-C and SMFQ-C total score) using Spearman correlations with fidelity measures were calculated. The only variable that correlated significantly with anxiety and depression at postest ($p < .05$) (therapist kindness) was included in the models. Models were performed to determine how its impact on symptoms at postest, controlling for pretest scores, group and demographic measures (age and sex). Following the instructions of Jaccard and Turrisi (2003), all continuous predictor variables were centered.

Post-hoc statistical power analyses were conducted using G*Power 3.1 software (Faul et al., 2007, 2009) for the main analyses. For the between-group comparison of satisfaction levels using the Mann–Whitney *U*-test, the matched rank biserial correlation (r_b = 0.236) was converted into Cohen's *d* to estimate statistical power. For the mixed ANOVA and multiple regression models, effect sizes were transformed into *f* and f^2, respectively.

3. Results

3.1. Aim 1: IF levels and comparison between groups

Table 2 shows mean levels of fidelity measures and differences between the two intervention groups. The results showed that there were statistically significant differences in mean satisfaction levels in the traditional group, with a small effect size.

3.2. Aim 2: Track IF measures change over time

3.2.1. Session satisfaction

Using the Huynh–Feldt correction, there was a significant main effect of session number ($F(5.931, 563.477) = 2.266$, $p = .037$, $\eta_p^2 = .023$) and type of intervention ($F(1, 95) = 8.852$, $p = .004$, $\eta_p^2 = .085$) on satisfaction levels. However, the interaction between session satisfaction and type of intervention was not statistically significant, $F(5.931, 563.477) = 1.210$, $p = .244$. Bonferroni–corrected post hoc tests showed ~~a~~an significant increase in satisfaction levels from Session 7 to Session 8 ($p = 0.02$). Additionally, there were significant differences in satisfaction levels between groups in Session 1 ($p = 0.001$), Session 2 ($p = 0.005$), Session 3 ($p = 0.016$), Session 4 ($p = 0.037$), Session 7 ($p = 0.007$), and Session 8 ($p = 0.003$).

3.2.2. Session perceived usefulness

Using the Huynh–Feldt correction, there was a significant main effect of session number ($F(5.940, 564.312) = 4.888$, $p < .001$, $\eta_p^2 = .049$) and type of intervention ($F(1, 95) = 4.762$, $p = .032$, $\eta_p^2 = .048$) on perceived usefulness levels. However, the interaction between session satisfaction and type of intervention was not statistically significant, $F(75.940, 564.312) = 1.665$, $p = .322$. Bonferroni–corrected post hoc tests showed an increase in perceived usefulness levels between Session 1 and 5 ($p = 0.034$), Session 1 and 8 ($p < 0.001$), Session 2 and 8 ($p = 0.006$), and Session 4 and 8 ($p = 0.003$). Additionally, there were differences in perceived usefulness levels between groups in Session 1 ($p = 0.005$) and Session 3 ($p = 0.019$).

3.2.3. Pattern of change in Supertasks

Results indicated a significant main effect of session number on *Supertask* completion ($\chi^2 = 29.578$, p < .001), suggesting that the likelihood of completing the *Supertask* varied significantly over time. Completion rates were significantly lower in Sessions 4, 5, 6, and 7 compared to earlier sessions ($p < .001$). However, there were no significant differences between groups and the interaction between session number and type of intervention was not significant ($p > 0.05$). Model fit statistics indicated a reasonable fit to the data, with a deviance of 95.771, log likelihood of –84.672, AIC of 199.345, and BIC of 267.154. Fig. 5 shows consistently high completion rates for most sessions, with slight drops observed in Sessions 5 and 7.

Table 2
Mann–Whitney U–Tests Examining Differences Between Groups in Fidelity Measures.

Variable	Traditional ($n = 51$)		Computerized ($n = 46$)		*U–Test*	*p*	Rank–biserial correlation
	M	*SD*	*M*	*SD*			
Number of *Supertasks*	6.745	0.796	6.696	0.591	1299.500	0.169	–
Satisfaction	9.459	0.646	8.804	1.418	1449.500	0.043	0.236
Perceived usefulness	9.246	0.893	8.698	1.531	1281.500	0.432	–
Recommendation to others	8.941	2.258	8.609	2.124	1332.000	0.189	–
Therapist kindness	9.902	0.361	9.630	1.199	1263.500	0.237	–

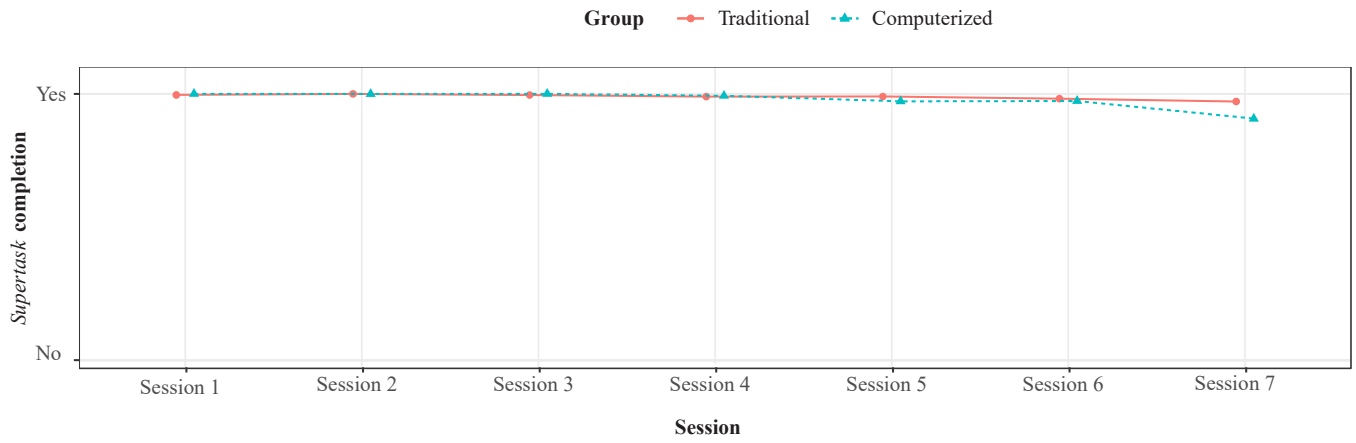

Fig. 5. *Supertask Completion Rates Across Sessions by Group. Note.* Each dot represents an individual children's *Supertask* completion for a given session. Completion was coded as Yes or No.

3.3. Aim 3: Analyze the relationships between fidelity dimensions

3.3.1. Traditional group

There was a direct, strong, and statistically significant relationship between the mean level of session satisfaction and perceived usefulness, mean level of session satisfaction and recommendation to others, and perceived usefulness and recommendation to others (Table 3). Additionally, there was a direct, moderate, and statistically significant relationship between the recommendation to others and therapist kindness. The number of *Supertasks* did not correlate with any of the other variables, and therapist kindness did not correlate with mean session satisfaction and perceived usefulness levels.

3.3.2. Computerized group

There was a direct, strong, and statistically significant relationship between the mean level of session satisfaction and perceived usefulness and recommendation to others, and a direct, moderate, and statistically significant relationship between the recommendation to others and mean perceived usefulness and therapist kindness (Table 3). The number of *Supertasks* did not correlate with any of the other variables, and therapist kindness did not correlate with mean session satisfaction and perceived usefulness levels.

3.4. Aim 4: Impact of fidelity dimensions on symptoms change

Only therapist kindness correlated significantly with anxiety (ρ = –0.239; $p < .018$) and depression (ρ = –0.314; $p < .002$) outcomes at postest. Multiple linear regression models for predicting anxiety and depressive symptoms at postest were significant and provided explained variance rates of 43.5 % [$F(5; 96) = 14.011$; $p < 0.001$] and 42.8 % [$F(5; 96) = 13.642$; $p < 0.001$], respectively. In both models, therapist kindness reported by children was related to lower anxiety (β = –0.170, t = –2.099, p = 0.039) and depressive scores (β = –0.237, t = –2.905, p = 0.005) after receiving the program, regardless of the type of intervention.

3.5. Power analysis

Post-hoc power analyses confirmed that the study had adequate statistical power to detect the observed effects in the main analyses. Specifically, the estimated power was 0.62 for the between-group comparison of satisfaction levels, 0.98 for the main effect of intervention type in the mixed ANOVA on satisfaction, and 0.97 for the main effect of session number. For the mixed ANOVA on perceived usefulness, the power was 0.83 for the group effect and 0.99 for the effect of session number. All correlational analyses yielded power estimates above 0.98, and the multiple linear regression models predicting post-intervention anxiety and depression scores showed power levels greater than 0.99.

4. Discussion

This study sought to enhance the understanding of the role of IF in the effectiveness of the SSL program for the first time in its individual version for Spanish children aged 8–12 years with emotional problems and provide insights for optimizing its implementation. Its objectives included evaluating the level of IF measures and comparing it between modalities, tracking changes over sessions, analyzing their relationships, and examining their impact on changes in symptomatology. The consistency and strength of the results were reinforced by statistical power analyses, supporting the reliability of the main findings.

Overall, results indicated that the IF measures related to *responsiveness* and *quality of delivery* were high across both formats. Children completed an average of over 6.5 *Supertasks* (range 0–7) in both versions, showing high compliance. Satisfaction levels were above 8.80 (on a 0–10 scale) in both groups, with the traditional format achieving significantly higher satisfaction, suggesting that face-to-face interaction may enhance engagement. Interestingly, despite the assumption that today's generation might prefer digital interactions over face-to-face

Table 3
Correlations of Fidelity Dimensions.

	Traditional					Computerized				
Variable	1	2	3	4	5	1	2	3	4	5
1. Number of *Supertasks*	–					–				
2. Satisfaction level	−0.123	–				−0.149	–			
3. Perceived usefulness	−0.096	0.808***	–			−0.175	0.750***	–		
4. Recommendation to others	−0.086	0.624***	0.511***	–		0.009	0.523***	0.374*	–	
5. Therapist kindness	−0.089	0.103	0.099	0.312*	–	0.026	0.038	−0.006	0.483***	–

*p < .05,**p < .01,***p < .001

relationships, the traditional program achieved higher satisfaction levels. The full involvement of the therapist in the traditional intervention, compared to the computerized intervention that relies solely on multimedia content, appears to create a more personalized and fulfilling experience by prioritizing the child's needs and interests. These findings align with prior research highlighting that a child-centered therapeutic relationship fosters trust in the therapist and contributes to a more positive therapeutic experience (Nuñez et al., 2022). However, there were no significant differences in perceived usefulness, recommendation to others, and therapist kindness, with high mean scores (above 8.60) suggesting that both formats were well-received.

Although their studies focused on group formats, our findings are consistent with those of Melero, Orgilés, Fernández-Martínez, et al. (2021) and Fernández-Martínez, Orgilés, Espada, Essau, & Morales (2021), who also reported high IF levels for the group version of SSL. This suggests that, regardless of the delivery format (group or individual), SSL can be implemented with high fidelity when appropriate support and monitoring strategies are in place. While our study did not aim to compare group and individual formats directly, this parallel reinforces the program's adaptability across settings. Supporting this, a recent meta-analysis that compared individual versus group CBT for anxiety in children and adolescents found no clear preference for either format among younger children (Guo et al., 2021), suggesting comparable effectiveness. However, the same meta-analysis noted that ~~found~~ adolescents tended to favor individual therapy (Guo et al., 2021). This aligns with studies involving adults with emotional disorders, who often prefer individual therapy due to privacy concerns and difficulties with self-disclosure in group settings (Osma et al., 2019). Such findings suggest a developmental trend: as individuals mature, their preference may shift toward individual formats, likely reflecting increased needs for privacy and emotional safety (Shirk & Karver, 2003). A practical implication for future SSL research is to explore whether adolescents show a stronger preference for individual over group delivery and whether this impacts the intervention's efficacy.

A key difference in our study is that we ensured all children attended all eight sessions, unlike Melero, Orgilés, Fernández-Martínez, et al. (2021) (7.26 sessions on average) and Fernández-Martínez, Orgilés, Espada, Essau, & Morales, (2021) (17.9 % attended less than seven sessions). This was made possible by the individual format's scheduling flexibility: if a child missed a session due to illness or any other reason, it was rescheduled—an option not feasible in group formats. Strategies such as shortening sessions, splitting them into multiple shorter sessions, and scheduling them during school hours to facilitate family access have been found helpful for improving attendance (Watt & Dadds, 2007). These adaptations have previously been used in anxiety treatment programs for children in school contexts, such as Coping Cat (Mychailyszyn et al., 2011), and are considered reasonable modifications (Kendall et al., 2023). Additionally, children in our study completed slightly more *Supertasks* than in group studies and the fully self-applied intervention (Orgilés et al., 2023). This is particularly relevant given that previous research has noted that digital mental health interventions, while effective in controlled trials, often suffer from poor engagement and high attrition rates (Bakker, Kazantzis, Rickwood, & Rickard, 2016; Lattie et al., 2019; Stiles-Shields, Montague, Lattie, Kwasny, & Mohr, 2017).

The second aim was to track how IF measures changed over the sessions. Satisfaction and perceived usefulness increased in later sessions, which could reflect participants becoming more comfortable and familiar with the program content and structure. The traditional format consistently showed higher satisfaction levels than the computerized format, suggesting the importance of therapist-led sessions for engagement. Only Sessions 5 and 6 showed no differences in satisfaction. Session 5's relaxation techniques were similar across formats, while Session 6's social abilities content was less liked in the traditional format, possibly due to role-plays, which may have induced social anxiety. Analysis of *Supertask* completion revealed overall high rates, indicating that the reinforcement strategy with stamps on the *Good Job!* sheet is effective. However, slight drops occurred in Sessions 5 and 7. Session 5's task required practicing a relaxation technique three times, demanding effort and time. Session 7's task involved filling out a problem resolution sheet, requiring higher cognitive involvement because children usually found it challenging to anticipate the consequences of possible solutions. These findings are particularly promising, as between-session homework is often underutilized in psychological interventions despite being a critical component of evidence-based mental health care (Kazantzis & Deane, 1999; Kazantzis, Deane, & Ronan, 2005). Homework completion has been shown to facilitate skill generalization, strengthen learning, and improve clinical outcomes (Hudson & Kendall, 2002; Kazantzis et al., 2016). However, adherence rates in youth interventions are typically low, ranging from 39 % to 63 % (Berkovits, O'Brien, Carter, & Eyberg, 2010; Danko, Brown, Van Schoick, & Budd, 2016; Simons et al., 2012), with barriers such as unclear instructions, insufficient family support, and inadequate reinforcement contributing to these challenges (Bunnell et al., 2021). The consistently high task completion rates observed in our study underscore the effectiveness of strategies like personalized reinforcement tools, such as the *Good Job!* sheet with stamps, while also highlighting the importance of parental involvement and clear instructions for tasks. Providing parents with handouts that detail homework assignments and emphasizing their importance to children could further enhance adherence. Nonetheless, the reduced completion rates observed in Sessions 5 and 7 suggest that these tasks may require additional reinforcement or modifications. Future research on the SSL program should explore ways to increase engagement, such as enhancing rewards for these tasks, simplifying their structure, or incorporating greater family support to address environmental barriers. These findings offer practical implications for improving the implementation of socio-emotional interventions for children, particularly those that rely on between-session homework.

The study's third aim was to explore the relationships between IF measures. Significant positive correlations were observed between session satisfaction, perceived usefulness, and willingness to recommend the program. However, the association between recommendation to other children and perceived usefulness was stronger in the traditional group, suggesting that children in this format might have placed more value on the perceived direct benefits of the sessions when making recommendations. In contrast, in the computerized format, recommendations correlated more strongly with therapist kindness, highlighting its importance even when therapists played a more secondary role. These findings corroborate the interrelated nature of the five dimensions of implementation fidelity proposed by Gázquez Pertusa et al. (2011). Their work suggests that while the dimensions are often linked, not all must be simultaneously high for effective implementation. Nonetheless, as in previous group IF studies with SSL, all measures of IF in this study were consistently high, making it challenging to draw more nuanced conclusions about how individual components uniquely affect outcomes. Moreover, to our knowledge, no international studies evaluating SSL have specifically explored relationships between these fidelity components, limiting our ability to compare results across different contexts. This suggests that future research could benefit from deliberately incorporating variability into fidelity measures by studying diverse implementation contexts or designing interventions that test specific fidelity components individually. Such approaches may help clarify which elements of fidelity exert the most significant influence on program success. Interestingly, *Supertasks* completion did not correlate with other IF measures, indicating that task completion may be influenced by factors other than those affecting session satisfaction and perceived usefulness. The result regarding the *Supertask* completion contrasts with previous SSL studies (Melero, Orgilés, Espada, et al., 2021), which noted modest correlations between *Supertask* completion and program acceptance. Further research is warranted to understand the factors influencing *Supertask* completion.

The fourth aim was to examine how IF influences symptom changes.

Results indicated that therapist kindness was significantly associated with reductions in anxiety and depressive symptoms, underscoring the importance of a supportive therapist-patient relationship in therapy. This finding echo research that highlights the therapeutic relationship as critical to effective outcomes (Norcross & Wampold, 2011). The absence of significant associations for other IF measures might be attributed to their uniformly high levels across our sample, limiting the variability necessary to detect their effects. These results suggest that the quality of interaction between the therapist and the child plays a pivotal role in achieving therapeutic outcomes. Therefore, therapist training should emphasize not only intervention techniques but also fostering empathetic relationships with children.

This study provides a solid foundation for applying the SSL program in diverse contexts and offers valuable guidance for future research in the area of IF. The high fidelity achieved in the Spanish SSL context the role of implementation quality in ensuring the program's effectiveness. Findings from this and previous Spanish research highlight that high-fidelity implementation is associated with better outcomes, particularly in group formats. However, the global reach of SSL through the train-the-trainer strategy raises important questions about its implementation across diverse settings. While this strategy has proven effective in scaling evidence-based practices to schools and communities, limited research has examined its cost-effectiveness or the level of support needed to sustain fidelity in implementations (Eiraldi et al., 2024). Future research should explore how SSL is being delivered in other countries, the degree to which fidelity is maintained in these settings, and the program's effectiveness under varying cultural and logistical conditions. Such studies will be critical for optimizing SSL's global scalability while preserving its evidence-based rigor, with the Spanish research serving as a reference point for implementation research.

4.1. Strengths and limitations

This study has several notable strengths. First, it provides a comprehensive analysis of IF in both traditional and computerized individual formats of the SSL program, which is a novel contribution to the field. By directly comparing these two formats, we offer valuable insights into how different delivery methods impact IF and intervention outcomes. Second, we included a measure of *quality of delivery* (therapist kindness) and multiple measures of *responsiveness* (session satisfaction, perceived usefulness, degree of recommendation, and number of *Supertasks* completed), enhancing the understanding of these IF dimensions and their influence on treatment effectiveness. Third, focusing on children aged 8–12 years, a critical developmental stage for socio-emotional development, makes our findings particularly relevant for early interventions aimed at preventing the progression of emotional disorders.

However, this study also has several limitations. First, we focused only on the dimensions of *responsiveness* and *quality of delivery*, which may limit a broader understanding of fidelity. Additionally, it relied on children's self-reports, susceptible to bias, with only one objective measure (*Supertask*). Second, families who agreed to participate may differ from those who declined, potentially affecting the representativeness of the sample. We did not collect detailed information on those who chose not to participate, which limits our ability to assess potential recruitment biases. Moreover, the same team of researchers conducted the recruitment, randomization, and delivery of the interventions, which may have introduced some bias due to the lack of allocation concealment, despite adherence to the randomization procedure. Lastly, a small sample size restricted to one Spanish region also limits generalizability. Nevertheless, the sample size proved sufficient to detect meaningful effects in key outcomes, supporting the relevance of the findings despite the study's limited scope. Future research should include other IF dimensions from the beginning and plan for high variability between participants (Fernández-Martínez, Orgilés, Espada, Essau, & Morales, 2021) and other evaluation methods such as interviews with facilitators and observation. Furthermore, expanding the scale to include multiple regions and larger sample sizes would enhance the generalizability of findings and provide a more comprehensive understanding of implementation fidelity across different contexts.

4.2. Conclusions

This study provides valuable insights into the IF of the SSL program in both traditional and computerized individual formats, showing high fidelity across various measures and emphasizing the importance of the therapist attitude in reducing anxiety and depression. While computerized formats may enhance accessibility, integrating face-to-face elements boost satisfaction and engagement. Future research should aim to address the identified limitations by incorporating comprehensive evaluations of all fidelity dimensions, utilizing multiple assessment methods, and expanding sample sizes to enhance the generalizability of the findings.

Ethical information

The study received approval from the ethical committee of the authors' institution (ref. DPS.MOA.01.22 PROV) and the Institute of Health and Biomedical Research of Alicante (ref. CEIm: 2022–013). It was registered in ClinicalTrials.gov (Identifier: NCT05574491).

Funding

This work was supported by the Ministry of Universities under Grant FPU20/00893; the European Regional Development Fund (ERDF) under Grant PSI2017–85493-P/AEI/10.13039/501100011033

CRediT authorship contribution statement

Orgilés Mireia: Visualization, Resources, Funding acquisition, Validation, Project administration, Conceptualization, Writing – review & editing, Supervision, Methodology. **Galán-Luque Teresa:** Writing – original draft, Software, Methodology, Data curation, Visualization, Resources, Investigation, Conceptualization, Writing – review & editing, Validation, Project administration, Formal analysis.

Declaration of Competing Interest

We have no conflicts of interest to disclose.

Acknowledgements

We would like to express our gratitude to all the families and therapists who participated in this study.

References

Angold, A., Costello, E. J., Messer, S. C., & Pickles, A. (1995). Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. *International Journal of Methods in Psychiatric Research, 5*(4), 237–249.

Bakker, D., Kazantzis, N., Rickwood, D., & Rickard, N. (2016). Mental health smartphone apps: Review and evidence-based recommendations for future developments. *JMIR Mental Health, 3*(1), e7. https://doi.org/10.2196/mental.4984

Barlow, D. H., Allen, L. B., & Choate, M. L. (2016). Toward a unified treatment for emotional disorders – Republished article. *Behavior Therapy, 47*(6), 838–853. https://doi.org/10.1016/j.beth.2016.11.005

Berkel, C., Mauricio, A. M., Schoenfelder, E., & Sandler, I. N. (2011). Putting the pieces together: An integrated model of program implementation. *Prevention Science, 12*(1), 23–33. https://doi.org/10.1007/s11121-010-0186-1

Berkovits, M. D., O'Brien, K. A., Carter, C. G., & Eyberg, S. M. (2010). Early identification and intervention for behavior problems in primary care: a comparison of two abbreviated versions of parent-child interaction therapy. *Behavior Therapy, 41*(3), 375–387. https://doi.org/10.1016/j.beth.2009.11.002

Botvin, G. J., & Griffin, K. W. (2004). Life Skills Training: Empirical findings and future directions. *The Journal of Primary Prevention, 25*(2), 211–232. https://doi.org/10.1023/B:JOPP.0000042391.58573.5b

Botvin, G. J., Griffin, K. W., Botvin, C., Murphy, M., & Acevedo, B. (2018). Increasing implementation fidelity for school-based drug abuse prevention: Effectiveness of enhanced training and technical assistance. *Journal of the Society for Social Work and Research, 9*(4), 599–613. https://doi.org/10.1086/700972

Breitenstein, S. M., Gross, D., Garvey, C. A., Hill, C., Fogg, L., & Resnick, B. (2010). Implementation fidelity in community-based interventions. *Research in Nursing Health, 33*(2), 164–173. https://doi.org/10.1002/nur.20373

Brignardello-Petersen, R., Ioannidis, J. A., Tomlinson, G., & Guyatt, G. (2015). Surprising results of randomized trials. In G. Guyatt, D. Rennie, M. O. Meade, & D. J. Cook (Eds.), *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice* (3rd ed). McGraw-Hill. ⟨https://jamaevidence.mhmedical.com/content.aspx?bookid=847&sectionid=69031469⟩ (Education).

Bunnell, B. E., Nemeth, L. S., Lenert, L. A., Kazantzis, N., Deblinger, E., Higgins, K. A., & Ruggiero, K. J. (2021). Barriers associated with the implementation of homework in youth mental health treatment and potential mobile health solutions. *Cognitive Therapy and Research, 45*(2), 272–286. https://doi.org/10.1007/s10608-020-10090-8

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science, 2*(1), 40. https://doi.org/10.1186/1748-5908-2-40

Carthy, T., Horesh, N., Apter, A., Edge, M. D., & Gross, J. J. (2010). Emotional reactivity and cognitive regulation in anxious children. *Behaviour Research and Therapy, 48*(5), 384–393. https://doi.org/10.1016/j.brat.2009.12.013

Chu, B. C. (2012). Translating transdiagnostic approaches to children and adolescents. *Cognitive and Behavioral Practice, 19*(1), 1–4. https://doi.org/10.1016/j.cbpra.2011.06.003

Christ, T. W. (2013). Scientific-Based Research and Randomized Controlled Trials, the "Gold" Standard? Alternative Paradigms and Mixed Methodologies. *Qualitative Inquiry, 20*(1), 72–80. https://doi.org/10.1177/1077800413508523

Danko, C. M., Brown, T., Van Schoick, L., & Budd, K. S. (2016). Predictors and correlates of homework completion and treatment outcomes in parent–child interaction therapy. *Child & Youth Care Forum, 45*(3), 467–485. https://doi.org/10.1007/s10566-015-9339-5

Daviss, W. B., Birmaher, B., Melhem, N. A., Axelson, D. A., Michaels, S. M., & Brent, D. A. (2006). Criterion validity of the Mood and Feelings Questionnaire for depressive episodes in clinic and non-clinic subjects. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 47*(9), 927–934. doi: 10.1111/j.1469-7610.2006.01646.x.

Diego-Castaño, S., Morales, A., & Orgilés, M. (2023). Benefits of Super Skills for Life in a randomized controlled trial in clinical settings for Spanish children with comorbid conditions. *Development and Psychopathology*, 1–10. https://doi.org/10.1017/S0954579423001189

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*(3–4), 327–350. https://doi.org/10.1007/s10464-008-9165-0

Dusenbury, L. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research, 18*(2), 237–256. https://doi.org/10.1093/her/18.2.237

Ecclestone, K., & Hayes, D. (2009). Changing the subject: The educational implications of developing emotional well-being. *Oxford Review of Education, 35*(3), 371–389. https://doi.org/10.1080/03054980902934662

Eiraldi, R., Lawson, G. M., Glick, H. A., Khanna, M. S., Beidas, R., Fishman, J., … Jawad, A. F. (2024). Implementation fidelity, student outcomes, and cost-effectiveness of train-the-trainer strategies for Masters-level therapists in urban schools: results from a cluster randomized trial. *Implementation Science, 19*(1), 4. https://doi.org/10.1186/s13012-023-01333-9

Espada, J. P., Gonzálvez, M. T., Fernández-Martínez, I., Orgilés, M., & Morales, A. (2022). Spanish validation of the Short Mood and Feelings Questionnaire (SMFQ) in children aged 8-12. *Psicothema, 34*(4), 610–620. https://doi.org/10.7334/psicothema2022.54

Essau, C. A., Lewinsohn, P. M., Lim, J. X., Ho, M. R., & Rohde, P. (2018). Incidence, recurrence and comorbidity of anxiety disorders in four major developmental stages. *Journal of Affective Disorders, 228*, 248–253. https://doi.org/10.1016/j.jad.2017.12.014

Essau, C.A., & Ollendick, T.H. (2013). The Super Skills for Life Programme. *University of Roehampton: London, UK*.

Essau, C. A., Sasagawa, S., Anastassiou-Hadjicharalambous, X., Guzmán, B. O., & Ollendick, T. H. (2011). Psychometric properties of the Spence Child Anxiety Scale with adolescents from five European countries. *Journal of Anxiety Disorders, 25*(1), 19–27. https://doi.org/10.1016/j.janxdis.2010.07.001

Fernández-Martínez, I., Morales-Sabuco, A., Espada, J. P., Orgilés, M., & Essau, C. A. (2019). Effectiveness of the program Super Skills For Life in reducing symptoms of anxiety and depression in young Spanish children. *Psicothema*. https://doi.org/10.7334/psicothema2018.336

Fernández-Martínez, I., Orgilés, M., Espada, J. P., Essau, C. A., & Morales, A. (2021). Effects as a function of implementation fidelity of a transdiagnostic prevention program in young school-aged children. Evaluation and Program. *Planning, 89*, 102011. https://doi.org/10.1016/j.evalprogplan.2021.102011

Fernández-Martínez, I., Orgilés, M., Morales, A., Espada, J. P., & Essau, C. A. (2020). One-year follow-up effects of a cognitive behavior therapy-based transdiagnostic program for emotional problems in young children: A school-based cluster-randomized controlled trial. *Journal of Affective Disorders, 262*, 258–266. https://doi.org/10.1016/j.jad.2019.11.002

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160.

Gasol Foundation. (2022). *Estudio PASOS 2022*. ⟨https://gasolfoundation.org/es/estudio-pasos/⟩.

Galán-Luque, T., Cabello, E., & Orgilés, M. (2024a). Comparative randomized trial addressing childhood emotional problems: Computerized vs. traditional transdiagnostic Super Skills program. *Journal of Affective Disorders, 365*, 95–104. https://doi.org/10.1016/j.jad.2024.08.081

Galán-Luque, T., Cabello, E., & Orgilés, M. (2024b). Evaluating Super Skills for Life program's impact on transdiagnostic variables in children with emotional disorders: A comparative randomized trial. *Translational Issues in Psychological Science, 10*(4), 388–407. https://doi.org/10.1037/tps0000429

Gázquez Pertusa, M., García del Castillo, J. A., & Ruiz Alfaro, I. (2011). Importance of fidelity in school programs implementation for prevention of drug consumption. *Health and Addictions/Salud York Drogas, 11*(1). https://doi.org/10.21134/haaj.v11i1.22

Goldberg, D. P., Krueger, R. F., Andrews, G., & Hobbs, M. J. (2009). Emotional disorders: Cluster 4 of the proposed meta-structure for DSM-V and ICD-11. *Psychological Medicine, 39*(12), 2043–2059. https://doi.org/10.1017/S0033291709990298

Guo, T., Su, J., Hu, J., Aalberg, M., Zhu, Y., Teng, T., & Zhou, X. (2021). Individual vs. group cognitive behavior therapy for anxiety disorder in children and adolescents: A meta-analysis of randomized controlled trials. *Frontiers in Psychiatry, 12*, 674267. https://doi.org/10.3389/fpsyt.2021.674267

Hansen, W. B., Graham, J. W., Wolkenstein, B. H., & Rohrbach, L. A. (1991). Program integrity as a moderator of prevention program effectiveness: results for fifth-grade students in the adolescent alcohol prevention trial. *Journal of Studies on Alcohol, 52*(6), 568–579. https://doi.org/10.15288/jsa.1991.52.568

Hudson, J. L., & Kendall, P. C. (2002). Showing you can do it: Homework in therapy for children and adolescents with anxiety disorders. *Journal of Clinical Psychology, 58*(5), 525–534. https://doi.org/10.1002/jclp.10030

Jaccard, J., & Turrisi, R. (2003). *Interaction effects in multiple regression*. SAGE Publications, Inc.. https://doi.org/10.4135/9781412984522

JASP Team. (2024). *JASP* (0.18.3).

Kazantzis, N., & Deane, F. P. (1999). Psychologists' use of homework assignments in clinical practice. *Professional Psychology: Research and Practice, 30*(6), 581–585. https://doi.org/10.1037/0735-7028.30.6.581

Kazantzis, N., Deane, F. P., & Ronan, K. R. (2005). Assessment of homework completion. In N. Kazantzis, F. P. Deane, K. R. Ronan, & L. L'Abate (Eds.), *Using homework assignments in cognitive behavior therapy* (pp. 61–72). Routledge/Taylor & Francis Group.

Kazantzis, N., Whittington, C., Zelencich, L., Kyrios, M., Norton, P. J., & Hofmann, S. G. (2016). Quantity and quality of homework compliance: A meta-analysis of relations with outcome in cognitive behavior therapy. *Behavior Therapy, 47*(5), 755–772. https://doi.org/10.1016/j.beth.2016.05.002

Kendall, P. C., Ney, J. S., Maxwell, C. A., Lehrbach, K. R., Jakubovic, R. J., McKnight, D. S., & Friedman, A. L. (2023). Adapting CBT for youth anxiety: Flexibility, within fidelity, in different settings. *Frontiers in Psychiatry, 14*, 1067047. https://doi.org/10.3389/fpsyt.2023.1067047

Kent, L., Vostanis, P., & Feehan, C. (1997). Detection of major and minor depression in children and adolescents: evaluation of the Mood and Feelings Questionnaire. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 38*(5), 565–573. https://doi.org/10.1111/j.1469-7610.1997.tb01543.x

Lattie, E. G., Adkins, E. C., Winquist, N., Stiles-Shields, C., Wafford, Q. E., & Graham, A. K. (2019). Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: Systematic review. *Journal of Medical Internet Research, 21*(7), Article e12869. https://doi.org/10.2196/12869

Leventhal, A. M. (2008). Sadness, depression, and avoidance behavior. *Behavior Modification, 32*(6), 759–779. https://doi.org/10.1177/0145445508317167

Long, E. E., Young, J. F., & Hankin, B. L. (2018). Temporal dynamics and longitudinal co-occurrence of depression and different anxiety syndromes in youth: Evidence for reciprocal patterns in a 3-year prospective study. *Journal of Affective Disorders, 234*, 20–27. https://doi.org/10.1016/j.jad.2018.02.074

López-Martín, E., & Ardura, D. (2023). The effect size in scientific publication. *EducaciónXX1, 26*(1), 9–17. https://doi.org/10.5944/educxx1.36276

Melero, S., Orgilés, M., Espada, J. P., & Morales, A. (2021). Spanish version of Super Skills for Life in individual modality: Improvement of children's emotional well-being from a transdiagnostic approach. *Journal of Clinical Psychology, 77*(10), 2187–2202. https://doi.org/10.1002/jclp.23148

Melero, S., Orgilés, M., Fernández-Martínez, I., Espada, J. P., & Morales, A. (2021). Influence of implementation fidelity on the effectiveness of a T-CBT program targeting emotional problems in childhood. *Studies in Educational Evaluation, 68*, Article 100975. https://doi.org/10.1016/j.stueduc.2021.100975

Mychailyszyn, M. P., Beidas, R. S., Benjamin, C. L., Edmunds, J. M., Podell, J. L., Cohen, J. S., & Kendall, P. C. (2011). Assessing and treating child anxiety in schools. *Psychology in The Schools, 48*(3), 223–232. https://doi.org/10.1002/pits.20548

National Research Council. (2002). Scientific research in education. In R. J. Shavelson, & L. Towne (Eds.), *Committee on Scientific Principles for Education Research*. Washington, DC: National Academy Press.

Nelson, P. M., Pulles, S. M., Parker, D. C., & Klaft, J. (2020). Implementation fidelity for math intervention: Basic quality ratings to supplement adherence. *School Psychology, 35*(1), 72–79. https://doi.org/10.1037/spq0000338

Norcross, J. C., & Wampold, B. E. (2011). Evidence-based therapy relationships: Research conclusions and clinical practices. *Psychotherapy, 48*(1), 98–102. https://doi.org/10.1037/a0022161

Nuñez, L., Fernández, S., Alamo, N., Midgley, N., Capella, C., & Krause, M. (2022). The therapeutic relationship and change processes in child psychotherapy: A qualitative, longitudinal study of the views of children, parents and therapists. *Research in Psychotherapy (Milano), 25*(1), 556. https://doi.org/10.4081/ripppo.2022.556

Oosthuizen, C., & Louw, J. (2013). Developing program theory for purveyor programs. *Implementation Science, 8*(1), 23. https://doi.org/10.1186/1748-5908-8-23

Orgilés, M., Morales, A., Fernández-Martínez, I., Méndez, X., & Espada, J. P. (2023). Effectiveness of a transdiagnostic computerized self-applied program targeting children with emotional problems: A randomized controlled trial. *Journal of Affective Disorders, 338*, 155–162. https://doi.org/10.1016/j.jad.2023.06.004

Orgilés, M., Rodríguez-Menchón, M., Fernández-Martínez, I., Morales, A., & Espada, J. P. (2019). Validation of the parent report version of the Spence Children's Anxiety Scale (SCAS-P) for Spanish children. *Clinical Child Psychology and Psychiatry, 24*(4), 776–790. https://doi.org/10.1177/1359104519835579

Osma, J., Suso-Ribera, C., Peris-Baquero, Ó., Gil-Lacruz, M., Pérez-Ayerra, L., Ferreres-Galan, V., … Domínguez, O. (2019). What format of treatment do patients with emotional disorders prefer and why? Implications for public mental health settings and policies. *PloS one, 14*(6), Article e0218117. https://doi.org/10.1371/journal.pone.0218117

R Core Team. (2018). *RStudio: integrated development for R (Version 1.1.453)*. ⟨https://www.R-project.org/⟩.

Rosen, J. B., & Schulkin, J. (1998). From normal fear to pathological anxiety. *Psychological Review, 105*(2), 325–350. https://doi.org/10.1037/0033-295X.105.2.325

Shirk, S. R., & Karver, M. (2003). Prediction of treatment outcome from relationship variables in child and adolescent therapy: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 71*(3), 452–464. https://doi.org/10.1037/0022-006X.71.3.452

Simons, A. D., Marti, C. N., Rohde, P., Lewis, C. C., Curry, J., & March, J. (2012). Does homework "matter" in cognitive behavioral therapy for adolescent depression? *Journal of Cognitive Psychotherapy, 26*(4), 390–404. https://doi.org/10.1891/0889-8391.26.4.390

Spence, S. H. (1998). A measure of anxiety symptoms among children. *Behaviour Research and Therapy, 36*(5), 545–566. https://doi.org/10.1016/S0005-7967(98)00034-5

Spence, S. H. (2021). *The Spence Children'States Anxiety Scale Website*. ⟨https://www.scaswebsite.com⟩.

Stiles-Shields, C., Montague, E., Lattie, E. G., Kwasny, M. J., & Mohr, D. C. (2017). What might get in the way: Barriers to the use of apps for depression. *Digital Health, 3*. https://doi.org/10.1177/2055207617713827, 2055207617713827.

Watt, B. D., & Dadds, M. R. (2007). Facilitating treatment attendance in child and edolescent mental health services: A community study. *Clinical Child Psychology and Psychiatry, 12*(1), 105–116. https://doi.org/10.1177/1359104507071089

Wood, A., Kroll, L., Moore, A., & Harrington, R. (1995). Properties of the Mood and Feelings Questionnaire in adolescent psychiatric outpatients: A research note. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 36*(2), 327–334. https://doi.org/10.1111/j.1469-7610.1995.tb01828.x