

**Effects as a function of Implementation Fidelity of a Transdiagnostic Prevention
Program in Young School-Aged Children**

Fernández-Martínez, I., Orgilés, M., Espada, J.P., Essau, C.A., & Morales, A.

Accepted for publication in “Evaluation and Program Planning”

(September 16, 2021)

Abstract

Super Skills for Life (SSL) is a transdiagnostic prevention program for school-aged children with internalizing problems. The current study aimed to examine the short- and long-term (12-month follow-up) effectiveness of SSL, depending on implementation fidelity. Participants were 123 Spanish-speaking children aged 6-8 years with internalizing symptoms and their parents. High ($n = 41$) and low fidelity ($n = 26$) groups were established and compared with a control group (CG) ($n = 56$). Compared to CG, the high-fidelity group (HFG) achieved better outcomes than the low-fidelity group (LFG) at short and long term, including symptoms of anxiety, depression, and internalizing problems, and also externalizing problems at long-term. The LFG improved key symptoms at short term (depression) and long term (anxiety and depression). Findings suggest that SSL is beneficial even if not delivered at maximum fidelity, although considering the influence of implementation fidelity is critical because the high-fidelity implementation yielded the greatest effects when compared to the CG. This research provides valuable data, analyzing for the first time the influence of fidelity on SSL outcomes in young school-aged children exhibiting internalizing symptoms, while there is still a lack of studies of these characteristics concerning transdiagnostic prevention programs targeting childhood anxiety and depression.

Keywords: Implementation fidelity; Internalizing symptoms; Super Skills for Life; Transdiagnostic program; Young children.

1. Introduction

Implementation fidelity (IF), namely the degree to which a program or intervention is delivered as originally planned or intended, has been identified as key in driving sound evidence-based practice (Carroll et al., 2007). Its systematic evaluation is also considered important, as it enhances the validity of studies, it is a good estimate of the quality of implementation, and is useful to avoid Type-III errors (Breitenstein et al., 2010). Furthermore, the influence of IF on the impact of an intervention or program has been supported. Thus, generally better fidelity correlates with better outcomes, and low fidelity increases the likelihood of weaker outcomes (Hill & Erickson, 2019). According to Carroll et al. (2007), it is critical to attend to IF, as it can moderate the link between an intervention and its results, and also avoid erroneous conclusions about the effectiveness of interventions and lead to better results. However, following other authors in this area, the extent to which IF impacts intervention results is typically ignored and scarcely reported (Nelson, Pulles, Parker, & Kluft, 2020; Rojas-Andrade & Bahamondes, 2019).

Knowledge of IF has grown in past years, leading to different frameworks, data collection methods, or fidelity dimensions (see Breitenstein et al., 2010; Carroll et al., 2007; Dusenbury, Brannigan, Falco, & Hansen, 2003; Hasson, 2010). One or more dimensions could be considered to measure IF, while there is no agreement regarding inclusion, measurement, and definitions of such dimensions. Some principal dimensions may be dose or exposure (e.g., number of sessions delivered or received by participants), adherence (e.g., delivery of prescribed intervention content), or participant responsiveness (e.g., engagement, involvement in the program's activities) (Carroll et al., 2007; Dusenbury et al., 2003). More recent studies such as the one by Bettencourt, Gross, and Breitenstein (2019) evaluated the IF of a school-based program including

some dimensions (e.g., participant satisfaction, weekly practice completion, adherence, attendance). Other school-based trials examined the effectiveness of preventive interventions according to the degree of IF, including acceptance in terms of participant satisfaction, adherence, or dose (Escribano, Espada, Orgilés, & Morales, 2016; Escribano, Morales, Orgilés, & Espada, 2015). In addition, a review of mental health programs delivered in schools found that IF dimensions concerning participants (e.g., number of sessions attended or exposure) were more strongly linked to outcomes than those more focused on provider practices (e.g., adherence) (Rojas-Andrade & Bahamonde, 2019).

In the recent literature, some studies have focused on the measurement of IF or issues related to the implementation process of interventions in school settings (e.g., Bettencourt et al., 2019; Biel et al., 2020; Martinsen, Kendall, Stark, & Neumer, 2016; Rasmussen et al., 2019). Other review studies have addressed the establishment of implementation strategies for mental health interventions in the school environment or have shown the links between IF and the outcomes of such interventions (Cook, Lyon, Locke, Waltz, & Powell, 2019; Rojas-Andrade & Bahamondes, 2019). Moreover, results of school-based trials have revealed that the effects of prevention programs may be influenced by the degree of IF, finding that a higher degree of IF is related to better results when comparing high and low fidelity groups with a control group (Escribano et al., 2015, 2016). These findings reveal the importance of examining this variable and the growing interest it attracts.

However, in the field of prevention of internalizing problems in children, data on the IF of preventive interventions addressing childhood depression and anxiety, including those applied at the school level, are still scarce, although it has been underlined that its study should be a priority to identify its impact on outcomes and

parameters to improve preventive effects (see Werner-Seidler et al., 2017). In this regard, it is well established that internalizing problems such as anxiety and depression are frequent and comorbid among children (e.g., Ghandour et al., 2019). From early ages, symptoms of these internalizing problems can follow stable and increasing trajectories and present homotypic and heterotypic continuities (Lewis, Sae-Koew, Toumbourou, & Rowland, 2020; Wichstrøm, Belsky, & Steinsbekk, 2017). Thus, effective preventive programs for anxiety and depression should be developed and applied at an early age (Bayer & Beatson, 2013; Bernaras, Jaureguizar, & Garaigordobil, 2019). Preventive interventions provided in school settings have been shown to have positive effects on child mental health, including internalizing problems. However, their effectiveness will depend on their proper implementation, as they may not be effective if they are not applied with accuracy and fidelity (Paulus, Ohmann, & Popow, 2016; Weare & Nind, 2011).

Complementarily, the need to advance the study of the effectiveness of anxiety and depression prevention programs for school children has been raised, considering long-term effects, and with particular focus on transdiagnostic programs (i.e., addressing both conditions within the same protocol), as they can be potentially beneficial (see Johnstone, Kemps, & Chen, 2018). One of the few existing transdiagnostic prevention programs is the Super Skills for Life (SSL; Essau & Ollendick, 2013; Essau et al., 2014). SSL is an 8-session, cognitive-behavioral therapy (CBT)-based program, which is usually delivered in schools and targets schoolchildren who exhibit initial symptoms of anxiety and/or depression. SSL is effective in the short and long term (up to 12-month follow-up) in reducing symptoms of anxiety and depression in schoolchildren aged 8-12 years (Essau, Sasagawa, Jones, Fernandes, & Ollendick, 2019; Orgilés, Fernández-Martínez, Espada, & Morales, 2019), but also in

younger schoolchildren aged 6-8 years (Author et al., 2019, 2020). Furthermore, the results of SSL have shown a positive impact on symptoms of externalizing problems (Essau et al., 2014; Orgilés et al., 2019). However, to date, no randomized control trials (RCT) have been conducted to examine the influence that the degree of IF can have on the effectiveness of SSL when compared with a control group.

The present study was part of a school-based cluster-randomized controlled trial (Author et al., 2019, 2020) aimed at evaluating, in a sample of young Spanish-speaking children aged 6-8 years with internalizing symptoms, the short-term (post-test) and long-term (12-month follow-up) effectiveness of the SSL program according to the degree of IF in reducing anxiety and depression symptoms and other secondary outcomes (i.e., internalizing and externalizing problems). Therefore, these were secondary analyses of an already reported outcome study (Author et al., 2019, 2020), which, in turn, is the only trial of the SSL program with young children including a control group to date. Therefore, concerning previous studies on the above-mentioned SSL program, this study aimed to provide new evidence on how IF influences program efficacy in the short and long term with young children, comparing high and low degrees of IF and a control group. Considering prior research (e.g., Escribano et al., 2015, 2016), this study evaluated IF in three ways or fidelity dimensions, namely dose (measured as participants' attendance to SSL sessions), participant responsiveness (measured as participants' completion of SSL weekly activities or homework), and acceptance (measured as participants' level of satisfaction with the program). Fidelity groups were derived based on the use of a cut-off (i.e., the 50th percentile) to establish a high or low level in each IF dimension. It was hypothesized that the effects of the SSL program on outcomes would be greater in a group with a higher degree of IF compared to a control group.

2. Methods

2.1. Study design and Participants

This trial involved a total of 123 young children aged 6-8 years and their father or mother at the baseline. Children involved in the study were in the 1st, 2nd, and 3rd grades of primary school, with a mean age of 6.89 ($SD = 0.79$) of whom 21.1% were boys. Most of the children had been born in Spain (98.4%), all of them were Spanish-speaking. The 10 participating schools were from the south-east area of Spain, and were randomly distributed to the following experimental conditions: experimental (SSL) group or waiting-list control group (CG). Most parents reported being married ($n = 106$; 78.9%) and having higher education ($n = 62$; 50.4%). A medium-high socio-economic status predominated. Parents gave their informed consent for their children's participation in the study and completed the assessment instruments that were included in this study at baseline, post-test, and 12-month follow-up.

All children whose parent-report assessments were available at baseline, post-test, and 12-month follow-up were included in the analyses. At baseline, 54.5% ($n = 67$) were included in the SSL group, and 45.5% ($n = 56$) were in the CG. Considering the degree of IF (i.e., the 50th percentile was used as a cutoff), the SSL group was divided into two new conditions: high-fidelity group (HFG) or low-fidelity group (LFG).

The HFG comprised those children who scored above 50% on all fidelity dimensions ($n = 41$, 33.3%). Therefore, children in this condition attended 7 to 8 of the 8 SSL sessions (dose), completed weekly homework in 5 to 7 out of 7 sessions where homework was required (participant responsiveness), and had a high level of participant satisfaction with the program by scoring 9 to 10 on a 10-point Likert scale (acceptance). Children categorized in the LFG ($n = 26$, 21.1%) did not accomplish the 50% criterion for any fidelity dimension. They attended from 0 to 6 sessions, did the homework in 0

to 4 sessions, and scored 0 to 8.99 on the 10-point Likert scale. Table 1 shows the sociodemographic characteristics and parent-reported measures of their children at baseline according to intervention condition.

2.2. *Intervention*

The transdiagnostic SSL program is based on the principles of CBT, social skills training, and behavioral activation. It is highly structured and includes an intervention manual that is led by facilitators during the eight 45-minute sessions. This manual contains all the details about the structure and contents of the sessions (i.e., objectives, timeline, activities, explanations, answers, materials required). Also, SSL provides participants with a workbook containing explanations of relevant concepts, activities, and homework. The methodology of SSL is based on positive reinforcement and the combination of different types of activities (i.e., group and individual exercises, games, readings, role-playings, video feedback with cognitive preparation, homework), promoting the learning of the target strategies (i.e., psychoeducation, cognitive restructuring, behavioral activation, training in social skills and relaxation, self-monitoring, social problem-solving) through a supportive and stimulating environment. SSL has generally been implemented weekly, in reduced groups, and in the school context. Further details about SSL have been provided in previous studies (see Essau et al., 2014, 2019; Orgilés et al., 2019).

2.3. *Measures*

2.3.1. *Primary outcomes*

2.3.1.1. *Depression.* The Mood and Feelings Questionnaire – parent version (MFQ-P; Angold et al., 1995) is a 34-item tool developed to assess symptoms of childhood depression through parents. The items are scored using a 3-point scale ranging from 0 (*not true*) to 2 (*true*). The MFQ-P provides a single total score (i.e., called *Depression*

in this study), calculated by summing all the item scores. The higher the score, the more severe the depressive symptoms. The Spanish version of the MFQ-P showed good psychometric properties (Fernández-Martínez, Morales, Espada, & Orgilés, 2020). In this study, total-score reliability was high (ordinal $\alpha = .94$).

2.3.1.2. Anxiety. The Spence Children's Anxiety Scale – parent version (SCAS-P; Nauta et al., 2004) evaluates anxiety symptoms in children and adolescents from their parents' perspective. Parents assess their child's anxiety on a four-point scale ranging from 0 (*never*) to 3 (*always*) across 38 items. The total SCAS score was used in this study (i.e., called *Anxiety*), obtained from summing all the items. Higher scores denote a higher severity of symptoms. The Spanish version of the SCAS-P is a valid and reliable instrument (Orgilés, Rodríguez-Menchón, Fernández-Martínez, Morales, & Espada, 2019). In the current study, total-score reliability was high (ordinal $\alpha = .89$).

2.3.2. Secondary outcomes

2.3.2.1. Internalizing and externalizing problems. The Strengths and Difficulties Questionnaire – parent version (SDQ-P; Goodman, 2001) evaluates children's emotional and behavioral problems through their parents by computing four subscales, which contain 5 items each and are scored from 0 (*not true*) to 2 (*not true*). The composite/subscale of *Internalizing Problems* is obtained by computing two subscales: Emotional Symptoms (i.e., anxiety and depression) and Peer Relationship Problems. The composite/subscale of *Externalizing Problems* is obtained by computing two subscales: Hyperactivity/Inattention and Conduct Problems. The scores of each composite can range from 0 to 20, with higher scores denoting greater difficulties. The psychometric properties of the Spanish SDQ-P have been supported (Rodríguez-Hernández et al., 2012). In this study, total-score reliability (i.e., the sum of the four subscales' scores) was satisfactory (ordinal $\alpha = .85$).

2.4. Implementation fidelity

Three fidelity dimensions were evaluated in this study. The acceptance dimension was evaluated through the children's response to a question referring to the level of satisfaction with the program sessions ("How much did you like today's session?"), which was answered at the end of each SSL session on a Likert scale ranging from 0 (*not at all*) to 10 (*very much*). Facilitators assessed the dose and participant responsiveness dimensions, recording the participants' attendance at each SSL session (i.e., dose) and whether the children completed the weekly homework (i.e., participant responsiveness) from the second to the last session (as the first weekly homework had to be done between the first and second session). For each of their groups' participants, the facilitators answered dichotomously (yes/no) two questions included in a form that they completed after the sessions (i.e., "Did the child attend today's session?", "Did the child complete this week's homework?").

2.5. Procedure

In this trial, 10 schools located in urban areas of the southeastern region of Spain were recruited based on their potential to represent the socio-economic structure of the Spanish population, with the participation of public, private, and state-assisted private schools. Schools were the unit of randomization, and each experimental condition, SSL or CG, was composed of five schools. Through the schools, parents received the information about the study and voluntarily accessed a link to an online form to complete the assessment instruments and provide the sociodemographic data. Children were selected based on the following inclusion/exclusion criteria: a) being Spanish-speaking children in the age range of 6 to 8 years, b) exhibiting emotional symptoms based on scores greater than or equal to 4 on the subscale of Emotional Symptoms (i.e., anxiety and depression) as measured by the Strengths and Difficulties Questionnaire-

Parent version (SDQ-P; Goodman, 2001), c) not undergoing psychiatric or psychological treatment; and (d) not having developmental disabilities or severe learning difficulties. In addition to the first evaluation conducted, the parents of the selected children completed the same measures through the online form after the eight-week SSL intervention (post-test) and at the 12-month follow-up. All three assessments were completed by the same parent of each child (i.e., mother or father). As reported in the previous studies about the trial from which the present study is derived (Fernández-Martínez et al., 2019; Fernández-Martínez et al., 2020), participants lost to follow-up in this study because of not completing the post-test and 12-month follow-up assessment when required were as follows: At post-test, seven participants (10.4%) from the intervention group and 14 (25%) from the control group. At the 12-month follow-up assessment, six participants (9%) from the intervention group and 10 (17.9%) from the control group. The flow of participants throughout the trial at baseline, post-test, and 12-month follow-up can be found in the studies by Authors et al. (2019, 2020).

Facilitators were seven psychologists with a Psychology Master's degree, mostly in psychological therapy with children and adolescents. They received intensive one-day training in the SSL program at the authors' institution. Additionally, weekly meetings during the intervention were held with the facilitators to monitor the functioning of the implementation, provide necessary materials, and exchange relevant information. At the end of each session, the facilitators registered relevant data about each group (e.g., attendance, homework completion, contents delivered, comments/observations), which served to track the progress of the intervention and assess the implementation accuracy. The eight SSL sessions were held weekly, at schools and in small groups. All groups received the full eight 45-minute sessions as

planned. The approval of the Ethics Board of the authors' institution was received for this research.

2.6. Data Analysis

Descriptive statistics were calculated to determine sample characteristics (children and parents), the study variables (depression, anxiety, internalizing and externalizing problems), and the dimensions of IF.

In the current study, IF included three components: dose (attendance at the sessions), participant responsiveness (doing the homework for each session), and acceptance (level of satisfaction with the program). Following Escribano et al. (2016), a new variable was created using a cut-off (i.e., the 50th percentile) to establish a high or low level in the IF dimension. In the SSL group, children who scored above the 50th percentile in dose, participant responsiveness, and acceptance were assigned to the high-fidelity group (HFG). The rest were assigned to the low-fidelity group (LFG). Both experimental groups (HFG and LFG) were compared with the CG to evaluate the effects of the intervention according to the degree of IF.

Baseline equivalence among the three conditions (HFG, LFG, and CG) was explored using ANOVA for quantitative variables and cross-tables for qualitative variables. We ran Generalized Estimating Equations (GEE; Liang & Zeger, 1986) to study the effects of SSL to reduce anxiety, depression, and internalizing and externalizing problems based on the level of IF. Each variable was analyzed independently. All analyses were adjusted for baseline differences among the three conditions, baseline score for the outcome, gender, age, and school. Following Rosenthal and Rosnow (1985), planned comparisons for pre-specified hypotheses were tested. To test the effects of HFG, one contrast compared HFG with the CG, and another contrast compared HFG with LFG. To test the effects of LFG, a contrast

compared LFG with the CG. Standardized mean differences (d) were estimated as a measure of effect size, where .20 is small, .50 medium, and .80 large, following Cohen (1988). Analyses were performed using SPSS v25.

3. Results

3.1. Evaluation of the IF of the SSL program

Related to the fidelity dimensions (i.e., dose, participant responsiveness, acceptance) evaluated in the SSL group, the results and the distribution in HFG or LFG were as follows. In the dose dimension, 17.9% of the children in the SSL group attended 0-6 out of 8 sessions and were allocated to the LFG, whereas 82.1% attended all or almost all sessions (7 – 8 sessions) and were assigned to the HFG. In the participant responsiveness dimension, 28.4% of the children in the SSL group completed their weekly homework in 0 to 4 sessions out of 7 (LFG), whereas 71.6% did so in 5 to 7 sessions (HFG). In the acceptance dimension, 19.4% of the children obtained scores of satisfaction with the program ranging from 0 to 8.99 (LFG), whereas 80.6% obtained scores of 9 to 10 (HFG).

3.2. Baseline differences in sociodemographic characteristics and outcomes by experimental condition

Considering the sociodemographic and dependent variables or outcomes of the study, few significant differences were found between experimental conditions (i.e., HFG, LFG, CG) at baseline (Table 1). Regarding the sociodemographic characteristics, the mean age of the participating parents was significantly higher in the HFG ($M_{\text{age}} = 43.14$, $SD = 5.45$) than in the GC ($M_{\text{age}} = 41.51$, $SD = 4.91$), according to the post hoc analysis ($p = .007$, $d = .72$). The parents' educational level also differed according to the experimental condition ($p = .009$), predominating a high educational level in HFG (63.4%) and a secondary educational level in GC (46.4%) compared to the other groups.

With respect to the dependent variables, differences were found in depression ($p = .03$) and internalizing problem scores ($p = .008$). Specifically, post hoc analyses revealed that both in depression ($p = .05$, $d = .58$) and internalizing problems ($p = .01$, $d = .37$), the scores in the LFG were significantly higher than those of the CG.

3.3. Effects of SSL according to the degree of IF

The effectiveness of the SSL program in the primary and secondary outcomes according to the degree of IF, which was established based on the 50th percentile of the dose, participant responsiveness, and acceptance dimensions, was assessed at posttest and 12-month follow-up (Table 2).

3.3.1. Primary outcomes

Concerning the symptoms of depression, at the posttest and 12-month follow-up, we found that symptoms generally tended to decrease in HFG and LFG throughout the assessments and to increase and be maintained in the CG. Compared to the baseline, there was a significant reduction in symptoms in the HFG and LFG compared to the CG in the short and long term, with the HFG showing the greatest positive impact (Table 2, Fig. 1). At posttest, significant differences were found between the HFG and GC ($p < .001$, $d = .65$), and the LFG and GC ($p = .03$, $d = .54$). These differences were similarly maintained at the 12-month follow-up between the HFG and GC ($p < .001$, $d = .56$), and the LFG and GC ($p = .008$, $d = .43$).

Regarding anxiety symptoms, the trend was to decrease in the HFG and LFG throughout the assessments and to increase and remain stable in the CG. Similar to the depression symptoms, the HFG showed a greater reduction of symptoms in the short and long term (Table 2, Fig. 2). At posttest, significant differences were observed between the HFG and CG ($p < .001$, $d = .42$), but not between the LFG and CG ($p =$

.08). However, at the 12-month follow-up, the results revealed significant differences between the HFG and GC ($p < .001$, $d = .44$), and the LFG and GC ($p = .005$, $d = .37$).

When the effects of the SSL program on depression and anxiety symptoms were compared in the HFG and LFG, no significant differences were found between the two conditions at the posttest and 12-month follow-up. However, although the improvement in symptoms did not differ significantly in the two groups, the marginal means indicated better short- and long-term scores in HFG for both symptoms (Table 2).

3.3.2. *Secondary outcomes*

Symptoms of internalizing problems as measured by the SDQ decreased significantly only for the HFG compared to baseline and CG. Results revealed significant differences between HFG and CG at posttest ($p < .001$, $d = .24$) and 12-month follow-up ($p = .001$, $d = .39$). No significant differences were found between LFG and CG or between HFG and LFG in any assessment.

The symptoms of externalizing problems as measured by the SDQ, compared to baseline and CG, decreased significantly at posttest only in the LFG ($p = .007$, $d = .56$), whereas at the 12-month follow-up, they only decreased in the HFG ($p = .01$, $d = .55$). No posttest differences were found between the HFG and CG ($p = .61$), or between the LFG and CG ($p = .11$) at follow-up. When the effects of the SSL program on symptoms of externalizing problems were compared in the HFG and LFG, significant differences were found between the two groups at posttest ($p = .01$, $d = .25$), but not at the 12-month follow-up ($p = .93$).

4. Discussion

The main goal of this study was to examine how the level of IF of the SSL influences SSL effectiveness, by examining IF for the first time through a school-based

cluster-randomized controlled trial with school children aged 6-8 years with internalizing symptoms.

Overall, the results indicated high IF of SSL in terms of dose, participant responsiveness, and acceptance as measured in this study, thereby showing that, of the total participants of the SSL program, 82.1% attended all or almost all the sessions, 71.6% did all or almost all of the assigned weekly homework, and 80.6% showed a very high degree of satisfaction with the program. This is in line with previous studies on school-based prevention programs (Bettencourt et al., 2019; Escribano et al., 2015, 2016) that reported a high level of fidelity. Research on another available transdiagnostic program to prevent childhood anxiety and depression, which was implemented during school hours, also reported high levels of child attendance and acceptability or satisfaction (Martinsen et al., 2016). It is noteworthy that the SSL program was implemented after school hours, suggesting that the program managed to maintain or increase the motivation of most of its participants. All these data also indicate that this type of intervention can be implemented within a school context. Although as noted by Rasmussen et al. (2019), successful implementation of interventions aimed at preventing anxiety and depression in the school settings could also be influenced by some school-related factors (e.g, level of involvement of schools, program information given to school staff).

To examine the effectiveness of SSL according to the IF degree, both experimental groups (HFG and LFG) were compared with the CG. As hypothesized, the greater the fidelity of the SSL program implementation, the greater was its effectiveness compared to the GC. Thus, complementing the efficacy studies of SSL in young children (Author et al., 2019, 2020), in the current study, SSL produced significant reductions at both short and long term in all primary outcomes (i.e., anxiety and

depression symptoms) and one secondary outcome— internalizing problems—, only for the HFG. The HFG also improved at long term in the other secondary outcome— externalizing problems. Moreover, findings also revealed that even if not applied with maximum fidelity (i.e., LFG), SSL yielded reductions in symptoms of depression and externalizing problems at short term, and both in anxiety and depression symptoms at long term. This study, therefore, suggests that, compared to CG, SSL shows benefits when applied both with low and high fidelity, but it produces the maximum benefits in the short and long term when applied with maximum fidelity. The benefits found in externalizing problems are in keeping with findings of SSL with older schoolchildren (Essau et al., 2014; Orgilés et al., 2019).

Given the lack of studies of childhood anxiety and depression prevention programs examining the fidelity of implementation and its impact on outcomes in a school setting (Werner-Seidler et al., 2017), it is difficult to compare our results on SSL with similar prior research and programs. However, our findings are consistent with those of school-based CRTs for prevention programs targeting other problems using a similar methodology (i.e., fidelity dimensions, experimental conditions) (Escribano et al., 2015, 2016). The current research results are also in line with a previous study with young children in a school-based mental health and emotional well-being promotion program (Clarke, Bunting, & Barry, 2014). These authors showed that fidelity directly impacted posttest scores of emotional literacy variables (e.g., self-regulation, self-awareness), with higher scores at higher levels of fidelity.

Interestingly, inconsistent with other studies (Dix, Slee, Lawson, & Keeves, 2012; Escribano et al., 2015, 2016), no difference in SSL effects was found when comparing the HFG with LFG at short and long term, except for short-term externalizing problems. Although there is no clear explanation, this suggests that

children benefit similarly from the intervention even if they do not receive the program with maximum fidelity. Nonetheless, marginal means indicated the expected trend, better scores for the HFG. Likewise, when compared with the CG, the HFG improved in all the symptoms assessed in the long term and in almost all symptoms at posttest, including anxiety and depression, with greater effect sizes at posttest and the 12-month follow-up than the LFG. Further, our results on the effects of SSL at the 12-month follow-up both for HFG and LFG may be of interest in the light of studies on preventive anxiety and depression programs in school-aged children that have found few or no effects in the follow-up period (Ahlen, Lenhard, & Ghaderi, 2019; Johnstone et al., 2018).

This study has some limitations that should be considered. First, the sample size is small and the sample is located in a specific geographical area of Spain. Future RCTs should overcome this limitation to ensure the generalizability of the results. Second, due to the young age of the children and the lack of reliable self-reports for younger children at the time of the study, parent-report measures were used to assess the child's psychological problems. However, future studies with young children should report results on SSL including reports from other informants if possible (e.g., self-reports, teacher reports). Furthermore, following previous studies (Escribano et al., 2015, 2016), IF was measured using self-report by children and facilitators, but other evaluation methods (e.g., interviews with facilitators, observation) could also be included in the future. Third, this study focuses on three fidelity dimensions, but IF should be further investigated (e.g., by analyzing other dimensions and factors that hinder or facilitate the implementation), examining their influence on SSL outcomes. A new study should be conducted in which fidelity is controlled from the beginning, implementing one of the arms with *real* low fidelity, and another with high fidelity, together with a control

group, unlike other studies in which IF is considered and analyzed after conducting a broader study.

5. Lessons learned

Overall, the results of this study indicate that children who receive the SSL program with high IF benefit more from the intervention at short and long term (12-month follow-up), including improvements in the key symptoms (i.e., anxiety and depression). These children showed significant improvement in a greater number of outcomes on all assessments than the LFG when compared to the CG. Furthermore, findings of this research reveal that participants may benefit similarly from the intervention even if they do not receive it with maximum fidelity, especially at long term, where children in either the high or low fidelity group improved symptoms of anxiety and depression when compared with the CG.

Thus, this research provides valuable data, as it is the first school-based cluster-randomized controlled trial analyzing the influence of fidelity on SSL outcomes in young schoolchildren exhibiting internalizing symptoms, while there is still a lack of studies of these characteristics concerning transdiagnostic prevention programs for childhood anxiety and depression. Finally, our study emphasizes the importance of evaluating and examining the IF of SSL in future studies, as it may affect its short- and long-term effects.

Figure captions

Fig. 1. Depression symptoms (means) for each time period across conditions.

Fig. 2. Anxiety symptoms (means) for each time period across conditions.

Acknowledgments

The authors wish to thank all study participants, including the schools, the facilitators, the parents, and the children.

Funding sources

This work was supported by the Ministry of Economy and Competitiveness (MINECO) of Spain [grant number PSI2014-56446-P]; and the Ministry of Education, Culture and Sport of Spain [grant number FPU14/03900]. The funding bodies had no further involvement in the research process.

References

- Ahlen, J., Lenhard, F., & Ghaderi, A. (2019). Long-term outcome of a cluster-randomized universal prevention trial targeting anxiety and depression in school children. *Behavior Therapy, 50*, 200–213.
<https://doi.org/10.1016/j.beth.2018.05.003>.
- Angold, A., Costello, E. J., Messer, S. C., Pickles, A., Winder, F., & Silver, D. (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*, 237–249.
- Bayer, J. K., & Beatson, R. (2013). Early intervention and prevention of anxiety and depression. In R. E. Tremblay, M. Boivin, RDeV. Peters (Eds.), R. M. Rapee (topic ed.), *Encyclopedia on Early Childhood Development*. Retrieved from <http://www.child-encyclopedia.com/anxiety-and-depression/according-experts/early-intervention-and-prevention-anxiety-and-depression>
- Bernaras, E., Jaureguizar, J., & Garaigordobil, M. (2019). Child and adolescent depression: A review of theories, evaluation instruments, prevention programs, and treatments. *Frontiers in Psychology, 10*, 543.
<https://doi.org/10.3389/fpsyg.2019.00543>.

- Bettencourt, A. F., Gross, D., & Breitenstein, S. (2019). Evaluating implementation fidelity of a school-based parenting program for low-income families. *The Journal of School Nursing, 35*, 325–336.
<https://doi.org/10.1177/1059840518786995>
- Biel, C. H., Buzhardt, J., Brown, J. A., Romano, M. K., Lorio, C. M., Windsor, K. S.,... Goldstein, H. (2020). Language interventions taught to caregivers in homes and classrooms: A review of intervention and implementation fidelity. *Early Childhood Research Quarterly, 50*, 140–156.
<https://doi.org/10.1016/j.ecresq.2018.12.002>
- 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*, 164-173. <https://doi.org/10.1002/nur.20373>.
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implement Science, 2*, 40.
<https://doi.org/10.1186/1748-5908-2-40>.
- Clarke, A. M., Bunting, B., & Barry, M. M. (2014). Evaluating the implementation of a school-based emotional well-being programme: a cluster randomized controlled trial of Zippy's Friends for children in disadvantaged primary schools. *Health Education Research, 29*, 786-798. <https://doi.org/10.1093/her/cyu047>
- Cohen, J. (1988). *Statistical power analyses for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Cook, C. R., Lyon, A. R., Locke, J., Waltz, T., & Powell, B. J. (2019). Adapting a Compilation of implementation strategies to advance school-based implementation research and practice. *Prevention Science, 20*, 914–935.
<https://doi.org/10.1007/s11121-019-01017-1>

- Dix, K. L., Slee, P. T., Lawson, M. J., & Keeves, J. P. (2012). Implementation quality of whole-school mental health promotion and students' academic performance. *Child and Adolescent Mental Health, 17*, 45–51. <https://doi.org/10.1111/j.1475-3588.2011.00608.x>
- Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health Education Research, 18*, 237–256. <https://doi.org/10.1093/her/18.2.237>
- Escribano, S., Espada, J. P., Orgilés, M., & Morales, A. (2016). Implementation fidelity for promoting the effectiveness of an adolescent sexual health program. *Evaluation and Program Planning, 59*, 81–87. <https://doi.org/10.1016/j.evalprogplan.2016.08.008>
- Escribano, S., Morales, A., Orgilés, M., & Espada, J. P. (2015). Influence of fidelity of implementation in the efficacy of a program to promote sexual health with adolescents. *Health and Addictions, 15*, 103–113.
- Essau, C.A., & Ollendick, T. H. (2013). *The Super Skills for Life programme*. University of Roehampton, London, UK.
- Essau, C. A., Olaya, B., Sasagawa, S., Pithia, J., Bray, D., & Ollendick, T. H. (2014). Integrating video-feedback and cognitive preparation, social skills training and behavioural activation the treatment of childhood anxiety. *Journal of Affective Disorders, 167*, 261–267. <https://doi.org/10.1016/j.jad.2014.05.056>.
- Essau, C. A., Sasagawa, S., Jones, G., Fernandes, B., & Ollendick, T. H. (2019). Evaluating the real- world effectiveness of a cognitive behavior therapy-based transdiagnostic programme for emotional problems in children in a regular

school setting. *Journal of Affective Disorders*, 253, 357–365.

<https://doi.org/10.1016/j.jad.2019.04.036>.

Fernández-Martínez, I., Morales, A., Espada, J. P., & Orgilés, M. (2020). Psychometric properties of the items and factorial structure of the Spanish version of the Parent-Report Mood and Feelings Questionnaire (MFQ-P). *Clinical Child Psychology and Psychiatry*, 25, 533–544.

<https://doi.org/10.1177/1359104519897939>.

Ghandour, R. M., Sherman, L. J., Vladutiu, C. J., Ali, M. M., Lynch, S. E., Bitsko, R. H., & Blumberg, S. J. (2019). Prevalence and treatment of depression, anxiety, and conduct problems in US Children. *The Journal of Pediatrics*, 206, 256–267.e3. <https://doi.org/10.1016/j.jpeds.2018.09.021>.

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40, 1337–1345. <https://doi.org/10.1097/00004583-200111000-00015>

Hasson, H. (2010). Systematic evaluation of implementation fidelity of complex interventions in health and social care. *Implementation Science*, 5, 67.

<https://doi.org/10.1186/1748-5908-5-67>

Hill, H. C., & Erickson, A. (2019). Using implementation fidelity to aid in interpreting program impacts: A brief review. *Educational Researcher*, 48, 590–598.

<https://doi.org/10.3102/0013189X19891436>.

Johnstone, K. M., Kemps, E., & Chen, J. (2018). A meta-analysis of universal school-based prevention programs for anxiety and depression in children. *Clinical Child and Family Psychology Review*, 21, 466–481. <https://doi.org/10.1007/s10567-018-0266-5>.

- Lewis, A. J., Sae-Koew, J. H., Toumbourou, J. W., & Rowland, B. (2020). Gender differences in trajectories of depressive symptoms across childhood and adolescence: A multi-group growth mixture model. *Journal of Affective Disorders, 260*, 463–472. <https://doi.org/10.1016/j.jad.2019.09.027>.
- Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*(1), 13–22. <https://doi.org/10.1093/biomet/73.1.13>
- Martinsen, K. D., Kendall, P. C., Stark, K., & Neumer, S. P. (2016). Prevention of anxiety and depression in children: Acceptability and feasibility of the transdiagnostic EMOTION program. *Cognitive and Behavioral Practice, 23*, 1–13. <https://doi.org/10.1016/j.cbpra.2014.06.005>.
- Nauta, M. H., Scholing, A., Rapee, R. M., Abbott, M., Spence, S. H., & Waters, A. (2004). A parent report measure of children's anxiety. *Behaviour Research and Therapy, 42*, 813–839. [https://doi.org/10.1016/S0005-7967\(03\)00200-6](https://doi.org/10.1016/S0005-7967(03)00200-6).
- Nelson, P. M., Pulles, S. M., Parker, D. C., & Kluft, J. (2020). Implementation fidelity for math intervention: Basic quality ratings to supplement adherence. *School Psychology, 35*, 72–79. <https://doi.org/10.1037/spq0000338>
- Orgilés, M., Fernández-Martínez, I., Espada, J. P., & Morales, A. (2019). Spanish version of Super Skills for Life: short- and long-term impact of a transdiagnostic prevention protocol targeting childhood anxiety and depression. *Anxiety, Stress & Coping, 32*, 694–710. <https://doi.org/10.1080/10615806.2019.1645836>.
- 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*, 776–790. <https://doi.org/10.1177/1359104519835579>.

- Paulus, F. W., Ohmann, S., & Popow, C. (2016). Practitioner Review: School-based interventions in child mental health. *Journal of Child Psychology and Psychiatry*, *57*, 1337–1359. <https://doi.org/10.1111/jcpp.12584>.
- Rasmussen, L. M. P., Patras, J., Neumer, S. P., Adolfsen, F., Martinsen, K. D., Holen, S.,... Martinussen, M. (2019). Facilitators and Barriers to the Implementation of EMOTION: An Indicated Intervention for Young Schoolchildren. *Scandinavian Journal of Educational Research*.
<https://doi.org/10.1080/00313831.2019.1596976>
- Rodríguez-Hernández, P. J., Betancort, M., Ramírez-Santana, G. M., García, R., Sanz-Álvarez, E. J., & De las Cuevas-Castresana, C. (2012). Psychometric properties of the parent and teacher versions of the Strength and Difficulties Questionnaire (SDQ) in a Spanish sample. *International Journal of Clinical and Health Psychology*, *12*, 265–279.
- Rojas-Andrade, R., & Bahamondes, L.L. (2019). Is implementation fidelity important? A systematic review on school-based mental health programs. *Contemporary School Psychology*, *23*, 339–350. <https://doi.org/10.1007/s40688-018-0175-0>.
- Rosenthal, R., & Rosnow, R. L. (1985). *Contrast analysis: Focused comparisons in the analysis of variance*. CUP Archive.
- Weare, K., & Nind, M. (2011). Mental health promotion and problem prevention in schools: what does the evidence say? *Health Promotion International*, *26*(S1), 29–69. <https://doi.org/10.1093/heapro/dar075>.
- Werner-Seidler, A., Perry, Y., Calear, A.L. Newby, J. M., & Christensen, H. (2017). School-based depression and anxiety prevention programs for young people: A systematic review and meta-analysis. *Clinical Psychology Review*, *51*, 30–47. <https://doi.org/10.1016/j.cpr.2016.10.005>

Wichstrøm, L., Belsky, J., & Steinsbekk, S. (2017). Homotypic and heterotypic continuity of symptoms of psychiatric disorders from age 4 to 10 years: a dynamic panel model. *Journal of Child Psychology and Psychiatry*, 58, 1239–1247. <https://doi.org/10.1111/jcpp.12754>.

Table 1

Sociodemographic characteristics and parent-reported measures of their children at baseline by intervention condition

Characteristics	HFG (n = 41)	LFG (n = 26)	CG (n = 56)	Total (n = 123)	p	Post Hoc	Effect size
Sociodemographics							
Parents							
Female, N (%)	32 (78)	22 (84.6)	43 (76.8)	97 (78.9)	.71	-	-
Mean age (SD), years	43.14 (5.45)	42.35 (4.67)	39.76 (4)	41.51 (4.91)	.004	CG < HFG (p = .007)	.72
Family situation					.49	-	-
Married	33 (80.5)	24 (92.3)	49 (87.5)	106 (86.2)			
Separated or divorced	7 (17.1)	2 (7.7)	7 (12.5)	16 (13)			
Single	1 (2.4)	0 (0)	0 (0)	1 (0.8)			
Education							
Primary education	10 (24.4)	6 (23.1)	7 (12.5)	23 (18.7)	.009	-	0.48
Secondary education	5 (12.2)	7 (26.9)	26 (46.4)	38 (30.9)			
Higher education	26 (63.4)	13 (50)	23 (41.1)	62 (50.4)			
Socioeconomic level (euros)					.34	-	-
0-499	1 (2.9)	1 (4.2)	0 (0)	2 (1.9)			
500-999	2 (5.7)	2 (8.3)	3 (6.7)	7 (6.7)			
1000-1999	9 (25.7)	6 (25)	16 (35.6)	31 (29.8)			
2000-2999	8 (22.9)	10 (41.7)	15 (33.3)	33 (31.7)			
3000-4999	12 (34.3)	3 (12.5)	5 (11.1)	20 (19.2)			
5000 or more	2 (8.3)	3 (8.6)	6 (13.3)	11 (10.6)			
Children							
Female, N (%)	21 (51.2)	13 (50)	21 (37.59)	55 (44.7)	.33	-	-
Mean age (SD), years	6.83 (.77)	7.04 (.87)	6.88 (.78)	6.89 (8.89)	.56	-	-
School grade							

	1	17 (41.5)	10 (38.5)	28 (50)	55 (44.7)	.78	-	-
	2	16 (39)	9 (34.6)	17 (30.4)	42 (34.1)			
	3	8 (19.5)	7 (26.9)	11 (19.6)	26 (21.1)			
Nationality								
	Spanish	39 (95.1)	26 (100)	56 (100)	121 (98.4)	.13	-	-
	Other	2 (4.9)	0 (0)	0 (0)	2 (1.6)			
Mean number (<i>SD</i>) of siblings		.90 (.70)	1.15 (.96)	.91 (.61)	.96 (.72)	.31	-	-
Dependent variables								
Depression, <i>M (SD)</i>		12.80 (9.05)	14.46 (11.18)	9.28 (7.47)	11.55 (9.07)	.03	CG < LFG (<i>p</i> = .05)	.58
Anxiety, <i>M (SD)</i>		27.48 (10.93)	31.11 (15.61)	26.16 (10.39)	27.65 (11.89)	.21		
Externalizing problems, <i>M (SD)</i>		7.34 (4.22)	9.19 (4.07)	7.91 (3.81)	7.99 (4.03)	.18		
Internalizing problems, <i>M (SD)</i>		7.98 (3.44)	9 (4.69)	6.41 (3.22)	7.48 (3.76)	.008	CG < LFG (<i>p</i> = .01)	.37

Note. HFG = High Fidelity Group; LFG = Low Fidelity Group; CG = Control Group.

Table 2

Intervention effects on depressive and anxiety measures at posttest and 12-month follow-up by experimental condition: High Fidelity Group (HFG), Low Fidelity Group (LFG) and Control Group (CG)

Variable	Time	HFG	LFG	Control	HFG-CG				LFG-CG				HFG- LFG			
		(n = 41)	(n = 26)	group (n = 56)	Coefficient	SE	p	d	Coefficient	SE	p	d	Coefficient	SE	p	d
		Mean (SE)	Mean (SE)	Mean (SE)												
Depression	Time 2	5.98 (0.91)	6.56 (1.41)	11.94 (1.46)	-5.10 (-6.50, -3.71)	.71	< .001	.65	-4.48 (-8.69, -.27)	2.14	.03	.54	.50 (1.48, -2.40)	1.48	.73	
	Time 3	6.33 (1.48)	7.78 (1.57)	11.64 (1.27)	-4.75 (-7.28, -2.23)	1.28	< .001	.56	-3.30 (-5.75, -.85)	1.25	.008	.43	1.45 (1.37, -1.24)	1.37	.29	
Anxiety	Time 2	21.73 (1.12)	22.26 (2.63)	27.60 (2.31)	-4.86 (-6.55, -3.17)	.86	< .001	.42	-5.09 (-10.92, .70)	2.96	.08		-.35 (-5.74, 5.02)	2.74	.89	
	Time 3	19.63 (1.83)	20.39 (2.28)	27.25 (2.75)	-6.97 (-10.37, -3.55)	1.73	< .001	.44	-6.95 (-11.81, -2.09)	2.47	.005	.37	.01 (-5.44, 5.48)	2.78	.99	
Internalizing problems	Time 2	5.82 (.47)	6.04 (1.21)	6.99 (.76)	-1.73 (-2.40, -.98)	.37	< .001	.24	-1.60 (-4.18, .97)	1.31	.22		.07 (-2.08, 2.23)	1.10	.94	
	Time 3	5.28 (.75)	5.94 (1.11)	7.10 (.62)	-2.27 (-3.66, -.87)	.70	.001	.39	-1.69 (-4.06, .67)	1.20	.16		.57 (-2.35, 3.49)	1.49	.70	
Externalizing problems	Time 2	7.01 (.46)	6.32 (.48)	8.24 (.51)	-.21 (-1.04, .61)	.42	.61		-1.28 (-2.21, -.35)	.47	.007	.56	-1.04 (-1.92, -.21)	.43	.01	.25
	Time 3	6.19 (.50)	6.60 (.64)	8.04 (.46)	-1.02 (-1.81, -.23)	.40	.01	.55	-1 (-2.23, .23)	.63	.11		.02 (-.58, .63)	.31	.93	

Note. SE = Standard Error. Time 2 = Posttest; Time 3 = 12-month follow-up. Higher scores denote greater symptomatology. Each analysis was adjusted for baseline differences, gender, age, and school.

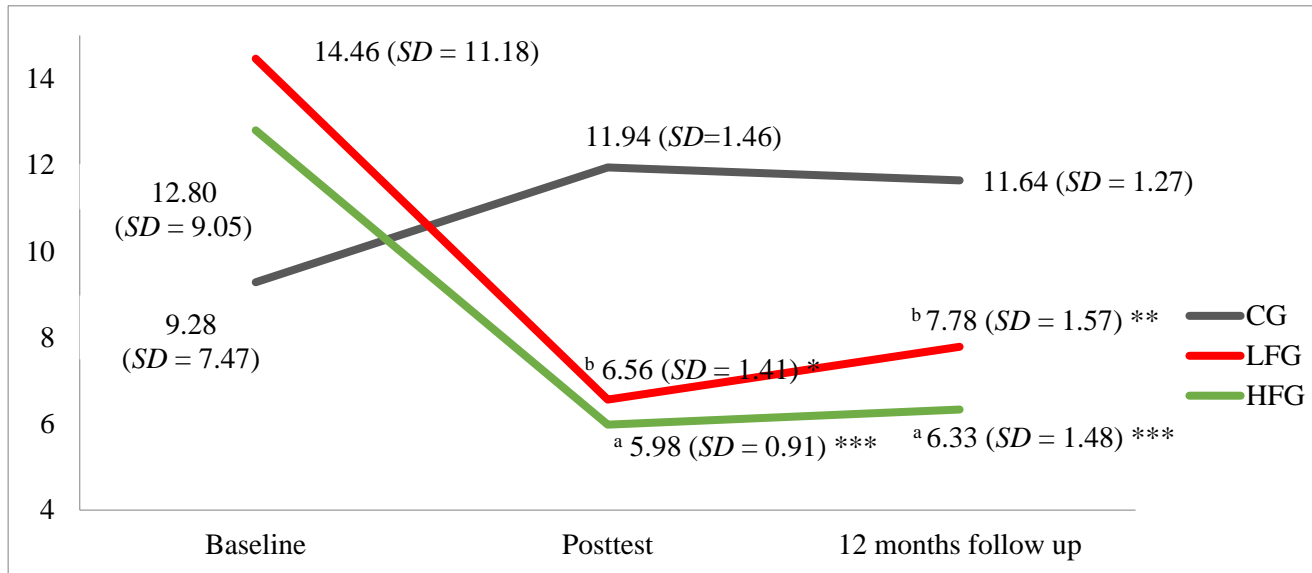


Fig. 1. Depression symptoms (means) for each time period across conditions.

Note. LFG = Low Fidelity Group; HFG = High Fidelity Group; CG = Control Group.

^a Comparison between HFG and CG.

^b Comparison between LFG and CG.

* $p < .05$; ** $p < .01$; *** $p < .001$.

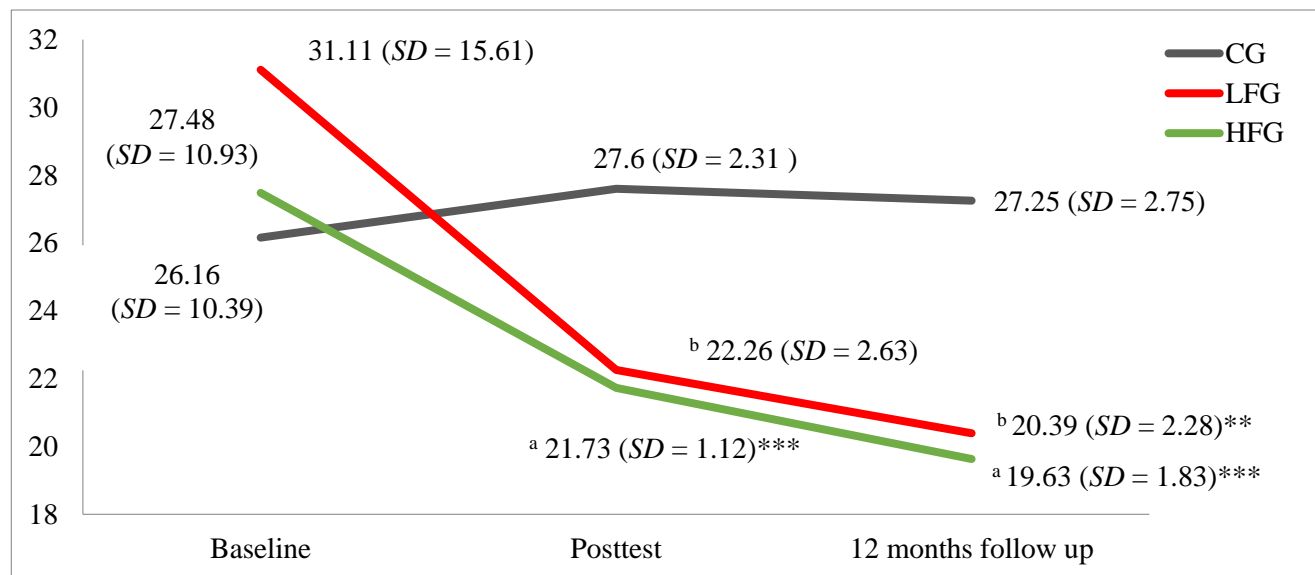


Fig. 2. Anxiety symptoms (means) for each time period across conditions.

Note. LFG = Low Fidelity Group; HFG = High Fidelity Group; CG = Control Group.

^a Comparison between HFG and CG.

^b Comparison between LFG and CG.

* $p < .05$; ** $p < .01$; *** $p < .001$.