Preventing Behavioral Disorders via Supporting Social and Emotional Competence at Preschool Age

Annika Schell*, Lucia Albers*, Rüdiger von Kries, Clemens Hillenbrand, Thomas Hennemann

SUMMARY

Background: 13–18% of all preschool children have severe behavioral problems at least transiently, sometimes with long-term adverse consequences. In this study, the social training program “Lubo aus dem All! – Vorschulalter” (Lubo from Outer Space, Preschool Version) was evaluated in a kindergarten setting.

Methods: 15 kindergartens were randomly assigned to either an intervention group or a control group, in a 2:1 ratio. The intervention was designed to strengthen emotional knowledge and regulation, the ability to take another person's point of view, communication skills, and social problem solving. The control group continued with conventional kindergarten activities. The primary endpoint was improvement in social-cognitive problem solving strategies, as assessed with the Wally Social Skills and Problem Solving Game (Wally). Secondary endpoints were improvement in prosocial behavior and reduction in problematic behavior, as assessed with the Preschool Social Behavior Questionnaire (PSBQ) and the Caregiver–Teacher Report Form (C-TRF). Data were collected before and after the intervention and also 5 months later. Mixed models were calculated with random effects to take account of the cluster design and for adjustment for confounding variables.

Results: 221 children in kindergarten, aged 5–6 years, were included in the study. Randomization was unsuccessful: the children in the intervention group performed markedly worse on the tests carried out before the intervention. Five months after the end of the intervention, the social-cognitive problem-solving strategies of the children in the intervention group had improved more than those of the children in the control group: the intergroup difference in improvement was 0.79 standard deviations of the Wally test (95% confidence interval [CI] 0.13–1.46). This effect was just as marked 5 months later (0.63, 95% CI 0.03–1.23). Prosocial behavior, as measured by the PSBQ, also improved more in the intervention group, with an intergroup difference of 0.37 standard deviations of the Wally test (95% CI 0.05–0.71).

Conclusion: An age-appropriate program to prevent behavioral disorders among kindergarten children improved both the children's knowledge of prosocial problem solving strategies and their prosocial behavior.

Cite this as:

*Annika Schell and Lucia Albers are joint first authors.
Department of Special Needs Education and Rehabilitation, Carl von Ossietzky University of Oldenburg: Dr. phil. Schell, Prof. Dr. phil. Hillenbrand
Institute of Social Paediatrics and Adolescent Medicine, Ludwig-Maximilians-Universität München: Dipl.-Math. Albers, Prof. Dr. med. von Kries
Faculty of Human Sciences, University of Cologne: Prof. Dr. phil. Hennemann

Recent study results show that as many as 13% to 18% of all preschool children show serious behavioral problems, at least for a time (1–8, e1). Among these children, those whose abnormal behavior starts at an early age (“early starters”) exhibit particularly pronounced and stable problematic behaviors (both externalizing problem behaviors, such as hyperactivity and aggressive/oppositional behavior, and internalizing behaviors, such as anxiety, unhappiness, and withdrawal). While some behavioral problems are transient, stable aggressive behavioral problems are seen in about 5% to 10% of a male birth cohort, and these have long-term negative consequences later on in life (9), such as mental disorders, dropping out of school, and delinquency (10–14).

Children with behavioral disorders more often have lower social and emotional skills. They find it difficult to regulate their emotions and have markedly fewer social problem solving strategies. They also have a lower ability to empathize and few supportive, positive friendships (15, 16). Related to this, their social-cognitive information processing is often quite inflexible, and they prefer aggressive behaviors in conflict situations (17). Meta-analyses have shown that social training programs can improve social and emotional skills and reduce behavioral problems in preschool children (18–21).

So far, however, most such programs have been developed and evaluated for use in elementary and secondary school children (22).

Few training programs exist for preschool children. Often they are adaptations of programs aimed at school-age children, which, because they generally have a strong cognitive element, do not always appear age-appropriate for preschool children (23, 24). In addition, there has been very little evaluation of them in high-quality studies.

It was for these reasons that the “Lubo from Outer Space!” program (25) was developed, on the basis of concepts introduced by Durlak and colleagues, to provide an age-appropriate, structured training program to be delivered by preschool teachers (26). We investigated the effectiveness of this program in a study initially designed as a cluster randomized study.
Methods

Study design

All of the roughly 500 kindergartens in the German city of Cologne were potential participants in the study. Based on an index of socioeconomic and educational status, the various neighborhoods of the city were assessed according to the social and educational status of children and young persons. A total of 100 kindergartens were randomized. The randomization process was based on the social/educational index. These kindergartens, situated in catchment areas where there was a low, intermediate, or high use of child and young persons’ care services/educational disadvantage, were sent written information about the study and asked to participate (27). Those that were willing to take part were allocated to the intervention or the control group in a 2:1 ratio, using the following randomization procedure: the kindergartens were numbered consecutively, and these numbers were randomly sampled to make up an intervention group of 10 kindergartens and a control group of five.

In Germany, children attend a kindergarten for about 3–6 years before they start school. Attending kindergarten is not obligatory, but most children do it for 2 to 3 years. The way kindergarten costs are regulated varies greatly between federal states.

Data collection relating to the effectiveness parameters was carried out shortly before the start of the intervention, shortly after it ended, and 5 months after it ended, just before the summer holidays in 2007.

How the sample size was calculated is described in the supplementary eMethods.

The “Lubo from Outer Space!” intervention in preschool children

The program is designed to promote knowledge, expression, and regulation of emotions, to teach social conflict and problem solving strategies, and to promote children’s abilities of perspective taking and building friendships. Theoretically, it is based on the model of social-cognitive information processing (SIP) (28, 29). It is made up of 34 clearly structured lessons, provided in the form of a manual. Each lesson lasts for 35–40 minutes. The lessons are carried out by a teacher under the guidance of a Lubo project worker over a period of about 12 weeks, three times a week, in small groups of 9–14 preschool children. The social training program is embedded in a story about “extraterrestrial Lubo” (a hand puppet), who travels to Earth to learn about feelings, friendship, and getting on together. In the process, “Lubo” keeps running into social problems, and the children help him—and thus also each other—to solve them. Age-appropriate methods such as cooperative games, role play, discussions, picture cards, creative methods, and a feedback system support the learning process in a stimulating way. To foster generalization of the learning contents, parents and kindergarten teachers were encouraged to use various rituals such as “Lubo’s problem solving circle”—a strategy for thinking about social problem solving in small steps—in everyday life. In addition, parents and teachers received monthly information letters explaining the key concepts to be taught the next month and containing suggestions on how to incorporate these in daily practice at home or at kindergarten. In the present study, the children in the control group followed the normal activities program for the kindergarten they attended. At the end of the study, these kindergartens and the parents involved were offered the opportunity to receive detailed information about the children’s social and emotional developmental process over the period of the study and use it for their own documentation of the children’s education.

Measuring instruments

The primary outcome parameter was the whole range of items of the Wally Problem Solving Test (Wally) (24, 30, 31). In this test, 13 illustrations of typical social problem situations are presented one by one to the children, who are asked for ideas about how to solve them. The responses are then assessed to determine if the child’s strategy was positive (16 potential categories) or negative (17 potential categories). The total score is calculated as the difference between the number of positive and the number of negative problem solving strategies. The entire test is freely available on the home page of its author, C. Webster-Stratton (http://incredibleyears.com/for-researchers/measures/). Aiming to record the effects of the intervention on everyday behavior in the kindergarten, we also used, as secondary outcome parameters, the German versions of the Preschool Social Behavior Questionnaire (PSBQ; [32, 33]) (scales: total problem behavior; externalizing disorder; prosocial behavior) and the Caregiver–Teacher Report Form (C-TRF; [34]) (scales: total problem behavior; externalizing disorder; internalizing disorder). These are standardized screening instruments which the teachers used to assess the children’s behavior.

Further details about these instruments may be found in the eMethods section.

Evaluation of possible confounding factors

Children’s intelligence quotients (IQ) were estimated using the Culture Fair Test 1 (CFT1) of Cattell, Weiß, and Osterland (35). Other potential confounding factors were assessed using some items in the C-TRF test. However, these questions were answered by the teachers solely on the basis of their own knowledge, not on the basis of any specific diagnostic testing or by asking the parents. Children were classified as at risk of behavioral problems if at the first assessment they were in the lowest 30% for social skills (based on the Prosocial Disorder Scale of the PSBQ) (33).

Statistical analysis

Descriptive statistics for demographic and socioeconomic variables and the baseline test values were calculated for the intervention group and the control group; differences between the groups were assessed...
using $\chi^2$ and $t$ statistics. To make the results of the various tests comparable, Z-scores were calculated for the pre–post (pre–follow-up) test results; in each case, the pooled standard deviation was used (eTable 1, 2). The effects of the intervention were evaluated in mixed linear regression models with pre–post Z-scores as outcome variable, group as independent variable, random effects for kindergartens (to take into account the study’s cluster structure), and stepwise adjustment for possible confounding factors (sex, age, disease/disability). In the final step, adjustment was made for the differing baseline test values in the two groups which had occurred due to the failure of the randomization process. The models’ beta coefficients can be interpreted as showing additional improvement in the intervention group (in standard deviations) compared to the spontaneous improvement seen in the control group. For the Wally test and the Prosocial Disorder Scale of the PSBQ, a positive coefficient means an improvement, whereas for the other test scales negative coefficients imply improvements.

**Results**

A total of 15 out of 100 randomly selected kindergartens in Cologne agreed to participate in the study. The parents of all 221 children in the 15 kindergartens gave their written consent. Ten kindergartens (126 children) were assigned to the intervention group and 5 kindergartens (95 children) to the control group (Figure). The loss of study participants up to the post-intervention test was small (n = 6, due to absence on the test day). Data for the follow-up observation period were lacking for 34 children: the questionnaires from one kindergarten were lost in the mail (n = 14), 6

---

**Deutsches Ärzteblatt International | Dtsch Arztebl Int 2015; 112: 647–54**
children had moved away, and 23 children were absent on the day of testing (Wally) (for these children, the teachers’ tests were also excluded from the analysis).

The randomization process failed. There are statistically significant differences between the groups in the baseline data for the primary and secondary measuring instruments. Differences between the groups in the distribution of demographic and socioeconomic factors and in baseline test values are shown in Table 1: the control group contained less children with a disease or disability, and overall this group showed better values for all outcome parameters except for prosocial behavior (PSBQ). These variables were included in the main analysis as possible confounding factors. The differences in mean cluster size between the groups are due to the fact that two kindergartens in the control group had a great many children.

The effects after the interventions are shown in Table 2. For the Wally test, clear effects of about one standard deviation can be seen—both in the univariate analysis and after adjusting for confounders. With regard to secondary outcomes, the effect on externalizing behavior disappears after adjusting for the baseline values.

Long-term effects after an observation period of 5 months are shown in Table 3. In the Wally test, the effects in the intervention group were as strong as they were immediately following the intervention. In addition, for prosocial behavior (measured with the PSBQ), effects of 0.37 standard deviations were observed (confidence interval did not include zero). The effect sizes for the scales Total Problem Behavior (PSBQ) and Internalizing Disorder (C-TRF) appeared as negative effects after adjusting for baseline values.

The intracluster correlation coefficients are relatively small (≤0.237) for all outcomes. For further details, see eTable 3.

There were no significant differences between at-risk children and children with normal social skills (effect modification testing; data not shown). The

<table>
<thead>
<tr>
<th>TABLE 1</th>
</tr>
</thead>
</table>

**Distribution of demographic and socioeconomic factors and baseline values for outcome variables in the intervention group and the control group**

<table>
<thead>
<tr>
<th></th>
<th>Intervention group</th>
<th>Control group</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 126</td>
<td>n = 95</td>
<td></td>
</tr>
<tr>
<td>Sex (female) % (n)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age in years M (SD)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intelligence quotient M (SD)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disease or disability % (n)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Special support % (n)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preschool teacher knows child (very) well % (n)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At-risk child at baseline % (n)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parents’ employment % (n)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wally at baseline M (SD)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSBQ at baseline M (SD)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C-TRF at baseline M (SD)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

PSBQ, Preschool Social Behavior Questionnaire; C-TRF, Caregive–Teacher Report Form; M, mean; SD, standard deviation; p-values <0.05, indicating a significant difference, are shown in bold.
intervention seemed to be equally effective in children of both sexes (similar effect sizes after the intervention; data not shown).

Discussion

Both immediately after the intervention and 5 months later, marked improvements in prosocial problem solving strategies were detected in the children who had been through the training program. This shows that the children had significantly more positive strategies available to them in dealing with problematic social situations than did the children in the control group. Interestingly, in the teachers’ assessments, the improvement in prosocial behavior in the intervention group was not visible until the 5-month follow-up. The effects were comparable in children at increased risk of behavioral disorders and those who were not. These positive results have been replicated very recently in studies in Troisdorf and Oldenburg (both in Germany), although the data have not yet been published.

In comparison with the calculated effect sizes in international meta-analyses on social skills training, the proximal effects (children’s social problem solving skills) of "Lubo from Outer Space!" were higher (26, 36, 37). Similarly, the effect sizes in terms of improved prosocial behavior appear to be as high or somewhat higher (36, 38). Lösel and Beelmann point out that the effects are often greater in smaller studies (<150), and for this reason assume the existence of a publication bias (20). Bearing in mind that almost 90% of studies included in the meta-analysis are small studies, the effect sizes reported in the present study are remarkable. In agreement with other studies on social skills training programs, the effects on the distal outcome parameters (children’s behavior) have proved to be smaller than those on the proximal parameters (social problem solving strategies) (20, 21, 38). The more positive results seen with “Lubo from Outer Space!” could therefore be due to its age-appropriate design, using activating and affective methods (26).

It may be assumed that the delayed improvement in prosocial behavior in the intervention group 5 months after the end of the program represents a “sleeper effect” (21). What this means is that children need to integrate newly learned behaviors into their behavior system and practice them before they are perceptible to others. In the same way, a deterioration in prosocial behavior in the period before starting elementary school, as observed in the control group, is not uncommon (21), so it may be assumed that the intervention prevented a deterioration of this nature.

Strengths and limitations of the study

At present few high-quality randomized controlled trials exist in children of preschool age (20), and outside the USA there are almost none. It is unarguable that more such studies are needed, since programs and results cannot be uncritically extrapolated from one country to another (20, 40). Moreover, many previous studies have methodological weaknesses in the evaluation of effects and data analysis (39). The present study is also one of the few with at least a relatively prolonged follow-up period (5 months). Another strength can be seen in its high external validity, since the intervention took place under everyday conditions in the kindergartens rather than in an experimental setting.

Although the results of the “Lubo from Outer Space!” program were mainly positive, it must be stressed that the failure of randomization and the lack

---

**TABLE 2**

Results immediately following the intervention, given as beta coefficients from mixed models with randomization effects for clusters and stepwise adjustment for confounding factors.

<table>
<thead>
<tr>
<th></th>
<th>Wally</th>
<th>PSBQ</th>
<th>C-TRF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Effect</td>
<td>Effect</td>
<td>Effect</td>
</tr>
<tr>
<td></td>
<td>[95% CI]</td>
<td>[95% CI]</td>
<td>[95% CI]</td>
</tr>
<tr>
<td>All items</td>
<td>0.98</td>
<td>–0.19</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>[0.28; 1.67]</td>
<td>[–0.73; 0.35]</td>
<td>[–0.41; 0.01]</td>
</tr>
<tr>
<td>+ adjusted for sex</td>
<td>0.98</td>
<td>–0.19</td>
<td>–0.21</td>
</tr>
<tr>
<td></td>
<td>[0.29; 1.68]</td>
<td>[–0.73; 0.35]</td>
<td>[–0.41; 0.00]</td>
</tr>
<tr>
<td>+ adjusted for age (in years)</td>
<td>1.00</td>
<td>–0.19</td>
<td>–0.20</td>
</tr>
<tr>
<td></td>
<td>[0.32; 1.68]</td>
<td>[–0.72; 0.35]</td>
<td>[–0.39; –0.02]</td>
</tr>
<tr>
<td>+ adjusted for disease/disability</td>
<td>1.03</td>
<td>–0.17</td>
<td>–0.23</td>
</tr>
<tr>
<td></td>
<td>[0.30; 1.77]</td>
<td>[–0.70; 0.36]</td>
<td>[–0.41; –0.04]</td>
</tr>
<tr>
<td>+ adjusted for baseline test value</td>
<td>0.79</td>
<td>–0.31</td>
<td>–0.06</td>
</tr>
<tr>
<td></td>
<td>[0.13; 1.46]</td>
<td>[–0.92; 0.30]</td>
<td>[–0.26; 0.15]</td>
</tr>
</tbody>
</table>
TABLE 3

Results at 5 months after intervention, given as beta coefficients from mixed models with randomization effects for clusters and stepwise adjustment for confounding factors

<table>
<thead>
<tr>
<th>Wally</th>
<th>PSBQ</th>
<th>C-TRF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All items</td>
<td>Prosocial behavior</td>
</tr>
<tr>
<td></td>
<td>Effect [95% CI]</td>
<td>Effect [95% CI]</td>
</tr>
<tr>
<td>Univariate analysis</td>
<td>0.90 [0.30; 1.49]</td>
<td>0.58 [0.15; 1.01]</td>
</tr>
<tr>
<td>+ adjusted for sex</td>
<td>0.89 [0.30; 1.49]</td>
<td>0.54 [0.13; 0.96]</td>
</tr>
<tr>
<td>+ adjusted for age (in years)</td>
<td>0.30 [0.34; 1.46]</td>
<td>0.55 [0.10; 0.96]</td>
</tr>
<tr>
<td>+ adjusted for disease/disability</td>
<td>0.91 [0.34; 1.49]</td>
<td>0.60 [0.18; 1.02]</td>
</tr>
<tr>
<td>+ adjusted for baseline test value</td>
<td>0.63 [0.03; 1.23]</td>
<td>0.37 [0.05; 0.71]</td>
</tr>
</tbody>
</table>

PSBQ: Preschool Social Behavior Questionnaire; C-TRF: Caregiver–Teacher Report Form; CI, confidence interval; *Confidence intervals were calculated on the basis of t-quantiles. Effects where the confidence interval does not include zero are shown in bold.

of blinding of the teachers in the assessment of the children do represent limitations of the study. Blinding was not possible because of the practical conditions under which the program was carried out, and this means that placebo effects on the secondary outcome parameters due to the teachers’ desire for the intervention to be effective cannot be entirely ruled out. Nevertheless, it may be assumed that placebo effects would have to occur universally, across test instruments, whereas in this study improvements were only shown in prosocial behavior. The improvement in prosocial problem solving strategies (primary outcome parameter) was assessed using a child-centered method (Wally) and so cannot have been influenced by biased answering on the part of the teachers.

Although we aimed at randomization, several baseline test values were unequally distributed between the groups. This uneven distribution reflects the fact that the test parameters are moderately to strongly correlated. The direction of shift between the intervention and the control group was such that children in the intervention group had the less favorable initial values. Although we tried to ensure that the cases and the controls were from similarly structured living environments, we did not succeed. In the intervention group there were three kindergartens where the children’s skills profile before the beginning of the study was markedly unfavorable, whereas in the control group there was a kindergarten with an above-average number of children without any problem markers. However, by including the baseline test values as a confounder in the analysis, we were able to prevent distortion of the effects. The effects in the Wally test and in prosocial behavior in the PSBQ test were somewhat weaker. For the follow-up effects of the Total Problem Behavior scale of the PSBQ test and the Internalizing Disorder Scale of the C-TRF test, the effects were reversed by qualitative confounding—that is to say, the reduced effect was more marked in the intervention group than in the control group, and in this case the main reason was taken to be the higher baseline values, so that the effect of the intervention was reversed.

The number of drop-outs at follow-up could be seen as a further limitation of this study, but the reasons for dropping out may be regarded as effect neutral (loss of questionnaires in the mail; families moving away; absence on the day of testing). It may be assumed that the fairly small number of participating kindergartens does not reduce the internal validity of the study. Only limited conclusions can be drawn about the longer-term effects of the training program, since it was not practicable to follow the children for longer as they were just about to start school. Likewise, no conclusions could be drawn about preschool children from a migrant background.

Conclusion

Our study shows that an age-appropriate social training program in kindergarten can not only improve children’s understanding of prosocial problem solving strategies, but also their prosocial behavior. The social training program “Lubo from Outer Space!” (25) can thus be seen as an effective preventive measure in the preschool setting for promoting social and emotional skills and preventing behavioral disorders before they become ingrained.

Registration

This study is registered with the German Register of Clinical Studies (Deutsches Register Klinischer Studien) under the identification number DRKS00007173.
Acknowledgment
We would like to thank all participating children, parents, and teachers for the opportunity to carry out this study.

Conflict of interest statement
Annika Schell, Clemens Hillenbrand, and Thomas Hennemann receive royalties from the sale of the manual published by Reinhardt Verlag.
Rüdiger von Kries and Lucia Albers declare that no conflict of interest exists.

Manuscript received on 26 January 2015, revised version accepted on 1 June 2015.

REFERENCES

KEY MESSAGES
- Behavioral problems such as hyperactivity, aggression, or anxiety are common in preschool children and can have negative consequences in the longer term.
- The preschool age group is particularly well suited to targeted preventive and interventional measures because of the developmental tasks associated with this age group, and because of their greater behavioral flexibility.
- To date there have been very few high-quality studies carried out in this age group to show that targeted skills training is effective.
- The present study is one of the first evaluations of a preschool program designed to promote children’s social and emotional skills.
- In the intervention group, at 5-month follow-up social problem solving had improved by 0.63 standard deviations (Wally test) and prosocial behavior by 0.37 standard deviations (Preschool Social Behavior Questionnaire, PSBQ).

ACKNOWLEDGMENTS
We would like to thank all participating children, parents, and teachers for the opportunity to carry out this study.

CONFLICT OF INTEREST STATEMENT
Annika Schell, Clemens Hillenbrand, and Thomas Hennemann receive royalties from the sale of the manual published by Reinhardt Verlag.
Rüdiger von Kries and Lucia Albers declare that no conflict of interest exists.

Manuscript received on 26 January 2015, revised version accepted on 1 June 2015.

REFERENCES

KEY MESSAGES
- Behavioral problems such as hyperactivity, aggression, or anxiety are common in preschool children and can have negative consequences in the longer term.
- The preschool age group is particularly well suited to targeted preventive and interventional measures because of the developmental tasks associated with this age group, and because of their greater behavioral flexibility.
- To date there have been very few high-quality studies carried out in this age group to show that targeted skills training is effective.
- The present study is one of the first evaluations of a preschool program designed to promote children’s social and emotional skills.
- In the intervention group, at 5-month follow-up social problem solving had improved by 0.63 standard deviations (Wally test) and prosocial behavior by 0.37 standard deviations (Preschool Social Behavior Questionnaire, PSBQ).
Supplementary material to:

**Preventing Behavioral Disorders Via Supporting Social and Emotional Competence at Preschool Age**

by Annika Schell, Lucia Albers, Rüdiger von Kries, Clemens Hillenbrand, and Thomas Hennemann

Dtsch Arztebl Int 2015; 112: 647–54. DOI: 10.3238/arztebl.2015.0647

**REFERENCES**

Supplementary material

**eMETHODS**

**Sample size calculation**
The sample size calculation was carried out for the primary outcome, which was measured with the Wally test. In a pilot study of five kindergartens, an intracluster correlation coefficient of 0.246 was calculated. With a cluster size of 14 children, a significance level of 5%, and a power of 80%, 112 children were needed to detect that problem solving skills in the intervention group had improved by four positive answers. This corresponds to 0.72 standard deviations on the Wally scale.

**The outcome tests**

**Wally**
The Wally Problem Solving Test is a child-centered individual test that measures social problem solving skills in children. It is based on the Preschool Problem Solving Test and the Child Social Problem Solving Test and was developed by Webster-Stratton and her study group. Thirteen illustrations of typical social problem situations are presented one after another to the children, who are asked for their ideas on how to solve them.

Themes are: “prohibition,” “making a mistake,” “being provoked/attacked,” “loneliness,” “being laughed at,” “disapproval and punishment by an adult.” The problem solving strategies listed in the test can be assigned to 16 positive and 17 negative categories. Examples of positive categories are: “forgiving,” “agreement with others,” “helping,” “sharing,” “compromising,” “waiting,” etc. Examples of negative categories might be: “lying,” “destroying (an object),” “hurting (a person),” “threatening (a person),” etc.

Good internal consistency was found for both positive and negative categories ($\alpha = 0.65$ and $\alpha = 0.54$, respectively), as was good construct validity for the total score (i.e., the difference in number between positive and negative problem solving strategies) of the Wally Problem Solving test ($r = 0.60$ and $r = 0.50$, respectively). The interrater reliability was calculated in a pilot study and achieved a Cohen’s kappa of 0.89.
### eTABLE 1

**Intracluster coefficients for primary and secondary outcome tests**

<table>
<thead>
<tr>
<th>Wally</th>
<th>PSBQ</th>
<th>C-TRF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prosocial behavior</td>
<td>Externalizing disorder</td>
</tr>
<tr>
<td>All items</td>
<td>0.188</td>
<td>0.137</td>
</tr>
</tbody>
</table>

PSBQ, Preschool Social Behavior Questionnaire; C-TRF, Caregiver–Teacher Report Form; ICC, intracluster correlation coefficient.

### eTABLE 2

**Mean Z-scores for the pre–post differences in primary and secondary outcome tests in the intervention and control groups**

<table>
<thead>
<tr>
<th></th>
<th>Wally</th>
<th>PSBQ</th>
<th>C-TRF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All items</td>
<td>Prosocial behavior</td>
<td>Externalizing disorder</td>
</tr>
<tr>
<td>Intervention</td>
<td>1.2 (1.17)</td>
<td>−0.17 (0.92)</td>
<td>−0.25 (0.79)</td>
</tr>
<tr>
<td>Control group</td>
<td>0.2 (0.98)</td>
<td>−0.01 (1.01)</td>
<td>−0.02 (0.46)</td>
</tr>
</tbody>
</table>

PSBQ, Preschool Social Behavior Questionnaire; C-TRF, Caregiver–Teacher Report Form; M, mean; SD, standard deviation.

### eTABLE 3

**Mean Z-scores of the pre–follow-up differences in primary and secondary outcome tests in the intervention and control groups**

<table>
<thead>
<tr>
<th></th>
<th>Wally</th>
<th>PSBQ</th>
<th>C-TRF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All items</td>
<td>Prosocial behavior</td>
<td>Externalizing disorder</td>
</tr>
<tr>
<td>Intervention</td>
<td>1.13 (1.05)</td>
<td>0.07 (1.28)</td>
<td>−0.25 (1.03)</td>
</tr>
<tr>
<td>Control group</td>
<td>0.36 (1.03)</td>
<td>−0.5 (1.63)</td>
<td>−0.06 (0.78)</td>
</tr>
</tbody>
</table>

PSBQ, Preschool Social Behavior Questionnaire; C-TRF, Caregiver–Teacher Report Form; M, mean; SD, standard deviation.