



Differential Effects of an Early Childhood Care Preventive Intervention Program on Behavior and Emotional Problems

Maryam Zarra-Nezhad¹ · Ali Moazami-Goodarzi² · Joonas Muotka³ · Markus Hess⁴ · Sari Havu-Nuutinen¹ · Herbert Scheithauer⁵

Accepted: 9 August 2023 / Published online: 21 August 2023
© The Author(s) 2023, corrected publication 2023

Abstract

Early childhood is a crucial time for developing social-emotional skills, offering the opportunity to lay the foundation for healthy development. However, early childhood behavior problems are risk factors for problems and disorders in later life phases, including depression, antisocial behavior, and substance abuse. This study uses a longitudinal design to examine the differential impact of a preventive intervention and social-emotional learning program (Papilio-3to6) on children's social-emotional development. The program was evaluated using a 1-year-randomized controlled trial design with an intervention (IG) and control group (CG) that provided data at a pretest and post-test. Teachers rated 627 children's behaviors ($M_{AGE} = 56.77$ months at pretest) from 50 daycare center groups using the Strengths and Difficulties Questionnaire (SDQ). To investigate differential effectiveness, a latent profile analysis at the pretest identified three distinct subgroups of children with different patterns in SDQ subscales: 1) Internalizers with high internalizing problems, 2) Prosocials with high prosocial behaviors, and 3) Externalizers with high externalizing problems. Latent change score analysis showed significantly reduced externalizing problems and increased prosocial behaviors in the IG compared to the CG. Further, the children categorized as Internalizers, Prosocials, and Externalizers benefited from the program regarding externalizing problems. However, no effect was found regarding internalizing problems.

Keywords Preventive intervention program · Early childhood · Social-emotional development · Behavior problems · Prosocial behaviors

Highlights

- Taking a person-centered approach, a three-class model of SDQ subscales, including Internalizers, Prosocials, and Externalizers, was chosen as the best-fitting model.
- Children in the intervention group showed a greater reduction in externalizing problems and a greater increase in prosocial behaviors compared to the control group of children.
- The children categorized as Internalizers, Prosocials, and Externalizers benefited from the program regarding externalizing problems.
- No intervention effects were found regarding internalizing problems.

✉ Maryam Zarra-Nezhad
maryam.zarra-nezhad@uef.fi

¹ School of Applied Education Science and Teacher Education, University of Eastern Finland, P.O. Box 111, FI-80101 Joensuu, Finland

² Department of Psychology and Speech-Language Pathology, University of Turku, Assistentinkatu 7, Publicum, FI-20014 Turku, Finland

³ Department of Psychology, University of Jyväskylä, PO Box 35, FI-40014 Jyväskylä, Finland

⁴ Department of Health, German University of Health and Sports, Ranklinstraße 28-29, 10587 Berlin, Germany

⁵ Department of Education and Psychology, Freie Universität Berlin, Habelschwerdter Allee 45, 14195 Berlin, Germany

Early childhood is a crucial time for the development of various social-emotional competencies. It is a unique opportunity to lay the foundation for wellbeing and mental health, positive peer relationships, problem-solving skills, as well as improving the chances of early academic performance and successful adjustment at school (e.g., Collie et al., 2019; Cristóvão et al., 2017; Denham et al., 2015) and in later phases of life into adulthood (e.g., Jones et al., 2015; Moffitt et al., 2011). In early childhood, social-emotional competencies refer to children's skills and ability to function adaptively in interpersonal relationships and express, perceive, and regulate emotions in socially and culturally appropriate ways (for a review, see Mejía, & Kliewer, 2006). Children who are socially and emotionally competent gain the skills and knowledge needed to cope with challenges, regulate their emotions, build secure and stable social relationships with peers and adults, successfully adjust to school, and can solve problems (e.g., Housman et al., 2018; Jones et al., 2015; Nix et al., 2013). For example, children with highly prosocial skills (social behaviors characterized by helping, sharing, volunteering, and cooperating) are relatively well-adjusted and have better peer relationships and academic achievements than children with low prosocial skills (e.g., Caputi et al., 2012).

Developing social-emotional competencies in early childhood not only supports children's wellbeing and positive mental health but also affects development, adjustment, and mental health in adolescence and adulthood (for a review, see Herrenkohl et al., 2010; Jones et al., 2015; Moffitt et al., 2011). For example, the acquisition of social-emotional competencies in early childhood (5–6-years-old) has been linked to academic outcomes, substance abuse, criminal activity, mental health, and employment 13 to 19 years later, suggesting that children with higher social-emotional competencies are more likely to become well-adjusted adults (Jones et al., 2015). However, early childhood social interactions are not always well developed, and children face challenges in controlling their emotions and behavior. Children who do not experience appropriate social-emotional development are more likely to develop severe forms of behavioral problems (Izard, 2002). For example, low levels of social-emotional competencies can, on the one hand, contribute to externalizing behavior problems, such as conduct problems, aggressiveness, and antisocial behavior, in which negative emotions and social actions are often directed against others (Bornstein et al., 2010). On the other hand, low social-emotional competence levels can contribute to internalizing behavior problems, such as depression and anxiety, where negative emotions are directed inwards, and social actions are rather inhibited (Bornstein et al., 2010).

Despite the apparent distinctiveness, the behavioral patterns resulting from low social-emotional competencies

often include both externalizing and internalizing proportions to a varying degree (Polier et al., 2012; Willner et al., 2016). Theories have posited that comorbidity occurs when the symptoms of one disorder increase the risk of the development of another (Caron & Rutter, 1991; see also Willner et al., 2016). The presence of externalizing symptoms may raise the risk of later internalizing symptoms, which may appear in response to adverse consequences of externalizing behaviors such as rejection by peers (e.g., Van Lier & Koot, 2010; see also Willner et al., 2016). Further, the presence of internalizing symptoms may raise the risk of the emergence of externalizing symptoms. For example, anxiety symptoms are often associated with increased aggressive behaviors in children and adolescents (e.g., Polier et al., 2012).

In the short term, deficits in social and emotional competencies may put children at high risk of negative consequences such as attention deficit hyperactivity disorder (ADHD), social withdrawal, conduct problems, and aggressiveness (for a review, see Herrenkohl et al., 2010). Another risk factor related to deficits in social-emotional competence is the occurrence of peer relationship problems. During early childhood, positive peer relationships provide opportunities for enjoyable play and help children to regulate their behavior, which can have longitudinal impacts on children's positive social-emotional development, cognitive development, adjustment, and psychological wellbeing (for a review, see Rubin et al., 2013a, 2013b). Positive peer relationships provide a sense of security, self-value, and personal validation and offer the social-emotional support needed during childhood and adolescence (for a review, see Rubin et al., 2013a, 2013b). Children who experience peer relationship problems are considered at risk of emotional problems and poor social adjustment in adolescence (e.g., Shin et al., 2016). In the long-term, children suffering from maladaptive social-emotional development may be at high risk of negative consequences such as mental illness, depression, delinquency, antisocial behavior, substance abuse, poor achievement in school, becoming school dropouts, and poor physical health (for a review, see Herrenkohl et al., 2010; Jones et al., 2015; Moffitt et al., 2011).

Promoting Social-Emotional Learning and Preventing Behavior Problems in Early Childhood

Child development is fundamentally contextual, meaning that without the specific and active support of a nurturing environment, no child would thrive (Pluess, 2015). However, individuals differ substantially in sensitivity and responsivity to the environment, with some children being more susceptible than others to the same environmental

influences (Belsky & Pluess, 2009). This highlights the need for early childhood education (ECE) and care to reduce negative environmental impacts and ensure optimal environments and experiences for young children (Silkbeumer et al., 2016).

The transition from growing up mainly in the family environment to spending a substantial amount of time in institutional daycare centers that take usually place in Germany at around three years of age (recently care for under-threes has also been expanded and more children from the age of one attend daycare centers) has been identified as a vulnerable stage for many children due to several challenges, such as entering a new social environment, spending a significant amount of time outside the family for the first time, meeting the complex demands of teachers, establishing new peer relationships, entering group play, and following directions (see Denham et al., 2012). However, many children are not yet well-equipped—emotionally, cognitively, and socially—to face these challenges and lack the social-emotional skills necessary for successful individual development in the daycare setting, which makes early social-emotional learning (SEL) essential for supporting children’s healthy development (Buscemi et al., 1996). Furthermore, early social-emotional competencies, such as attentional skills, problem-solving skills, and behavior regulation, are critical for children’s academic achievements (Cristóvão et al., 2017; Denham et al., 2012; Denham et al., 2015). Therefore, it is essential for early education and daycare providers to be involved in developing and implementing SEL programs.

According to the World Health Organization, schools are the key setting when it comes to mental health promotion, being the most cost-effective investment that a nation can make. Even though several SEL intervention programs have been developed to tackle behavior problems in early childhood (for reviews, see Bierman & Motamedi, 2015), interventions that support teachers to promote social-emotional competencies in early childhood are limited (e.g., Bierman et al., 2008; Bierman, 2002; Conners-Burrow et al., 2017; Fox et al., 2011; Han, 2014; McLaren et al., 2009) particularly in Europe (for a review, see Jensen & Rasmussen, 2019). Large scale implementation of SEL intervention programs for specific groups such as socio-economically disadvantaged children (e.g., Head Start Redi; Bierman et al., 2008), or children at risk for conduct disorder (e.g., Fast Track; Bierman, 2002), have shown that early interventions can have positive effects on socio-economically disadvantaged children. However, these studies do not tell us about benefits for children in universal daycare systems, where all children participate regardless of a risk status (see Jensen et al., 2017), as is the case in Germany with more than 90 % attendance rate among children aged 3–6 years (Statistisches Bundesamt, 2020).

Further, a relatively limited number of universal interventions evaluating different SEL programs in ECE have been undertaken in different cultural contexts (Koivula & Huttunen, 2018). In particular, there are still not many comparable European universal early childhood prevention and intervention programs, such as Papilio-3to6, that have been implemented and evaluated (for reviews, see Jensen & Rasmussen, 2019; Koivula & Huttunen, 2018; Von Klitzing et al., 2011). The advantage of universal interventions, such as the Papilio-3to6 program, compared to targeted interventions (which are only delivered to those at high risk of developing mental health problems, disorders respectively), is that they target all children, independently of their risk status, and thus avoid stigma while attempting to reduce or prevent a variety of behavior problems and promote a broad range of protective factors (Greenberg & Abenavoli, 2017). Components of the Papilio-3to6 program are implemented simultaneously and continuously in early education care facilities, and the measures developed are “easy-to-apply” and can easily be integrated into the daily pedagogical routine. In Europe (for a review, see Jensen & Rasmussen, 2019), especially in Germany, there are not many comparable, evidence-based early childhood prevention and intervention programs implemented large-scale in early education facilities (e.g., the programs EFFEKT [Lösel et al., 2006], or FAUSTLOS [Schick & Cierpka, 2003]). For example, Jensen et al. (2017) found that the VIDA intervention significantly improved children’s emotional development and caused a reduction in emotional and behavioral problems (measured by SDQ) among Danish children aged 3–5 years. Hutchings et al. (2013) also found that the IY-TCM intervention had a more significant effect on children with behavioral problems (measured by SDQ) and reduced negative behavior among children aged 3–7 years from Wales. Schick and Cierpka (2003) evaluated the German “Faustlos” curriculum for early education care and elementary schools in a 3-year control-group design study with 44 school classes. Their results prove significant positive changes in 6–9-year-old children’s emotional competencies and prosocial development, including decreased internalizing behavior.

Investigating Differential Effects of Universal Preventive Intervention Programs

Yet, it is likely that the effects of universal interventions vary due to variations in pre-existing children’s social-emotional and behavioral characteristics (Farrell et al., 2013). Evidence shows that the variation in intervention responsiveness, in fact, can be predicted by baseline social-emotional or behavioral risks (e.g., Bradshaw et al., 2015; Farrell et al., 2013; Spilt et al., 2013; Schochet et al., 2014).

Despite the important efforts in establishing main intervention effects and unraveling the proximal processes behind these effects, the question “for whom does it work?” has mainly remained unanswered (Spilt et al., 2013). Therefore, it is essential to evaluate universal interventions regarding their differential impact above and beyond their overall effectiveness.

Employing a person-centered data-driven approach, such as latent profile analysis (LPA), as an alternative to more norm-based approaches, which group persons together according to specific cut-off values regarding behavioral characteristics, appears to be a particularly innovative and appropriate approach for examining the variation in impacts of universal interventions (Bradshaw et al., 2015; Lanza & Rhoades, 2013; Schochet et al., 2014). Research has mostly investigated children’s development from a variable-centered perspective. This approach emphasizes universal explanations for developmental outcomes by describing the relations between variables and assesses development at the whole group level, leading to a gap in our understanding of young children and our ability to promote their development and prevent their behavior problems (Denham et al., 2012). Therefore, a limited focus on variable-centered approaches fails “to capture the configurations of factors that jointly explain behavior processes” (Bauer & Shananhan, 2007, p. 256).

In contrast, a person-centered approach classifies children into distinct subgroups, classes respectively, defining meaningful subgroups characterized by unique developmental pathways (Bergman & Magnusson, 1997). Such subgroups would replicate “types” of children for whom the creation of intervention and prevention programs could be more strategic (Denham et al., 2010). As it is difficult to determine if and how interventions work for all participants within a universal prevention approach, a person-centered analytical approach identifying specific subgroups based on baseline behavioral data would clarify universal interventions’ differential effectiveness (Greenberg & Abenavoli, 2017). For example, in a study by Bradshaw et al. (2015), LPA was conducted on baseline ratings of primary school children’s social-emotional functioning, prosocial behavior, behavior problems, and concentration problems, which revealed four latent classes: high-risk, at-risk, normative, and socially-emotionally skilled. Findings suggested that at-risk and high-risk children may benefit most from exposure to positive behavioral interventions and supports.

The importance of subgroup analysis is not in differential response to intervention but in identifying how to maximize the intervention benefits and alleviate risk (Rothwell, 2005). Consequently, our study addresses the growing interest in the impact of universal preventive interventions on subgroups of children with different social-emotional behaviors (Schochet et al., 2014; Supplee et al., 2013) and may

strengthen our understanding of personalized prevention and help to identify children who are most responsive to the programs (Bradshaw et al., 2015).

The Present Study

The present study aims to contribute to the field of ECE and development by using a longitudinal design to examine the overall and differential impact of a universal preventive intervention program, i.e., the *Papilio-3to6* program, on early social-emotional development (in terms of prosocial behaviors and internalizing and externalizing behavior problems). We sought to identify subgroups of children characterized by specific social-emotional behaviors and were particularly interested in understanding the impact of the *Papilio-3to6* program on each subgroup. As mentioned above, previous research mainly investigated children’s development using a variable-centered approach. Although a variable-centered analysis is helpful in mapping observed relations between the variables, information related to distinct types of individuals sharing common characteristics is often lost (see Denham et al., 2012). To gain more insight into the effectiveness of the universal intervention program, the present study used a combination of a person-centered and a variable-centered approach. Based on the research mentioned above, we were interested in the relation between baseline types of behavioral tendencies and changes in externalizing and internalizing behavior in children receiving the *Papilio-3to6* program. Determining the impact of the *Papilio-3to6* program on the studied children’s social-emotional development required understanding how social-emotional behaviors, as demonstrated in groups of individual children, would be addressed by the program and how the children would benefit from it. The *Papilio-3to6* program promotes social-emotional skills using an SEL approach and promotes positive peer relationships while remediating skill deficits and reducing children’s behavior problems. Children’s behavior, and behavior problems, respectively, were investigated in the current study using the Strength and Difficulties Questionnaire (SDQ; Klasen et al., 2003), which consists of five subscales (emotional problems, behavior problems, hyperactivity, peer relationship problems, and prosocial behaviors), representing the aimed outcomes of the *Papilio-3to6* program very well.

Using the same sample, the effectiveness of the *Papilio-3to6* program has been proven effective in earlier studies. For example, in Scheithauer et al. (2022), the black-box evaluation (effectiveness) results were published, in which the SDQ subscale scores—comparing the intervention group (IG) and waiting control group (CG, which also received the program one year after implementation in the IG) at the first and third measurement points—were used in

multilevel analysis. Additionally, teachers completed questionnaires measuring their levels of job satisfaction, self-efficacy (control variables), and program implementation. The paper also presents the results of the fidelity and process evaluations, respectively. In an earlier German publication, Scheithauer et al. (2016) reported simple mean differences between pre- and post-test on several SDQ subscales.

Consequently, the current study exceeds the previous findings using the same data. It represents a significant gain in knowledge regarding the effectiveness of the Papilio-3to6 program by identifying subgroups of children (characterized by specific social-emotional behaviors) and examining the program's impact on each subgroup. Our aims are: (1) following a person-centered approach to identify subgroups of children with different social-emotional behavior at baseline; (2) following a variable-centered process to determine the overall impact of the Papilio-3to6 program on SDQ subscales; and (3) to examine whether the program's impact varied as a function of the subgroup classification. Based on previous research on universal preventive intervention programs (e.g., Bradshaw et al., 2015; Farrell et al., 2013; Schochet et al., 2014; Spilt et al., 2013), we hypothesized that children with higher behavior problems would benefit most from the Papilio-3to6 program. Based on the research mentioned above, we were additionally interested in the relation between baseline types of behavioral tendencies and changes in externalizing and internalizing behavior resulting from participation in the Papilio-3to6 program.

Method

Participants and Procedure

The present study was carried out in Augsburg, a medium-sized city in Germany, and Augsburg County. A stratified sample in which we controlled for relevant socio-demographic variables was used for the current study. In the first step, 50 daycare center groups (25 daycare centers participated with two groups each) were approached and consented to participate in the present study. From 1,231 eligible children, those who were not available for further assessment (due to transfer to regular elementary school) were excluded, which resulted in a total sample of 796 children (3–6-year-olds) who received regular childcare from 109 daycare center teachers (100% female). The study was conducted following the ethical principles of the Finnish Advisory Board on Research Integrity and the ethical principles of the Helsinki Declaration and approved by the University of Bremen and the German Research Foundation. An informed consent form was sent to the parent(s)/

guardian(s) to sign before the intervention. The participants were recruited voluntarily, guaranteed complete anonymity, and could withdraw themselves and all their information from the study without reason. Components of the Papilio-3to6 program can be implemented as part of the daily pedagogical routine of every ECE center, and the benefits of the program are shared on an equal basis by all children. All teachers participated in the intervention during their working hours (as part of their work), and therefore it did not involve any additional work. The parents' decision to participate in the study does not affect whether children receive the SEL activity or materials for the intervention. Measurements included in the effectiveness study of the intervention were only collected from children whose parents consented.

Overall, 86% of parents consented to their children's participation in the study. Consequently, final data were available for 667 children at the pretest (before program implementation) and 646 children at the post-test (after full implementation of the program), resulting in a total analytic sample of 627 children ($M_{AGE} = 56.77$ months at pretest; 49% girls) with valid longitudinal data for pre- and post-test. Attrition was very low, and missing data analyses revealed that the overall attrition could largely be considered unsystematic (Scheithauer et al., 2022). Respective means, standard deviations, internal consistencies, and correlations for the observed variables are presented in Appendix A.

The program was evaluated using a randomized controlled trial design with an IG and CG. The daycare centers selected for the study were randomly (stratified randomization) assigned to the intervention or control group. Data were collected before the teacher training phase (pretest), during program implementation 7–8 months later, and post-test (after full implementation; one year after teacher training). The present study includes data from the pretest and post-test. As the intermediate test was realized during program implementation, 7–8 months after pretest/teacher training, and post-test after full implementation one year after teacher training, the present study focuses on the pre-post-test design to report long-term effects after the program has been—fully—implemented. In the IG, daycare center teachers received standardized Papilio-3to6 training. After stratified selection based on relevant socio-demographic variables, recruitment, and random assignment (daycare centers) to the IG and CG, 24 daycare center groups (from 12 daycare centers) belonged to the IG ($n = 301$; 48%) and 26 daycare center groups (from 13 daycare centers) belonged to the CG ($n = 326$; 52%) during the evaluation period. Both groups were comparable, as supported by small effect sizes regarding differences in relevant socio-demographic variables such as age (57.10 months in the CG vs. 56.43 months in the IG; $d = 0.09$) and gender (49% girls in the CG vs. 49% girls in

the IG; $d = 0.00$). A detailed comparison between the IG and CG is presented in Appendix A.

Teachers of the CG did not receive any treatment, training, etc., during the evaluation period. In the CG facilities, teachers continued with “care as usual”. At pre- (IG and CG) and post-test (CG only, to control for any training or continuing education experiences they may have had between the measurement occasions), teachers were asked to answer a questionnaire (self-development) about advanced training experiences regarding topics relevant to the content of the program, e.g., behavioral disorders and improvement of their pedagogical work, and their interest in the program. A comparison between daycare center teachers of the IG and CG revealed no (significant) differences in the content-related or professional reasons (example items: “because it is important to promote the social competence of children” or “because it is an opportunity to gain further professional qualifications”). Further, a comparison between IG and CG revealed no significant difference with regard to teachers’ prior experiences with workshops on behavioral problems.

Measures

The Papilio-3to6 program aimed to improve early socio-emotional competencies by enhancing the pedagogical sensitivity of early education care teachers. The present study focused on the teachers’ reports, enabling an unbiased external assessment of children’s behavior within their daily social setting. Teacher reports on preschool children’s behavior and skills (in contrast to parents’ reports) are associated with observations of children’s behavior in the classroom (e.g., Winsler & Wallace, 2002). Results on the program’s effectiveness using other/additional information sources have been reported in Scheithauer et al. (2016, 2022).

The Strength and Difficulties Questionnaire

The same daycare teachers completed the German version (Klasen et al., 2003) of the Strength and Difficulties Questionnaire (SDQ) at the pretest and post-test. The SDQ consists of 25 items and five subscales, each containing five items: emotional problems (e.g., “many fears and easily scared”), conduct problems (e.g., “often fights with other children or bullies them”), hyperactivity/inattention (e.g., “restless, overactive, and cannot stay still for long”), peer relationship problems (e.g., “rather solitary and tends to play alone”), and prosocial behaviors (e.g., “shares readily with other children”). Daycare center teachers were asked to rate each child on a 3-point rating scale (1 = not true, 2 = somewhat true, 3 = certainly true). Subscales were created by calculating the sum score of relevant items. Thus, the

total problem sum score consisted of emotional problems, conduct problems, hyperactivity, and peer relationship problems. The Cronbach’s alpha values of all subscales at both the pretest and the post-test ranged from 0.73 to 0.86. The validity of the German SDQ, the teacher version for preschoolers, has been proven and recommended as a convenient and valid screening instrument to assess preschool-age children’s positive and negative behavior (e.g., Petermann et al., 2010; Koglin et al., 2007).

There is theoretical and preliminary empirical support for using broader internalizing (i.e., emotional subscale and peer relationship problems subscale) and externalizing (i.e., conduct problems and hyperactivity subscales) SDQ subscales scores, alongside the prosocial behaviors subscale (i.e., the fifth SDQ subscale), for analyses in low-risk samples, while retaining all five SDQ subscales when screening for disorders (Goodman, Lamping, & Ploubidis, 2010). Thus, in this study, we combined the SDQ’s hypothesized subscales concerning emotional problems and peer relationship problems into an “internalizing” subscale and the hypothesized subscales on conduct problems and hyperactivity/inattention into an “externalizing” subscale (see Goodman et al., 2010). The Cronbach’s alpha values for the internalizing and externalizing subscales at both the pretest and the post-test ranged from 0.71 to 0.81. To evaluate the intervention’s effects, a change score for internalizing and externalizing subscales alongside prosocial behaviors (the fifth SDQ subscale) was calculated, i.e., the pretest score was subtracted from the post-test score so that the resulting score represented the change during the evaluation period.

Preventive Intervention Program

The teachers involved in the program were trained by Papilio-3to6 coaches in its measures and implementation. Teachers also had regular face-to-face meetings, including “update days,” to maintain the training quality and implementation fidelity. On these “update days”, teachers had the opportunity to clarify questions regarding the program implementation and discuss with each other and with the Papilio coaches how to solve problems (e.g., how to deal with a bad room situation in the institution). The Papilio-3to6 program includes three manualized and standardized intervention components that are mainly meant to promote children’s social-emotional competence, group-orientated and prosocial behaviors, and integration into their peer group, which can be applied in everyday care (see Mayer et al., 2016). The first component includes child-focused interventions in the form of a variety of activities such as a “Paula and the pixies in the box,” a problem-solving story in which children learn to recognize, perceive, and regulate their own emotions and those of others (sadness, anger, fear,

Table 1 Model fit indices for one- to five-class profiles of SDQ subscales and profile prevalence (%) from LPA

| Model | df | Scaling | AIC | BIC | aBIC | Entropy | LMR | BLRT | Profile prevalence | | | | |
|----------|-----------|--------------|----------------|----------------|----------------|--------------|--------------|--------------|--------------------|-------------|-------------|------|------|
| | | | | | | | | | 1 | 2 | 3 | 4 | 5 |
| 1 | 10 | 1.078 | 14518.83 | 14573.22 | 14535.1 | na | na | | | | | | |
| 2 | 18 | 1.373 | 9075.89 | 9157.42 | 9100.27 | 0.841 | <0.001 | <0.001 | 0.74 | 0.26 | | | |
| 3 | 26 | 1.521 | 8920.62 | 9038.39 | 8955.83 | 0.858 | 0.050 | 0.002 | 0.69 | 0.21 | 0.10 | | |
| 4 | 34 | 1.742 | 8753.65 | 8827.90 | 8907.65 | 0.876 | 0.370 | 0.233 | 0.60 | 0.20 | 0.15 | 0.05 | |
| 5 | 42 | 1.954 | 8708.88 | 8899.11 | 8765.76 | 0.876 | 0.656 | 0.550 | 0.60 | 0.18 | 0.12 | 0.06 | 0.04 |

The values in the LMR and BLRT columns are the *p* values related to LMR and BLRT in comparing fit between models

df the number of free parameters, *AIC* Akaike Information Criterion, *BIC* Bayesian Information Criterion, *aBIC* adjusted Bayesian Information Criterion, *LMR* Lo-Mendell-Rubin, *BLRT* Bootstrapped Likelihood Ratio Tests

and happiness, each represented by a pixie in the story) to foster their emotional awareness, knowledge, and competencies; b) a “Toys-go-on-holiday day” when, once a week, the children play interactive games without toys to enhance their positive social interactions and social skills; c) the “Mine-yours-yours-ours game,” in which the children learn to adhere to social rules in small groups to address and/or prevent behavior problems and to foster prosocial behaviors. The second component is a teacher-focused intervention that supports daily teacher-child interactions and constructs a supportive educational atmosphere in the day-care center group. Teachers are trained in positive interaction and communication forms with children and handling challenging and conflict situations in a constructive, positive way (e.g., avoiding punishments and focusing on positive behavior). Teachers are provided with information on topics relevant to the program’s aims, such as the early prevention of behavior problems, and general developmental psychological background information, such as normative behavior regulation, social interaction, and interpersonal communication. In addition, teachers are trained to strengthen their positive interaction skills and support children’s social-emotional development during daily interactions. The third component is a parent-focused intervention which consists of regular sessions at the day-care center. Teachers introduce parents to the educational procedures of the child-focused intervention and support them in their educational skills in the home environment. However, the third component was not yet fully developed at the time of the evaluation study and thus was not a topic for the current study and was not implemented in the IG during the term of the study.

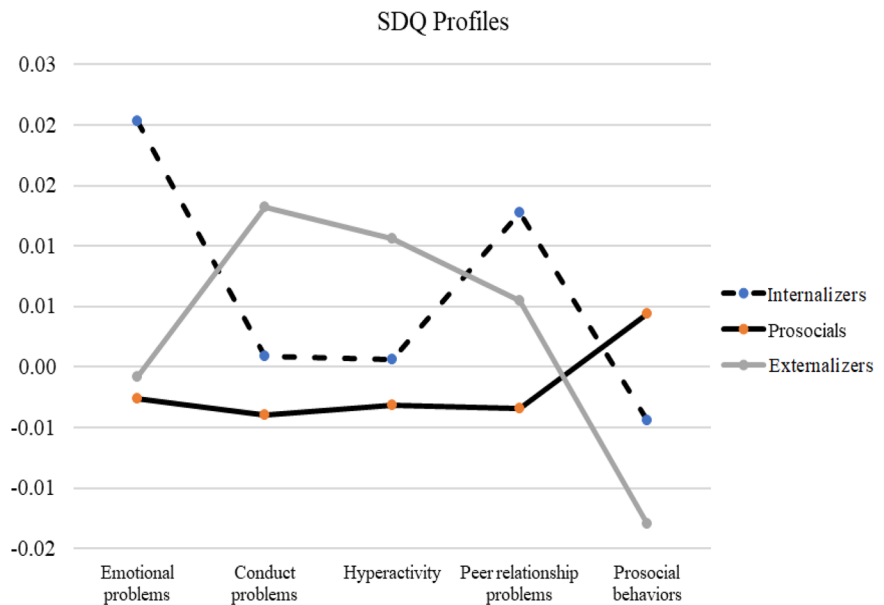
Teacher training included basic and in-depth skills training (7 + 2 days). Papilio trainers are trained educators with experience in teaching in a daycare context. As this is a qualification measure for educators, there are no further requirements to be met to participate in the program. To become an accredited trainer or Papilio teacher, all training content must be completed, including supervision, reporting, etc. The

implementation of the program, as well as the teacher training, was accompanied by measures to ensure fidelity and implementation of measures according to the manual. A detailed description of the measures implemented, results from the fidelity evaluation, and the relation to program effectiveness are described in a study by Scheithauer et al. (2022).

Data Analyses

An intent-to-treat analysis was applied as an analytic strategy, including all children assigned to the IG with pre-and post-test data independent of their true level of participation in the treatment sessions (Gupta, 2011). An intent-to-treat approach, in general, produces conservative estimations regarding intervention effects (Gupta, 2011). This procedure reflects real-life circumstances and produces rather conservative values regarding the effectiveness of interventions. The statistical analyses were performed in two steps. First, LPA was used in Mplus 8.0 to explore the most likely number of subgroups based on the children’s social-emotional behaviors at baseline (Muthén & Khoo, 1998). This analysis included all the children that were included in the pretest. The mean item scores for each SDQ subscale were calculated to ensure that the subscales were comparable and to make the latent profiles more interpretable. As shown in Table 1, the SDQ subscales were not highly inter-correlated, i.e., most of the coefficients were less than 0.70. Hence, they were suitable for use as component variables. Six model fit indices were adopted to evaluate the optimal model resulting from the LPA. These were: the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), adjusted Bayesian Information Criterion (aBIC), Lo-Mendell-Rubin (LMR), Bootstrapped Likelihood Ratio Test (BLRT), and entropy. The AIC, BIC, and aBIC are commonly used for comparing different counterpart models, with the lowest value on each indicator suggesting a best-fitting model (Carragher et al., 2009). LMR and BLRT were used to compare solutions with different numbers of classes; a low [<0.05] *p*-value indicates that the *k*-1 class

Fig. 1 Three profiles of SDQ of the best-fitting three-class pattern



model has to be rejected in favor of a model with at least k classes (Muthén, 2004). Entropy measures the quality of individuals' classification into latent classes, with values exceeding 0.80 preferred and approaching 1.0 demonstrating much clearer results (Muthén, 2004). To further examine the results' accuracy, the discriminant analyses from the LPA were reviewed to identify group assignment probabilities of being classified into one specific group. Post hoc tests were used to compare the SDQ subscales within each classification. The parameters were estimated using the full-information maximum likelihood method (MLR).

Second, latent change score (LCS) analysis using structural equation modeling (SEM) techniques with Wald test was used to analyze the group \times time interaction, i.e., whether the two study groups (IG; CG) changed differently from the pretest to post-test. LCS was estimated by using pre-and post-measurements (observed variables). In addition, the group \times time \times class membership interaction was analyzed using this method; in case of a difference, post hoc tests were conducted. Due to the hierarchical data (children were in different classes), intraclass correlations (ICCs) were computed to see between-classes variation. The complex method was used in the models to take into account the clustered sample by correcting the standard errors using a sandwich estimator, thus giving more reliable p values. Analyses were carried out with full-information maximum likelihood estimation. Estimation accounts for missing values at random (MAR) and includes all the available data. Cohen's d was adopted to measure effect size using estimated values from the pretest to the post-test within-groups and corrected between-groups. A within-group effect size of 0.5 is considered small, 0.8 medium, and 1.1 large, and a corrected between-group effect size of 0.2 is considered small, 0.5 medium, and 0.8 large (Cohen, 1988).

Table 2 Average latent profile class probabilities for the most likely class allocation (row) by latent class (column)

| Latent Classes | Class allocation | | | M (SD.) Total score for SDQ |
|----------------|------------------|--------------|--------------|-----------------------------|
| | C1 (n = 68) | C2 (n = 479) | C3 (n = 138) | |
| 1 | 0.858 | 0.087 | 0.055 | 21.90 \pm 5.45 |
| 2 | 0.019 | 0.955 | 0.026 | 13.81 \pm 3,78 |
| 3 | 0.023 | 0.067 | 0.910 | 18.65 \pm 5.39 |

The columns refer to the latent class, and the rows refer to the most likely profile membership

Results

Baseline Pattern of LPA for SDQ Subscales

The five subscales from the SDQ were used to determine the latent classes, groups respectively. The results of five models (from one-class to five-class patterns), respectively, are presented in Table 1, indicating that the three-class solution yielded optimal model values (see Fig. 1). It demonstrated lower AIC, BIC, and aBIC values compared to model 2 and an acceptable value of entropy, although the entropy value for models 4 and 5 was better. In addition, the p -value for LMR and BLRT in model 3 also demonstrated that it had a good fit ($p < 0.001$), compared to the p -value for LMR and BLRT in models 4 and 5 that were not statistically significant. Ultimately, based on the five model fit indices' results, the three-class model of SDQ subscales was the best fitting model.

As shown in Table 2, the probabilities of classes 1, 2, and 3 of the three-class models were 85%, 95%, and 91%, respectively, suggesting a good discriminability and a reliable result of the LPA with a three-class model.

Table 3 Mean, Standard Deviation, and Cohen's *d* for SDQ subscales of the three latent classes of SDQ

| SDQ subscales | C1 (n = 68) | | C2 (n = 479) | | C3 (n = 138) | | Cohen's <i>d</i> | | |
|----------------------------|-------------|------|--------------|------|--------------|------|------------------|--------------|--------------|
| | M | SD | M | SD | M | SD | <i>d</i> 2-1 | <i>d</i> 3-1 | <i>d</i> 3-2 |
| Emotional problems | 2.03 | 0.90 | -0.26 | 0.69 | -0.08 | 0.80 | 2.16 | 2.47 | 0.24 |
| Conduct problems | 0.09 | 0.98 | -0.39 | 0.61 | 1.32 | 0.97 | 0.59 | 1.26 | 2.11 |
| Hyperactivity | 0.06 | 0.87 | -0.31 | 0.82 | 1.06 | 0.89 | 0.44 | 1.13 | 1.06 |
| Peer relationship problems | 1.28 | 1.11 | -0.34 | 0.71 | 0.55 | 1.06 | 1.73 | 0.67 | 0.98 |
| Prosocial behaviors | -0.45 | 0.90 | 0.44 | 0.69 | -1.30 | 0.70 | 1.10 | 1.05 | 2.50 |

The mean Z scores of five SDQ subscales are provided

C1 Internalizers, C2 Prosocials, C3 Externalizers

Table 4 Mean scores and standard deviations at pre-and post-measurement in the IG and CG

| Subscales | Group | Pre-M (SD) | Post-M (SD) | Cohen's <i>d</i> within-group | Cohen's <i>d</i> corrected between-group | Wald test (df = 1) group × time |
|---------------|-------|-------------|-------------|-------------------------------|--|---------------------------------|
| Externalizing | IG | 5.11 (4.30) | 3.23 (3.66) | 0.47 | 0.23 | 9.65** |
| | CG | 5.25 (4.23) | 4.36 (4.18) | 0.21 | | |
| Prosocial | IG | 6.58 (2.91) | 7.60 (2.50) | 0.38 | 0.21 | 6.28* |
| | CG | 6.67 (2.72) | 7.11 (2.42) | 0.17 | | |
| Internalizing | IG | 3.71 (3.55) | 2.30 (2.83) | 0.44 | 0.10 | 3.36 |
| | CG | 3.86 (3.65) | 2.81 (3.21) | 0.31 | | |

** = $p < 0.01$, * = $p < 0.05$

As shown in Fig. 1, the participants assigned to class 1 ($n = 68$, 9.9%) were characterized as having higher scores for emotional problems ($M = 2.03$) and peer relationship problems ($M = 1.28$). Moreover, they expressed a very low level of prosocial behaviors ($M = -0.45$), a neutral level of conduct problems ($M = 0.09$), and hyperactivity ($M = 0.06$) (the members of class 1 are referred to as the *Internalizers*). The participants assigned to class 2 ($n = 479$, 69.9%) were characterized as having the lowest scores on emotional problems ($M = -0.26$), conduct problems ($M = -0.39$), hyperactivity ($M = -0.31$), and peer relationship problems ($M = -0.34$). However, they scored highest in prosocial behaviors ($M = 0.45$) compared to participants allocated to other classes (the members of class 2 are referred to as the *Prosocials*). Class 3 ($n = 138$, 20.2%) reflected the individuals who endorsed the highest level of conduct problems ($M = 1.32$) and hyperactivity ($M = 1.06$) and the lowest level of prosocial behaviors ($M = -1.30$) compared to the other groups (the members of class 3 are referred to as the *Externalizers*).

Table 3 shows the means, standard deviations, and Cohen's *d* values comparing the five SDQ subscales across the three latent groups. Pairwise comparisons indicated that the children categorized as Internalizers scored higher on the SDQ subscales than those categorized as Prosocials, except for the hyperactivity subscale, which showed a moderate effect size ($d = 0.44$). Further, pairwise comparisons also showed that the children categorized as Externalizers scored higher on the SDQ subscales than those

categorized as Internalizers, except for the peer relationship problems subscale, which showed a medium effect size ($d = 0.67$). Furthermore, pairwise comparisons showed that the children categorized as Externalizers scored higher on the SDQ subscales than those categorized as Prosocials, except for prosocial behavior and emotional problems ($d = 0.24$).

Intervention Effects on Social-Emotional Behaviors

To determine the overall impact of the Papilio-3to6 program on children's social-emotional behaviors, represented by the broader internalizing and externalizing subscales, alongside the prosocial behavior subscale (i.e., the fifth SDQ subscale), group × time interactions were tested at a significance level of 0.05. The following covariates were included in the analysis: baseline (pretest) scores of the outcome variables, age, gender, and origin (German/non-German). Significant group × time interactions were found regarding the prosocial (Wald test (1) = 6.275, $p = 0.012$) and externalizing (Wald test (1) = 9.645, $p = 0.002$) subscales. The change on the prosocial subscale was significant ($\beta = 1.02$, $p < 0.001$) in the IG. The change was smaller in the control group but still significant ($\beta = 0.43$, $p < 0.003$). The change on the externalizing subscale was significant ($\beta = -1.88$, $p < 0.001$) in the IG. The change was smaller in the control group but still significant ($\beta = -0.89$, $p < 0.001$) (see Table 4).

Table 5 Mean scores and standard deviations at pre- and post-measurement in the IG and CG

| Subscales | Class | Control Pre-M (SD) | Control Post-M (SD) | Cohen's <i>d</i> within-group | Intervention Pre-M (SD) | Intervention Post-M (SD) | Cohen's <i>d</i> within-group | Cohen's <i>d</i> corrected between-group | Wald test (df = 2) group × time × class |
|---------------|---------------|--------------------|---------------------|-------------------------------|-------------------------|--------------------------|-------------------------------|--|---|
| Externalizing | Internalizers | 5.74 (3.74) | 5.95 (4.32) | 0.05 | 5.29 (3.44) | 2.19 (2.19) | 1.10 | 0.92 | 6.02* |
| | Prosocials | 3.70 (2.91) | 3.01 (3.18) | 0.23 | 3.27 (2.86) | 2.18 (2.90) | 0.38 | 0.14 | |
| | Externalizers | 11.07 (3.68) | 8.72 (4.26) | 0.59 | 10.65 (3.43) | 6.93 (4.01) | 1.00 | 0.39 | |
| Prosocial | Internalizers | 5.26 (2.20) | 5.97 (2.33) | 0.32 | 5.52 (2.86) | 7.61 (2.54) | 0.78 | 0.55 | 3.49 |
| | Prosocials | 7.78 (1.98) | 7.81 (2.02) | 0.02 | 7.95 (1.87) | 8.41 (1.90) | 0.25 | 0.22 | |
| | Externalizers | 3.11 (1.96) | 4.94 (2.38) | 0.85 | 2.88 (1.93) | 5.17 (2.45) | 1.05 | 0.24 | |
| Internalizing | Internalizers | 11.10 (2.88) | 6.41 (4.48) | 1.27 | 10.87 (3.09) | 4.65 (3.91) | 1.78 | 0.52 | 1.79 |
| | Prosocials | 2.51 (2.39) | 2.12 (2.60) | 0.16 | 2.41 (2.17) | 1.696 (2.30) | 0.32 | 0.14 | |
| | Externalizers | 4.99 (2.76) | 3.40 (2.97) | 0.55 | 4.69 (3.26) | 3.20 (3.14) | 0.47 | 0.03 | |

* = $p < 0.05$

Group (IG/CG) × Time × Latent Class Interaction

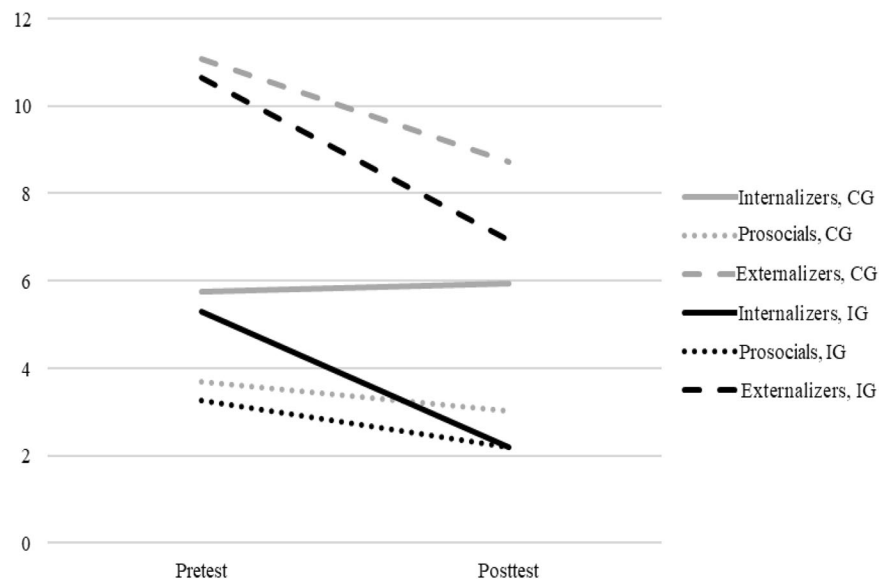
A stepwise procedure was applied in the subgroup analysis because of the low sample sizes in the Internalizers subgroup and the unequal distribution of participants in different groups (Altman, 2015). In step 1, no covariates were included at all. In step 2, the baseline scores of the outcome variables were added (Twisk et al., 2018). In step 3, the remaining demographic background variables, age, gender, and origin, were included in the group × time × latent class interaction analysis. An overall significant group × time × latent class membership interaction was found only regarding the externalizing subscale in steps 1 and 2 (only the results for step 2 are reported here: Wald test (2) = 6.024, $p = 0.049$). The group × time interaction was significant for the Internalizers, i.e., class 1, (Wald test (1) = 10.086, $p = 0.002$), the Prosocials, i.e., class 2, (Wald test (1) = 4.476, $p = 0.034$), and the Externalizers, i.e., class 3, (Wald test (1) = 4.923, $p = 0.027$). For Class 1, the change was only significant for the IG ($\beta = -3.097$, $p < 0.001$) and not for the CG ($\beta = 0.216$, $p = 0.838$). A significant drop in the rate of externalizing problems over time was found only for the IG. For class 2, the change was significant for both the IG ($\beta = -1.094$, $p < 0.001$) and CG ($\beta = -0.685$, $p < 0.001$), however, the change was bigger in the IG. For class 3, the change was significant for both the IG ($\beta = -3.720$, $p < 0.001$) and CG ($\beta = -2.352$, $p < 0.001$), however, the change was bigger in the IG. That is, a significant drop in the rate of externalizing problems over time was found for both the IG and CG, with a stronger decrease in the IG (see Table 5; Fig. 2). Thus, the results revealed that the Papilio-3to6 program affects all three latent classes (see Table 5).

Discussion

This study's main aim was to examine the impact of a universal preventive intervention program, i.e., the Papilio-3to6 program, on the development of social-emotional behaviors (in terms of prosocial behavior and internalizing and externalizing behavior problems) from both a variable and person-centered perspective. We identified subgroups of children characterized by SDQ subscales. We were particularly interested in examining the program's impact on each subgroup in the IG children compared to the CG children. Results suggest that the Papilio-3to6 program may be regarded as a universal preventive intervention program that improves children's social-emotional behaviors and prevents behavior problems successfully in early education and daycare settings—from the teachers' perspective.

This study aimed to adopt a person-centered approach to create subgroups, classes respectively, of children with

Fig. 2 The group \times time interaction in all three latent classes with externalizing subscale



different social-emotional behaviors at baseline. Using an LPA at the pretest, from the five model fit indices, a three-class model of SDQ subscales was the best fitting model, including the following subgroups:

(1) The children categorized as Internalizers scored higher on emotional and peer relationship problems. This group's characteristics are internal distress, anxiety, loneliness, depressive symptoms, social withdrawal, and somatic complaints. Internalizers have higher peer relationship problems. They may show limited social contact, social and academic skill problems, excessive need for feedback and reassurance, limited activity and expressions, and avoidance, which may further inhibit their ability to participate in peer activities. (2) The children categorized as Prosocials showed higher scores on prosocial behaviors. The characteristic of this group is the tendency to help, share, and take care of others, which evokes positive responses from others and promotes positive social relationships. (3) The children categorized as Externalizers showed higher scores on conduct problems and hyperactivity. This group's characteristic is aggressive, agitated, impulsive, and disruptive behaviors. They show high activity levels, poor regulation of impulses, and aggression toward peers. A closer look at the profiles revealed that whereas the children categorized as Internalizers and Externalizers showed a distinct pattern regarding emotional problems, conduct problems, and hyperactivity, the remaining subscales describing peer problems and prosocial behaviors in both groups pointed in the same direction compared to the dominant Prosocials. The effect sizes revealed only a medium difference (Cohen, 1988) in peer problems between the children categorized as Externalizers and Internalizers, with the Externalizers displaying fewer peer problems than the Internalizers. This result is even more interesting, considering that the children categorized as Externalizers were less

prosocial than those categorized as Internalizers, based on teacher reports. These results seem to suggest that despite their externalizing behavioral tendencies, the children categorized as Externalizers in our sample were better integrated into peer contexts than those categorized as Internalizers. These results are in accordance with Rubin (2021), who already summarized evidence that social withdrawal in childhood is mainly related to poor peer relations.

It is essential that teachers understand the characteristics and needs of children, particularly those categorized as Internalizers and Externalizers, given that they spend several hours a day with the children. For example, teachers can be prepared to help children categorized as Internalizers by recognizing each behavior and maladaptive thought associated and providing environments, modeling, and explicit instruction that support developing more adaptive thinking, such as: supporting children's self-efficacy, modeling positive and adaptive thinking, teaching and supporting self-monitoring, encouraging specific positive thoughts and behaviors, teaching children how to use self-affirmation, creating a checklist to highlight the tasks children completed, encouraging children's recall of when they have been successful (Stormont et al., 2015). For children categorized as Externalizers, teachers can support children by providing the appropriate discipline strategies, practical limit setting, and rewarding positive behavior. Teachers can focus, for example, on giving children skills for interacting with others (such as making friends, accepting others, or expressing individual needs) and targeting basic cognitive skills for social problem solving, such as perspective taking and self-control (for a review, see Schindler et al., 2015).

The second aim was to adopt a variable-centered approach to determine the longitudinal impact of the Papilio-3to6 program on the internalizing, externalizing,

and prosocial SDQ subscales. Theoretical and (preliminary) empirical support for our findings can be found in the literature, suggesting that broader SDQ subscales should be used when analyzing a low-risk sample based on a variable-centered approach (Goodman et al., 2010). This is: (1) merging the subscales for emotional problems and peer relationship problems into one subscale to describe internalizing problems; (2) merging the subscales for conduct problems and hyperactivity into one subscale to describe externalizing problems; (3) leaving the subscale for prosocial behaviors as it is. Results revealed that children who received the Papilio-3to6 program (IG) showed a more significant reduction in externalizing problems (e.g., conduct problems, aggressiveness, and antisocial behavior) and a more significant increase in prosocial behaviors (e.g., helping, sharing, and cooperating) compared to the CG. No intervention effects were found regarding internalizing problems. One explanation for this finding might be that teachers are, to a lesser extent, aware of children's internalizing problems (Van der Ende et al., 2012). Internalizing problems, such as feelings of depression and loneliness, are presumably less observable and typically occur as part of 'normal' behaviors such as sadness, worrying, and withdrawal, making them less visible to teachers. Internalized children are often labeled as 'good' or 'forgotten' children in the classroom, while externalizing behaviors such as fighting or overt forms of bullying are more disturbing in the classroom and dominate a teacher's attention (Van der Ende et al., 2012). Thus, as a result of our present study, and considering the negative consequences of early internalizing problems, it is essential to educate teachers on the manifestations and perception of internalizing problems (Onchwari, 2010), which will also be a necessary topic for further program development. Another explanation for this result could be that the program measures more adequately address externalizing problems and prosocial behaviors than internalizing problems. Additional analyses are planned to investigate this in more detail.

The third aim of this study was to determine the impact of the Papilio-3to6 program on children's behavior after being allocated to the three identified subgroups and classes, respectively. A significant group (IG/CG) \times time \times class membership interaction was only found regarding the externalizing subscale. Group (IG/CG) \times time interaction was significant for the Internalizers (i.e., class 1) and Externalizers (i.e., class 3). For the children categorized as Internalizers, externalizing problems decreased over time only in the IG. For the children categorized as Prosocials and Externalizers, a more significant reduction in externalizing problems over time was evident in the IG compared to the CG. These results indicate that the Papilio-3to6 program is most effective when it comes to reducing externalizing behaviors for all three classes. In an earlier study conducted by Vazsonyi et al.

(2004), a similar approach was undertaken to inform the effectiveness of a universal violence prevention program regarding different subgroups. The participants were grouped into different risk groups according to the same baseline measures that were later used as outcomes. As expected, high-risk groups profited the most from the intervention regarding reducing aggression and promoting social competence.

Based on Cohen's benchmark (1988), the effects of the Papilio-3to6 program for the children categorized as Internalizers reached a high magnitude, while for those categorized as Externalizers and Prosocials, the program had medium and small effects, respectively. According to Lipsey et al. (2012), even small effects of a universal intervention program in the education field could be considered large. This corresponds with the findings of meta-analyses that found an average effect size of 0.28 SD among a sample of 124 randomized intervention trials in education research (Lipsey et al., 2012).

Our findings might not be surprising regarding the children categorized as Externalizers since they are characterized by the highest externalizing behaviors. As the latent profiles showed, even those categorized as the Prosocials and Internalizers at young ages exhibited a certain amount of externalizing behavior (i.e., conduct problems and hyperactivity). Regarding the children categorized as Internalizers, the results—at first sight—seem to be surprising. However, whereas the children classified as Externalizers show a more evident behavior pattern, the behavioral profile is more complicated regarding those categorized as Internalizers. The children classified as the Internalizers group represent the smallest group in the present sample, and they might receive less attention in day-care centers due to their less disturbing behavior than those classified as the Externalizers. This rather untypical and, to some degree, an ambivalent behavioral pattern has been well established in the clinical setting. It has been found that internalizing problems, such as symptoms of depression, and externalizing problems, such as conduct problems, co-occur at higher rates than expected from early childhood into adulthood (for a review, see Capaldi & Kim, 2014; McElroy et al., 2018). It seems that the implemented Papilio-3to6 measures are best suited for children showing moderate levels of externalizing behavior, offering a helpful toolbox to manage these externalizing facets in an overall mixed behavioral pattern.

However, prevention and intervention programs targeting co-occurring symptoms of externalizing and internalizing problems remain limited (Capaldi & Kim, 2014). Given the widespread adjustment difficulties for those children with co-occurring problems (for a review, see Capaldi & Kim, 2014), future efforts in universal prevention should address this subgroup of children with adequate measures targeting both externalizing and internalizing problems. However, it is worth mentioning that interactions were non-significant when including all demographic variables as covariates. Although this result is difficult to interpret, it might be due to the loss of

statistical power because of the low sample size in the Internalizers group. The results of the LPA empirically determined the number of children in each class from the present data, and therefore it was not possible to predict the required sample size beforehand.

Limitations and Future Directions

The present study has some limitations that are worth mentioning. First, the children's behavioral outcomes in this study were assessed using only teacher reports. In general, the psychometric properties of the teacher reports form of the SDQ for children are good, and several studies have shown that self-reports of social behavior in early childhood are relatively weak regarding reliability. For example, Dahlberg et al.'s (2019) study with 3–5-year-olds children in a non-clinical sample revealed a good construct validity of the SDQ rated by daycare teachers. In addition, Sharp et al. (2005) examined the concurrent predictive validity of SDQ ratings from parents and teachers regarding elementary school children's help-seeking behavior and reported higher values for teachers than for parents' ratings. However, teacher ratings of child behavior are limited to the institutional context in which the children are observed and, to some degree, reflect teachers' beliefs and expectations (Stone et al., 2010). Future research can address this limitation by conducting direct behavioral observations or using age-appropriate self-reports to assess children's behavioral outcomes. In addition, we would like to point out that the teachers were also the main target group of the intervention, so an influence on the teachers' assessments is possible. Since a response shift bias is thus possible, we asked the teachers in a retrospective pretest design (Pratt et al., 2000) to rate the children's behavioral and emotional problems from their perspective on the scales of the SDQ, before the training (T1), after the training (T2), and retrospectively at T2 (manuscript reporting these results in preparation). Comparing teacher and parent reports, it is possible that teachers underrated children's internalizing behavior as it is less available for teacher observation than externalizing behavior (Rubin et al., 2013a, 2013b). Thus, a multi-methodological approach, including direct behavioral observations, play-oriented measures, or age-appropriate self-reports to assess data from children, would provide additional insights that allow analyzing the program's effectiveness in more detail.

Second, it would be interesting to measure the group climate based on teacher ratings, to examine whether the children classified as the Prosocials would benefit indirectly from a better group climate if the externalizing problems have been reduced in children classified as the Internalizers and Externalizers. Third, no additional analysis of gender effects was possible due to the small sample size in each subgroup. Thus, future studies could explore potential gender differences using a larger sample size. Fourth, the Papilio-3to6 program

consists of several measures that address different aspects of children's social-emotional development with different educational approaches. In the present study, the program's effectiveness was tested "as a whole" (so-called black box evaluation), and therefore it was not possible to disentangle the value each measure adds to the general effectiveness in different domains. Future studies applying more complex evaluation designs and testing for complex change models would help unravel each measure's contribution to the overall effect of the Papilio-3to6 program. Fifth, although blinding (e.g., teachers, assessors) is a valuable strategy to avoid bias, we were not permitted to mislead or fail to inform early education care providers, facilities, and teachers about the study's aims. Participation in the study - which, after all, included extensive program training over several days and monitoring in program implementation - would not have been possible in the context of a blinded design. Implementing psychosocial preventive (universal) intervention programs is barely possible or not practicable (Juil et al., 2021) using a blind- or even double-blind-controlled design. Thus, we were not able to include a blinding strategy in our study design. Sixth, the family components were not part of the implemented program components in the present study; however, the impact of the family will be addressed in an additional paper that deals with the question of 'whether family and early risk factors do have an impact on the effectiveness of the Papilio-3to6 intervention program in the early childhood education setting'. Finally, since the program implementation was limited to a relatively narrow region in Southern Germany, one must be cautious about generalizing the results. However, the sample included institutions in urban and rural settings, which increases the representativeness of the results.

Conclusions and Implications

Taking a person-centered approach, a three-class model of SDQ subscales was chosen as the best-fitting model in the present study. The three latent classes included: one group with higher scores describing emotional problems and peer relationship problems, i.e., Internalizers; a group with higher scores regarding prosocial behaviors, i.e., Prosocials; and a group showing higher scores concerning conduct problems and hyperactivity, i.e., Externalizers. The Papilio-3to6 program, as a universal preventive intervention program, has been found to reduce externalizing problems and promote prosocial behaviors in children aged 3 to 6. Further, the results suggest that the effects were valid for children from all three latent classes receiving the Papilio-3to6 program. The children classified as the Internalizers, Prosocials, and Externalizers all benefited from the program with a reduction in externalizing problems. However, in the present study, we did not find improvements concerning internalizing problems.

The results of the current study are particularly significant - for educators, parents/caregivers, and the implementation of preventive measures - as they clearly indicate that children with internalizing symptoms -- even and especially when they otherwise show prosocial behavior or are particularly conspicuous for externalizing symptoms -- receive too little attention. Internalizing behavior tends to be viewed as less problematic because such behaviors are most often characterized by quiet, internal distress rather than disruptive or socially negative behaviors. Such features may also make these behaviors more difficult to detect in young children who generally have less well-developed verbal skills and an even more limited capacity to describe internal feeling states (for a review, see Tandon et al., 2009). Several studies examining the persistence of internalizing problems from preschool age onward found that internalizing problems can be substantially stable or increase and often inflict impairment (e.g., Bufferd et al., 2012; Klein et al., 2019; Von Klitzing et al., 2014). Parents and educators should therefore pay particular attention to children who appear to be relatively quiet and adjusted, as this can be a problem if the children are too quiet and withdrawn compared to other children. In the same way, parents and educators should pay attention to children who were initially noticed for their externalizing behavior to ensure that they do not also show or have developed internalizing problems (e.g., when the externalizing behavior has decreased). Consequently, our findings might have implications for designing future preventive intervention programs or for further developments in the Papilio-3to6 program, respectively. For example, (1) inclusion of measures targeting internalizing problems; (2) inclusion of training components that train teachers to detect internalizing problems and address these behavioral tendencies in 3-to-6-year-old children.

Data availability

The data generated and/or analyzed during the current study are not publicly available for legal/ethical reasons but are available from the corresponding author on reasonable request.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10826-023-02655-4>.

Acknowledgements We would like to thank the former staff members at beta Institut gGmbH, especially Horst Erhardt and Beatriz Barquero-Santander, and all the kindergarten teachers, participating children, and parents. Heidrun Mayer, co-developer of the program Papilio-3to6 and co-founder of the association Papilio e.V., passed away unexpectedly in August 2016 due to a bicycle accident. We mourn the loss of Heidrun Mayer and thank her for the trusting and cordial cooperation.

Author contributions M.Z.-N.: Conceptualization, Methodology, Investigation, Writing – original draft preparation, Writing – review & editing.

A.M.-G.: Methodology, Formal analysis, Writing - review & editing. J.M.: Methodology, Formal analysis. M.H.: Conceptualization, Methodology, Investigation, Writing - review & editing, Visualization. S.H.-N.: Writing - review & editing. H.S.: Conceptualization, Methodology, Investigation, Writing - review & editing, Supervision, Project administration.

Funding This study was supported by a grant from the Academy of Finland to M.Z.-N. [grant number 330960]. The development of Papilio-3to6 and the evaluation study was conducted by beta Institut gGmbH für sozialmedizinische Forschung und Entwicklung, Augsburg, in cooperation with science partners at Universität Bremen (Franz Petermann, Ute Koglin, Herbert Scheithauer) and Universität Augsburg (Leonie Herwartz-Emden), since 2004 with science partners at Freie Universität Berlin (especially Herbert Scheithauer). Program development and the evaluation study was supported by funding from Bayerisches Staatsministerium für Umwelt, Gesundheit und Verbraucherschutz, betapharm Arzneimittel GmbH, BMW Group, and Rotary District 1840. Responsible for the distribution and implementation of the program Papilio-3to6 is the Papilio gGmbH, Augsburg, Germany. Open access funding provided by University of Eastern Finland (UEF) including Kuopio University Hospital.

Compliance with Ethical Standards

Conflict of interest The authors declare no competing interests.

Ethics approval The study was conducted following the ethical principles of the Finnish Advisory Board on Research Integrity, the German Research Foundation, and the ethical principles of the Helsinki Declaration and was approved by the University of Bremen.

Informed consent Informed consent was obtained from parent(s)/guardian(s) for all children participating in the study.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Altman, D. (2015). Subgroup analyses in randomized trials—more rigour needed. *Nature Reviews Clinical Oncology*, 12, 506–507. <https://doi.org/10.1038/nrclinonc.2015.133>.
- Bauer, D. J., & Shanahan, M. J. (2007). Modeling complex interactions: Person-centered and variable-centered approaches. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 255–284). Erlbaum
- Belsky, J., & Pluess, M. (2009). Beyond diathesis stress: Differential susceptibility to environmental influences. *Psychological Bulletin*, 135, 885–908. <https://doi.org/10.1037/a0017376>.

- Bergman, L. R., & Magnusson, D. (1997). A person-oriented approach in research on developmental psychopathology. *Development and Psychopathology*, *9*, 291–319. <https://doi.org/10.1017/S095457949700206x>.
- Bierman, K. (2002). The implementation of the Fast Track Program: An example of a large-scale prevention science efficacy trial. *Journal of Abnormal Child Psychology*, *30*, 1–17. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2756657/>
- Bierman, K. L., & Motamedi, M. (2015). SEL programs for preschool children. In J. A. Durlak, C. E. Domitrovich, R. P. Weissberg, & T. P. Gullotta (Eds.), *Handbook of social and emotional learning: Research and practice* (pp. 135–150). The Guilford Press.
- Bierman, K. L., Nix, R. L., Greenberg, M. T., Blair, C., & Domitrovich, C. E. (2008). Executive functions and school readiness intervention: Impact, moderation, and mediation in the Head Start REDI program. *Development and Psychopathology*, *20*, 821–843. <https://doi.org/10.1017/S0954579408000394>.
- Bornstein, M. H., Hahn, C. S., & Haynes, O. M. (2010). Social competence, externalizing, and internalizing behavioral adjustment from early childhood through early adolescence: Developmental cascades. *Development and psychopathology*, *22*, 717–735. <https://doi.org/10.1017/S0954579410000416>.
- Bradshaw, C. P., Waasdorp, T. E., & Leaf, P. J. (2015). Examining variation in the impact of school-wide positive behavioral interventions and supports: Findings from a randomized controlled effectiveness trial. *Journal of Educational Psychology*, *107*, 546–557. <https://doi.org/10.1037/a0037630>.
- Bufferd, S. J., Dougherty, L. R., Carlson, G. A., Rose, S., & Klein, D. N. (2012). Psychiatric disorders in preschoolers: Continuity from ages 3 to 6. *American Journal of Psychiatry*, *169*, 1157–1164. <https://doi.org/10.1176/appi.ajp.2012.12020268>.
- Buscemi, L., Bennett, T., Thomas, D., & Deluca, D. A. (1996). Head start: Challenges and training needs. *Journal of Early Intervention*, *20*, 1–13. <https://doi.org/10.1177/105381519602000101>.
- Capaldi, D. M., & Kim, H. K. (2014). Comorbidity of depression and conduct disorder. In C. S. Richards & M. W. O'Hara (Eds.), *The Oxford handbook of depression and comorbidity* (pp. 186–199). Oxford University Press
- Caputi, M., Lecce, S., Pagnin, A., & Banerjee, R. (2012). Longitudinal effects of theory of mind on later peer relations: The role of prosocial behavior. *Developmental Psychology*, *48*, 257–270. <https://doi.org/10.1037/a0025402>.
- Caron, C., & Rutter, M. (1991). Comorbidity in child psychopathology: Concepts, issues and research strategies. *Child Psychology & Psychiatry & Allied Disciplines*, *32*, 1063–1080. <https://doi.org/10.1111/j.1469-7610.1991.tb00350.x>.
- Carragher, N., Adamson, G., Bunting, B., & McCann, S. (2009). Subtypes of depression in a nationally representative sample. *Journal of Affective Disorders*, *113*, 88–99. <https://doi.org/10.1016/j.jad.2008.05.015>.
- Cohen J. (1988). *Statistical power analysis for the behavioral sciences*. Routledge Academic.
- Collie, R. J., Martin, A. J., Nassar, N., & Roberts, C. L. (2019). Social and emotional behavioral profiles in kindergarten: A population-based latent profile analysis of links to socio-educational characteristics and later achievement. *Journal of Educational Psychology*, *111*, 170–187. <https://doi.org/10.1037/edu0000262.supp>.
- Connors-Burrow, N. A., Patrick, T., Kyzer, A., & McKelvey, L. (2017). A preliminary evaluation of REACH: Training early childhood teachers to support children's social and emotional development. *Early Childhood Education Journal*, *45*, 187–199. <https://doi.org/10.1007/s10643-016-0781-2>.
- Cristóvão, A. M., Candeias, A. A., & Verdasca, J. (2017). Social and emotional learning and academic achievement in Portuguese schools: A bibliometric study. *Frontiers in Psychology*, *8*, 1913. <https://doi.org/10.3389/fpsyg.2017.01913>.
- Dahlberg, A., Ghaderi, A., Sarkadi, A., & Salari, R. (2019). SDQ in the hands of fathers and preschool teachers—Psychometric properties in a non-clinical sample of 3–5-year-olds. *Child Psychiatry and Human Development*, *50*, 132–141. <https://doi.org/10.1007/s10578-018-0826-4>.
- Denham, S. A., Bassett, H. H., Brown, C., Way, E., & Steed, J. (2015). “I Know How You Feel”: Preschoolers' emotion knowledge contributes to early school success. *Journal of Early Childhood Research*, *13*, 252–262. <https://doi.org/10.1177/1476718x13497354>.
- Denham, S. A., Bassett, H., Mincic, M., Kalb, S., Way, E., Wyatt, T., & Segal, Y. (2012). Social-emotional learning profiles of preschoolers' early school success: A person-centered approach. *Learning and Individual Differences*, *22*, 178–189. <https://doi.org/10.1016/j.lindif.2011.05.001>.
- Denham, S. A., Brown, C., & Domitrovich, C. (2010). “Plays nice with others”: Social-emotional learning and academic success. *Early Education and Development*, *21*, 652–680. <https://doi.org/10.1080/10409289.2010.497450>.
- Farrell, A. D., Henry, D. B., & Bettencourt, A. (2013). Methodological challenges examining subgroup differences: Examples from universal school-based youth violence prevention trials. *Prevention Science*, *14*, 121–133. <https://doi.org/10.1007/s11121-011-0200-2>.
- Fox, L., Hemmeter, M. L., Snyder, P., Binder, D. P., & Clarke, S. (2011). Coaching early childhood special educators to implement a comprehensive model for promoting young children's social competence. *Topics in Early Childhood Special Education*, *31*, 178–192. <https://doi.org/10.1177/0271121411404440>.
- Goodman, A., Lamping, D. L., & Ploubidis, G. B. (2010). When to use broader internalizing and externalizing subscales instead of the hypothesized five subscales on the Strengths and Difficulties Questionnaire (SDQ): Data from British parents, teachers and children. *Journal of Abnormal Child Psychology*, *38*, 1179–1191. <https://doi.org/10.1007/s10802-010-9434-x>.
- Greenberg, M. T., & Abenavoli, R. (2017). Universal interventions: Fully exploring their impacts and potential to produce population-level impacts. *Journal of Research on Educational Effectiveness*, *10*, 40–67. <https://doi.org/10.1080/19345747.2016.1246632>.
- Gupta, S. K. (2011). Intention-to-treat concept: A review. *Perspectives in Clinical Research*, *2*, 109–112. <https://doi.org/10.4103/2229-3485.83221>.
- Han, H. S. (2014). Supporting early childhood teachers to promote children's social competence: Components for best professional development practices. *Early Childhood Education Journal*, *42*, 171–179. <https://doi.org/10.1007/s10643-013-0584-7>.
- Herrenkohl, T. I., Kosterman, R., Mason, W. A., Hawkins, J. D., McCauley, C. A., & McCauley, E. (2010). Effects of childhood conduct problems and family adversity on health, health behaviors, and service use in early adulthood: Tests of developmental pathways involving adolescent risk taking and depression. *Development and Psychopathology*, *22*, 655–665. <https://doi.org/10.1017/S0954579410000349>.
- Housman, D. K., Denham, S. A., & Cabral, H. (2018). Building young children's emotional competence and self-regulation from birth: The “Begin to... ECSEL” approach. *International Journal of Emotional Education*, *10*, 5–25. <https://eric.ed.gov/?id=EJ1197894> Eric.
- Hutchings, J., Martin-Forbes, P., Daley, D., & Williams, M. E. (2013). A randomized controlled trial of the impact of a teacher classroom management program on the classroom behavior of children with and without behavior problems. *Journal of School Psychology*, *51*, 571–585. <https://doi.org/10.1016/j.jsp.2013.08.001>.
- Izard, C. E. (2002). Translating emotion theory and research into preventive interventions. *Psychological Bulletin*, *128*, 796–824. <https://doi.org/10.1037/0033-2909.128.5.796>.
- Jensen, B., Jensen, P., & Rasmussen, A. W. (2017). Does professional development of preschool teachers improve children's socio-

- emotional outcomes? *Labour Economics*, 45, 26–39. <https://doi.org/10.2139/ssrn.2589819>.
- Jensen, P. & Rasmussen, A. W. (2019). Professional development and its impact on children in early childhood education and care: A meta-analysis based on European studies. *Scandinavian Journal of educational research*, 63, 935–950. <https://doi.org/10.1080/00313831.2018.1466359>.
- Jones, D. E., Greenberg, M., & Crowley, M. (2015). Early social-emotional functioning and public health: The relationship between kindergarten social competence and future wellness. *American Journal of Public Health*, 105, 2283–2290. <https://doi.org/10.2105/ajph.2015.302630>.
- Juul, S., Gluud, C., Simonsen, S., Weischer Frandsen, F., Kirsch, I., & Jakobsen, J. C. (2021). Blinding in randomised clinical trials of psychological interventions: A retrospective study of published trial reports. *BMJ Evidence-Based Medicine*, 26, 109. <https://doi.org/10.1136/bmjebm-2020-111407>.
- Klasen, H., Woerner, W., Rothenberger, A., & Goodman, R. (2003). Die deutsche Fassung des Strengths and Difficulties Questionnaire (SDQ-Deu) – Übersicht und Bewertung erster Validierungs- und Normierungsbefunde [German version of the Strengths and Difficulties Questionnaire]. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 52, 491–502. Hdl <http://hdl.handle.net/20.500.11780/2700>.
- Klein, A. M., Schlesier-Michel, A., Otto, Y., White, L. O., Andreas, A., Sierau, S., Bergmann, S., Perren, S., & Von Klitzing, K. (2019). Latent trajectories of internalizing symptoms from preschool to school age: A multi-informant study in a high-risk sample. *Development and Psychopathology*, 31, 657–681. <https://doi.org/10.1017/S0954579418000214>.
- Koglin, U., Barquero, B., Mayer, H., Scheithauer, H., & Petermann, F. (2007). Deutsche version des Strengths and Difficulties Questionnaire (SDQ-Deu) [German version of the Strengths and Difficulties Questionnaire]. *Diagnostica*, 53, 175–183. <https://doi.org/10.1026/0012-1924.53.4.175>.
- Koivula, M., & Huttunen, K. (2018). Children's social-emotional development and its support: guest editorial. *Journal of Early Childhood Education Research*, 7, 177–183. <http://urn.fi/URN:NBN:fi:juyu-201901041071> Urn.
- Lanza, S. T., & Rhoades, B. T. (2013). Latent class analysis: An alternative perspective on subgroup analysis in prevention and treatment. *Prevention Science*, 14, 157–168. <https://doi.org/10.1007/s11121-011-0201-1>.
- Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K. S., & Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms*. National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education. <https://ies.ed.gov/ncser/pubs/20133000/>
- Lösel, F., Beelmann, A., Stemmler, M., & Jaurusch, S. (2006). Prävention von Problemen des Sozi-alverhaltens im Vorschulalter. Evaluation des Eltern- und Kindertrainings EFFEKT [Prevention of social behavior problems at preschool age: Evaluation of the Parent and Child Training Program Package EFFEKT]. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 35, 127–139. <https://doi.org/10.1026/1616-3443.35.2.127>.
- Mayer, H., Heim, P., Peter, C., & Scheithauer, H. (2016). *Papilio: Theorie und Grundlagen. Ein Programm für Kindertagesstätten zur Prävention von Verhaltensproblemen und zur Förderung sozial-emotionaler Kompetenz. Ein Beitrag zur Sucht- und Gewaltprävention* (4. überarbeitete und ergänzte Auflage) [Papilio: Theory and basic information]. Papilio Verlag
- McElroy, E., Shevlin, M., Murphy, J., & McBride, O. (2018). Co-occurring internalizing and externalizing psychopathology in childhood and adolescence: A network approach. *European Child & Adolescent Psychiatry*, 27, 1449–1457. <https://doi.org/10.1007/s00787-018-1128-x>.
- McLaren, E. M., Hall, P. J., & Fox, P. (2009). Kentucky's early childhood professional development initiative to promote social-emotional competence. *NHSA Dialog*, 12, 170–183. <https://doi.org/10.1080/15240750902774700>.
- Mejia, R., & Kliewer, W. (2006). Socio-emotional development during childhood: Theories and research trends. *CES Medicina*, 20, 49–54. Researchgate <https://www.researchgate.net/publication/277120341>.
- Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., Ross, S., Sears, M. R., Thomson, W. M., & Avshalom, C. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*, 108, 2693–2698. <https://doi.org/10.1073/pnas.1010076108>.
- Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345–368). Sage
- Muthén, B. O., & Khoo, S.-T. (1998). Longitudinal studies of achievement growth using latent variable modeling. *Learning and Individual Differences*, 10, 73–101. [https://doi.org/10.1016/s1041-6080\(99\)80135-6](https://doi.org/10.1016/s1041-6080(99)80135-6).
- Nix, R. L., Bierman, K. L., Domitrovich, C. E., & Gill, S. (2013). Promoting children's social-emotional skills in preschool can enhance academic and behavior functioning in kindergarten: findings from Head Start REDI. *Early Education and Development*, 24, 1000–1019. <https://doi.org/10.1080/10409289.2013.825565>.
- Onchwari, J. (2010). Early childhood inservice and preservice teachers' perceived levels of preparedness to handle stress in their students. *Early Childhood Education*, 37, 391–400. <https://doi.org/10.1007/s10643-009-0361-9>.
- Petermann, U., Petermann, F., & Schreyer, I. (2010). The German Strengths and Difficulties Questionnaire (SDQ): Validity of the teacher version for preschoolers. *European Journal of Psychological Assessment*, 26, 256–262. <https://doi.org/10.1027/1015-5759/a000034>.
- Pluess, M. (2015). Individual differences in environmental sensitivity. *Child Development Perspectives*, 9, 138–143. <https://doi.org/10.1111/cdep.12120>.
- Polier, G. G., Vloet, T. D., Herpertz-Dahlmann, B., Laurens, K. R., & Hodgins, S. (2012). Comorbidity of conduct disorder symptoms and internalizing problems in children: investigating a community and a clinical sample. *European Child & Adolescent Psychiatry*, 21, 31–38. <https://doi.org/10.1007/s00787-011-0229-6>.
- Pratt, C. C., McGuigan, W. M., & Katzev, A. R. (2000). Measuring program outcomes: Using retrospective pre-test methodology. *American Journal of Evaluation*, 21, 341–349. [https://doi.org/10.1016/s1098-2140\(00\)00089-8](https://doi.org/10.1016/s1098-2140(00)00089-8).
- Rothwell, P. M. (2005). Subgroup analysis in randomized controlled trials: Importance, indication and interpretation. *Lancet*, 365, 176–186. [https://doi.org/10.1016/s0140-6736\(05\)17709-5](https://doi.org/10.1016/s0140-6736(05)17709-5).
- Rubin, D. H., Althoff, R. R., Walkup, J. T., & Hudziak, J. J. (2013a). Cross-informant agreement on child and adolescent withdrawn behavior: A latent class approach. *Child Psychiatry & Human Development*, 44, 361–369. <https://doi.org/10.1007/s10578-012-0330-1>.
- Rubin, K. H. (2021). Social Withdrawal in Childhood: A Personal History of Theory and Research That Guided a Program of Developmental Research. In R. J. Coplan, J. C. Bowker, & L. J. Nelson (Eds.), *The handbook of solitude: Psychological perspectives on social isolation, social withdrawal, and being alone* (pp. 405–447). John Wiley & Sons

- Rubin, K. H., Bowker, J. C., McDonald, K. L., & Menzer, M. (2013b). Peer relationships in childhood. In P. D. Zelazo (Ed.), *The Oxford handbook of developmental psychology*, Vol. 2, (pp. 242–275). Oxford University Press
- Scheithauer, H., Bondü, R., Hess, M., & Mayer, H. (2016). Förderung sozial-emotionaler Kompetenzen im Vorschulalter: Ergebnisse der Augsburger Längsschnittstudie zur Evaluation des primärpräventiven Programms Papilio® (ALEPP). In T. Malti & S. Perren (Eds.), *Soziale Kompetenz bei Kindern und Jugendlichen – Entwicklungsprozesse und Fördermöglichkeiten* [Social competence in children and adolescents - developmental processes and opportunities for promotion] (2. bearb. & erw. Aufl.) (pp. 155–176). Kohlhammer
- Scheithauer, H., Hess, M., Zarra-Nezhad, M., Peter, C., & Wölfer, R. (2022). Developmentally appropriate prevention of behavioral and emotional problems, social-emotional learning, and developmentally appropriate practice for early childhood education and care: The Papilio-3to6 program. *International Journal of Developmental Science*, 16(3–4), 81–97. <https://doi.org/10.3233/DEV-220331>.
- Schick, A., & Cierpka, M. (2003). Faustlos: Evaluation eines Curriculums zur Förderung sozial-emotionaler Kompetenzen und zur Gewaltprävention in der Grundschule. *Kindheit und Entwicklung*, 12, 100–110. <https://doi.org/10.1026//0942-5403.12.2.100>.
- Schindler, H. S., Kholoptseva, J., Oh, S. S., Yoshikawa, H., Duncan, G. J., Magnuson, K. A., & Shonkoff, J. P. (2015). Maximizing the potential of early childhood education to prevent externalizing behavior problems: A meta-analysis. *Journal of School Psychology*, 53, 243–263. <https://doi.org/10.1016/j.jsp.2015.04.001>.
- Schochet, P. Z., Puma, M., & Deke, J. (2014). *Understanding variation in treatment effects in education impact evaluations: An overview of quantitative methods* (NCEE Report 2014–4017). National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education. Ies. <https://ies.ed.gov/ncee/pubs/20144017/>
- Sharp, C., Croudace, T. J., Goodyer, I. M., & Amtmann, D. (2005). The Strengths and Difficulties Questionnaire: Predictive validity of parent and teacher ratings for help-seeking behavior over one year. *Educational and Child Psychology*, 22(3), 28–44. Psycnet. <https://psycnet.apa.org/record/2005-14159-004>.
- Shin, K. M., Cho, S. M., Shin, Y. M., & Park, K. S. (2016). Effects of early childhood peer relationships on adolescent mental health: A 6-to 8-year follow-up study in South Korea. *Psychiatry Investigation*, 13, 383–388. <https://doi.org/10.4306/pi.2016.13.4.383>.
- Silkenbeumer, J., Schiller, E. M., Holodynski, M., & Kärtner, J. (2016). The role of co-regulation for the development of social-emotional competence. *Journal of Self-Regulation and Regulation*, 2, 17–32. <https://doi.org/10.11588/josar.2016.2.34351>.
- Spilt, J. L., Koot, J. M., & van Lier, P. A. C. (2013). For whom does it work? Subgroup differences in the effects of a school-based universal prevention program. *Prevention Science*, 14, 479–488. <https://doi.org/10.1007/s11211-012-0329-7>.
- Statistisches Bundesamt (2020). *Betreuungsquote von Kindern unter 6 Jahren nach Bundesländern*. Destatis. <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Soziales/Kinderta>
[gesbetreuung/Tabellen/betreuungsquote-2018.html;jsessionid=AF13CFEA70B0935AB2751D7ADE1C648.live721](https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Soziales/Kindertagesbetreuung/Tabellen/betreuungsquote-2018.html;jsessionid=AF13CFEA70B0935AB2751D7ADE1C648.live721)
- Stone, L. L., Otten, R., Engels, R. C. M. E., Vermulst, A. A., & Janssens, J. M. A. M. (2010). Psychometric properties of the parent and teacher versions of the Strengths and Difficulties Questionnaire for 4- to 12-year-olds: A review. *Clinical Child and Family Psychology Review*, 13, 254–274. <https://doi.org/10.1007/s10567-010-0071-2>.
- Stormont, M., Herman, K. C., & Reinke, W. M. (2015). The overlooked children: How teachers can support children with internalizing behaviors. *Beyond Behavior*, 24(2), 39–45. <https://doi.org/10.1177/107429561502400206>.
- Supplee, L. H., Kelly, B. C., MacKinnon, D. M., & Barofsky, M. Y. (2013). Introduction to the special issue: Subgroup analysis in prevention and intervention research. *Prevention Science*, 14, 107–110. <https://doi.org/10.1007/s11211-012-0335-9>.
- Tandon, M., Cardeli, E., & Luby, J. (2009). Internalizing disorders in early childhood: A review of depressive and anxiety disorders. *Child and Adolescent Psychiatric Clinics*, 18, 593–610. <https://doi.org/10.1016/j.chc.2009.03.004>.
- Twisk, J., Bosman, L., Hoekstra, T., Rijnhart, J., Welten, M., & Heymans, M. (2018). Different ways to estimate treatment effects in randomised controlled trials. *Contemporary Clinical Trials Communications*, 10, 80–85. <https://doi.org/10.1016/j.conctc.2018.03.008>.
- Van der Ende, J., Verhulst, F. C., & Tiemeier, H. (2012). Agreement of informants on emotional and behavioral problems from childhood to adulthood. *Psychological Assessment*, 24, 293–300. <https://doi.org/10.1037/a0025500>.
- Van Lier, P. A., & Koot, H. M. (2010). Developmental cascades of peer relations and symptoms of externalizing and internalizing problems from kindergarten to fourth-grade elementary school. *Development and Psychopathology*, 22, 569–582. <https://doi.org/10.1017/s0954579410000283>.
- Vazsonyi, A. T., Belliston, L. M., & Flannery, D. J. (2004). Evaluation of a school-based, universal violence prevention program: Low-, medium-, and high-risk children. *Youth Violence and Juvenile Justice*, 2, 185–206. <https://doi.org/10.1177/1541204003262224>.
- Von Klitzing, K., Doyle, O., Saías, T., Greacen, T., Sierau, S., & Perren, S. (2011). Early prevention and intervention programs in Europe. In H. E. Fitzgerald, K. Puura, M. Tomlinson, & C. Paul (Eds.), *International perspectives on children and mental health* (pp. 107–147). Praeger
- Von Klitzing, K., White, L. O., Otto, Y., Fuchs, S., Egger, H. L., & Klein, A. M. (2014). Depressive comorbidity in preschool anxiety disorder. *Journal of Child Psychology and Psychiatry*, 55, 1107–1116. <https://doi.org/10.1111/jcpp.12222>.
- Willner, C. J., Gatzke-Kopp, L. M., & Bray, B. C. (2016). The dynamics of internalizing and externalizing comorbidity across the early school years. *Development and Psychopathology*, 28, 1033–1052. <https://doi.org/10.1017/s0954579416000687>.
- Winsler, A., & Wallace, G. L. (2002). Behavior problems and social skills in preschool children: Parent-teacher agreement and relations with classroom observations. *Early Education and Development*, 13, 41–58. https://doi.org/10.1207/s15566935eed1301_3.