JECER

VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

The effects of ProVaka and Papilio interventions on children's behavior and educator feedback: A cluster-randomized controlled trial in early childhood education and care

Noora Heiskanen[a], Anne Karhu[b], Merja Koivula[c], Päivi Moisio[d], Hannu Savolainen[e], Ville Vauhkonen[f] & Vesa Närhi[g]

[a] University of Eastern Finland, corresponding author, email: noora.heiskanen@uef.fi, https://orcid.org/0000-0002-2785-5346,
[b] University of Eastern Finland, https://orcid.org/0000-0002-5071-8972
[c] University of Jyväskylä, https://orcid.org/0000-0002-2980-0031
[d] University of Jyväskylä, https://orcid.org/0009-0001-6196-5063
[e] University of Eastern Finland, https://orcid.org/0000-0002-1264-3746
[f] University of Jyväskylä, https://orcid.org/0009-0001-1456-1744
[g] University of Jyväskylä, https://orcid.org/0000-0002-2619-8364

ABSTRACT: Evidence-based interventions can develop social and emotional competencies and foster social relationships during early childhood. In this paper, we explored the efficacy, implementation fidelity, and social validity of two systemic, preventive social-emotional learning interventions in Finnish early childhood education and care (ECEC)—ProVaka and Papilio—after their initial nine weeks of implementation. ProVaka is an application of Positive Behavior Intervention and Support that provides general guidelines for the development of social-emotional skills. Alongside general guidelines, the Papilio intervention also utilizes ready-made lesson plans and play-based measures. In this study, 80 ECEC classes from 39 ECEC centers participated in a cluster-randomized controlled trial. We measured educators' feedback, children's behavior, social validity, and implementation fidelity. The results indicate effects particularly on educator-reported child behavior and educators' feedback on certain observed areas when the systemic interventions were

implemented with sufficient fidelity. Further, both interventions seem well suitable for Finnish ECEC context.

Introduction

The importance of developing and acquiring social and emotional competencies in early childhood is widely acknowledged, including emotion recognition and regulation, relationship skills, social awareness, and responsible decision-making (Collaborative for Academic, Social, and Emotional Learning [CASEL], 2013). These skills form the basis of children's interactions during different social encounters, such as with their families, at ECEC centers, with their peers, and in their communities; moreover, these skills are highly significant to academic learning (Blewitt et al., 2018; Weissberg et al., 2017). However, difficulties in the social-emotional development domain are relatively common, and the reported prevalence of clinically significant difficulties has typically varied from 10% to 20% in different studies (Kato et al., 2015). Therefore, strengthening children's social-emotional competencies is very important to their overall development.

Systemic interventions are typically designed in programmatic and manualized forms to help educators follow methods and implement them with fidelity (Voegler-Lee & Kupersmidt, 2011). Children's challenging behavior is thought to be preventable using effective social-emotional teaching strategies (such as identifying emotions and supporting self-regulation; see Fox et al., 2011) and proactive pedagogical practices (such as prompts and systematic praise statements; Fullerton et al., 2009; Stormont et al., 2007). Active instruction, modeling, and reinforcement (Blewitt et al., 2018; Fox et al., 2011) and behavior-specific praise (which specifically acknowledges appropriate behaviors; Dufrene et al., 2012; Floress et al., 2017; LaBrot et al., 2020; Spilt et al., 2016) form a solid basis for effective SEL support. The starting point for change is that, in order to children to change their behavior, educators need to change their own behavior and way of acting. There, the perspective of reacting to unwanted behavior is replaced by proactive teaching appropriate social behavior. While educators learn to recognize and praise children's appropriate social behaviors, they consequently also become better able to identify situations where children need more practice for their behavior. This is also the case concerning both the studied interventions, Papilio and ProVaka.

To be effective, interventions must be implemented with sufficient fidelity, which means they must be implemented as intended (Carr & Horner, 2007). To achieve high fidelity, SEL interventions must align with individual educators' and ECEC communities' attitudes and motivations, as well as the ECEC environment's contextual features—that is, they

must be socially valid (Lawson et al., 2019). When teachers find an intervention socially valid, they are more likely to implement it with fidelity (Strain et al., 2012). To ensure the implementation's best possible cultural and contextual fit, the implementation of the intervention's critical elements is crucial, while simultaneously enhancing educators' autonomous application of the methods (Cook & Odom, 2013; Marchant et al., 2013). In practice, interventions are often implemented less thoroughly than they were designed to be (i.e., they are not implemented with fidelity; Schulte et al., 2009), likely reducing their efficacy (Gitlin & Parisi, 2016). Educators may implement only a few of an intervention's intended components (Määttä et al., 2017) or apply an intervention intended for systemic-level prevention to just a few children (Fox et al., 2011). Consequently, the quality of evidence-based SEL interventions' implementations and the support children receive varies (Fox et al., 2011; Määttä et al., 2017; Weissberg et al., 2017).

In this study, we investigated the effectiveness, social validity, and implementation fidelity of two universal, evidence-based social-emotional learning (SEL) interventions, Papilio and ProVaka. Our investigation focused on change in children's behavior as well as positive and negative feedback the educators give to children. Both interventions aim to promote systemic changes in local ECEC settings to create preventive and supportive environments and practices for effective SEL for all children in an ECEC class. Such universal methods' importance has often been highlighted in previous studies (e.g., Carr & Horner, 2007). ProVaka employs the positive behavior support and intervention (PBIS) approach (Carr & Horner, 2007; Dunlap et al., 2014; see also Heiskanen et al., 2024), and it stems from the tradition of applied behavior assessment (ABA; Shepley & Grisham-Brown, 2019). The Papilio program is a universal, evidence-based intervention developed in Germany to support the development of children's social-emotional competencies and peer relationships while preventing behavioral problems among four-to-seven-year-old children (Mayer et al., 2016). Both of the interventions studied in the current research, ProVaka and Papilio, had been piloted at a Finnish ECEC and reported to be feasible and socially valid in prior studies (Heiskanen et al., 2024; Karhu et al., 2021; Koivula et al., 2020). However, no previous efficacy studies had been conducted on Positive Behavior Interventions and Support in the ECEC of Finland or Nordic countries or on Papilio outside Germany. Both interventions share the fundamental aim of building strong social-emotional support in ECEC by highlighting the importance of shared pedagogical practices, positive feedback, and educators' long-term commitment. While Papilio and ProVaka are both systemic interventions, in practice, they include different practices and processes.

Previous research on ProVaka and Papilio

ProVaka is a systemic approach to preventing and addressing children's challenging behavior and applying the PBIS approach to social-emotional development in ECEC settings. Additionally, ProVaka utilizes evidence-based educational methods to redesign ECEC centers' practices in order to support children's social behaviors and increase educators' use of proactive supportive methods (Dunlap et al., 2014). ProVaka does not include ready-made lesson plans; instead, educators' autonomy under the ProVaka approach enhances its contextual and cultural fit. In the Finnish ProVaka pilot study, the approach was considered suitable for and well-fitted to Finnish ECEC due to its high fidelity and social validity (Karhu et al., 2021). Further, ProVaka's meaningful behavior change is based on creating a shared vision among adults on pedagogical aims and practices as well as organizing an ECEC center's physical and social contexts to help develop children's social skills (Hemmeter et al., 2007). This approach's effectiveness in supporting children's behavior was demonstrated in randomized controlled trials at US schools (Bradshaw et al., 2015; Horner et al., 2010) and in ECEC settings (Hemmeter et al., 2015; Hemmeter et al., 2016; Mahon et al., 2020). Studies at both the school (Kincaid & Horner, 2017) and ECEC (Hemmeter et al., 2015) levels have suggested that effectiveness of PBIS is related to implementation fidelity. Moreover, PBIS's implementation has also been associated with increased organization and behavior management in ECEC classes (Stanton-Chapman et al., 2016), as well as early childhood educators' greater use of systemic practices (Carter et al., 2011).

Papilio's activities were designed specifically to reduce risk factors (e.g., difficulties in peer relationships), help educators address the first signs of behavioral problems (e.g., self-regulation and emotion-regulation difficulties), and foster resilience, inhibitory control, and social-emotional competence (Mayer et al., 2016). Papilio's effectiveness was tested through a large-scale, longitudinal, randomized intervention study in Germany (Scheithauer et al., 2007, 2016). This study revealed that the prosocial behaviors of children in Papilio classes significantly improved, while their problematic behaviors (e.g., hyperactivity, inattentiveness symptoms, and emotional, conduct, and peer relationship problems) were reduced compared to control-group children (see Scheithauer et al., 2007). Papilio was piloted and culturally adapted to Finnish ECEC; ECEC personnel's experiences implementing the program and the resulting cultural adaptations were studied (Koivula et al., 2020). The results suggest that Finnish ECEC educators considered the program to have high fidelity and social validity.

Study purpose

The current study's purpose was to examine the ProVaka and Papilio interventions' effectiveness on children's appropriate and challenging behavior and educators' positive and negative feedback at the class-level. The study was conducted during 2019–2020 in Finnish ECEC classes, and it was cut short due to the COVID-19 pandemic. As a result, the study was limited to assessing the interventions' effectiveness during their initial nine weeks of implementation. However, given the acknowledged importance of a successful initial implementation phase (e.g., McIntosh et al., 2016), our study contributes to efficacy research of SEL interventions by investigating the effectiveness in the crucial initial phases of implementation with the focus on the implementation fidelity and social validity (see also Strain et al., 2012).

The objective of the research project (Towards better support for social-emotional learning in early childhood education and care, KOPA) and the national funding the project was founded on by Finnish National Agency for Education was to produce evidence on a variety of effective and socially valid interventions to support children's SEL in ECEC. The studied interventions Papilio and ProVaka were both relatively new in Finland while the study took place, and although they had already been piloted, the adaptation to Finnish ECEC was still ongoing. Therefore, in this study, we investigate the effectiveness, social validity, and intervention fidelity of these two interventions separately with the aim of providing information of both of them as potential new means for evidence-based SEL learning.

We addressed the following research questions:

- What was the ProVaka and Papilio interventions' fidelity after their first nine weeks of implementation, and how did educators assess their social validity?
- Did intervention-group educators provide more positive feedback and less negative feedback to children than control-group educators?
- Did children's behavior improve (more appropriate behavior and less challenging behavior) compared to control-group children after nine weeks?
- Were the results pertaining to educators' feedback and children's behavior different when only those ECEC classes that achieved high implementation fidelity were investigated?
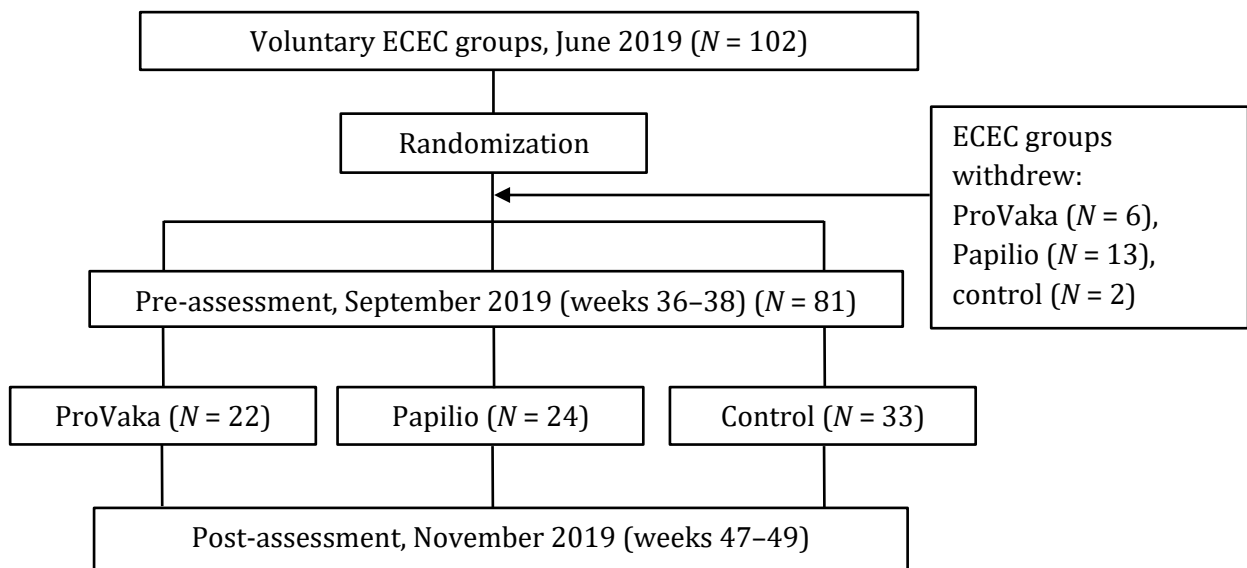
Materials and methods

Research design

This research project was launched in spring 2019. The intervention phase was conducted in autumn 2019 such that all classes could implement their intervention (either ProVaka or Papilio) in nine or ten weeks. To measure the change in children's behavior and educator feedback, pre-assessments were conducted during the two weeks preceding intervention implementation started, where educators in classes were also trained to implement either ProVaka or Papilio intervention.

We employed a cluster-randomized controlled trial design with two intervention groups (ProVaka and Papilio) and one waiting-list control group where each ECEC class was randomized into one of these groups. Participants were randomly assigned to three at the ECEC-center level via block randomization (Suresh, 2011) to avoid the likely spillover effect of several interventions conducted at the same center. Centers run by the same manager (40 administratively independent ECEC centers) were treated as single centers. A random number was assigned to each center using Excel's RAND function, and three same-size groups were created according to the random numbers' order. Two groups were randomly assigned to interventions, while the third group was assigned as a waiting-list control. To enable us to plan the practicalities related to training participants to implement the interventions, we conducted the study's randomization before pre-assessments. Participants were informed of their groups (ProVaka, Papilio, waiting-list control) after pre-assessments, where they also received information about the training schedule.

Between the randomization and pre-assessment, 21 classes from 14 centers withdrew from the study (see Figure 1) due to practical concerns, incomplete recruitment, and shortages of staff or unsuitable time for training. In total, 81 classes from 39 ECEC centers participated in the pre-assessment. After this stage began in September 2019, no classes dropped out of the study. The control group performed as usual during autumn 2019, without training or coaching. Those on the waiting list in the control group were to implement other SEL programs in spring 2020, about which they were informed after the pre-assessments. However, the COVID-19 pandemic made it difficult to systematically conduct these interventions in ECEC centers in spring 2020 and also made it impossible for researchers to conduct planned follow-up measurements, consequently, cutting the study short.

FIGURE 1 Flow chart of the participating ECEC classes and study design

```
┌─────────────────────────────────────────────────────────┐
│        Voluntary ECEC groups, June 2019 (N = 102)         │
└─────────────────────────────────────────────────────────┘
                           │
          ┌─────────────────────────────┐        ┌──────────────────────┐
          │        Randomization         │◄───────│ ECEC groups          │
          └─────────────────────────────┘        │ withdrew:            │
                           │                      │ ProVaka (N = 6),     │
   ┌─────────────────────────────────────────┐    │ Papilio (N = 13),    │
   │ Pre-assessment, September 2019            │    │ control (N = 2)      │
   │ (weeks 36–38) (N = 81)                    │    └──────────────────────┘
   └─────────────────────────────────────────┘
     │                │                 │
┌──────────────┐ ┌──────────────┐ ┌──────────────────┐
│ ProVaka       │ │ Papilio       │ │ Control (N = 33) │
│ (N = 22)      │ │ (N = 24)      │ │                  │
└──────────────┘ └──────────────┘ └──────────────────┘
     │                │                 │
┌─────────────────────────────────────────────────────────┐
│ Post-assessment, November 2019 (weeks 47–49)             │
└─────────────────────────────────────────────────────────┘
```

Study context

While supporting the development of social-emotional skills is central to ECEC worldwide, Finnish ECEC's unique characteristics create a singular context in which to study the systemic SEL interventions' effectiveness. "A Nordic model of ECEC," often defined as different compared to the Anglo-Saxon perspective (Urban et al., 2022), highlights a holistic approach to children's learning and development, learning environments, and systemic approaches to evaluation and assessment. Nordic educational systems emphasize play and focus on assessing pedagogical practices instead of performative child assessments (Urban et al., 2022). Moreover, Nordic educational systems are characterized by strong decentralization, which implies that municipalities and individual teachers play key roles in selecting and implementing pedagogical measures and programs (Urban et al., 2022). Consequently, Finnish ECEC educators should be well equipped to successfully apply holistic and systemic SEL interventions, and they should be accustomed to doing so.

SEL is a central component of Finnish ECEC and an important part of pedagogical process quality (Vlasov et al., 2018), beginning from the national ECEC curriculum that sets the general objectives for SEL pedagogy (Finnish National Agency for Education, 2022). In daily ECEC life, Finnish ECEC strongly emphasizes pedagogy and children's participation (Early Childhood Education Act, 540/2018). However, these elements are not typically implemented with lessons or standardized programs; all daily routines are considered potential pedagogical practices that focus on playful exploration, collaboration, and interaction (Finnish National Agency for Education, 2022). Taking a nap, dressing and

undressing for outdoor activities, and other transitions are also major parts of children's days and, consequently, SEL and teaching. A child's typical day includes adult-led activities, play, and outdoor activities, as well as breakfast, lunch, snacks, and midday nap time. The Finnish eligibility criteria for ECEC personnel[1], maximum class size (21 children with at least three educators), and multi-professional teamwork method support this emphasis. Furthermore, as pedagogical leaders of multi-professional teams, ECEC teachers play a central role in implementing the national curriculum. However, evidence-based SEL interventions are seldom implemented in Finnish ECEC (Määttä et al., 2017).

Participants

Hailing from three urban areas and 12 municipalities in Finland, 102 ECEC classes participated in this voluntary study. Participants were sought from the three geographical areas through personal contact with the people responsible for ECEC at individual municipalities. After municipalities signalled their interest, ECEC centers were selected based on managers' and staff's motivation. Our recruitment information indicated the study's focus on classes including three-to-six-year-old children.

The ECEC classes' mean size during pre-assessment was 19.1 children (SD = 3.9) for ProVaka, 17.7 children (SD = 5.5) for Papilio, and 18.5 children (SD = 6.6) for the control group. The mean proportion of six-year-old children in a class was 21.8% (SD = 36.3) for ProVaka, 19.1% (SD = 36.7) for Papilio, and 47.6% (SD = 48.1) for the control group. Further, the mean proportion of girls in a class was 48.2% (SD = 14.7) for ProVaka, 46.5% (SD = 10.7) for Papilio, and 47.8% (SD = 14.4) for the control group. The mean proportion of ECEC nurses in a class was 39.7% (SD = 16.5) for ProVaka, 39.7% (SD = 18.5) for Papilio, and 43.0% (SD = 20.1) for the control group. The mean number of classes at an ECEC center (ECEC center size) during pre-assessment was 4.6 (SD = 1.4) for ProVaka, 5.0 (SD = 2.8) for Papilio, and 5.7 (SD = 2.5) for the control group. Since we studied the interventions' group-level effects, we did not collect child-specific data.

Ethical considerations

Throughout this study's phases, ethical principles and legislation on operative privacy and data protection (Privacy Protection Act, 2018) were carefully followed. Before the research project began, an ethical pre-review was sought from University of Jyväskylä's Ethical Committee. Research permissions were requested from municipalities, and informed consent was obtained from educators. Educators could withdraw from the

---

[1] In Finnish ECEC, at least a third of staff must possess an ECEC teacher qualification (a bachelor's or master's degree; Early Childhood Education Act, 540/2018). Other staff members must either hold a higher education degree (a bachelor of social services) or have completed a three-year course in vocational education (ECEC nurses).

study at any point, and eight educators from two ECEC classes withdrew before post-assessment. Although all data were collected at the class level and no information on individual children was recorded, children's guardians were informed about the project and given an opportunity to withhold their children's participation; 18 guardians did not allow their children to be observed. Due to COVID-19 restrictions and ECEC lockdown measures in 2020, the waiting-list control group did not undergo an intervention phase as originally planned. Moreover, follow-up data collection in spring 2020 was impossible. The observation data were anonymously recorded. The (ECEC-SQ and social validity) questionnaire data were collected with secure online software (Webropol). All data were saved to a secure university information system.

ProVaka intervention

The ProVaka group was trained over two days for six hours each day. The first day of training took place before the implementation, while the second took place in the middle of the intervention phase. The first training day included all ECEC personnel, and it introduced the ProVaka system's theoretical background and main elements. The second day presented the leadership teams and their roles in supporting the implementation of ProVaka practices in ECEC classes. Further, systematic coaching protocols were developed to support ProVaka's implementation. Coaches, who were either researchers or ECEC practitioners with previous experience on ProVaka, visited each participating ECEC class after each training day (two in-person visits and one online check-in) to support the ProVaka leadership teams. These coaches also maintained contact with the teams and were reachable if necessary.

The ProVaka process at ECEC centers began by establishing leadership teams. The teams' first task was to to brainstorm exercises, plays, or games that could be used to reinforce educators' positive interaction practices. Then, the teams defined systemic-level behavioral expectations for at least one class location or situation along with other educators, children, and families. The teams were coached to focus first on aspects involving the most frequent problem behavior. These behaviours were identified through joint discussions and observations of educators working in the group. The educators were guided to use behavior-specific praise and reinforcement to help teach appropriate behaviors. These behaviors were taught to children through daily activities and during gatherings.

Papilio intervention

Before implementation, the Papilio intervention classes underwent two-day training (a total of 10 hours) from certified Papilio coaches on implementing the program's components. These teachers were instructed to follow the guidelines in the Papilio

teachers' handbook (Mayer et al., 2016), which had been translated and culturally adapted to the Finnish context. All classes received support and advice by telephone twice; the coaches called the teachers and discussed various aspects of the program's implementation.

At the start of the intervention, educators were encouraged to consider important group-management and interaction skills (e.g., providing positive feedback and praise to children, enforcing classroom rules, and supporting positive peer interactions and classroom atmospheres; Mayer et al., 2016). Then, they were instructed to implement the program's first child-focused component, Toys on a Holiday (ToH), which was conducted throughout the intervention. During ToH, one day a week, children played without toys. They played rule games, built huts, and immersed themselves in roleplay.

Next, the teachers began implementing Box Imp Stories alongside the ToH. Each week, they introduced a basic emotion to children (sadness, anger, fear, and joy) through four imps representing these emotions. The children listened to stories of the imps expressing their emotional states twice weekly before emotions were discussed with the children. The classes also created picture boards of the four imps. Every day, to support children's emotional awareness and emotional regulation, they were asked children to place their picture under a corresponding imp, based on how they felt.

Measures

Fidelity

We identified separate fidelity measures for ProVaka and Papilio since the programs' central components varied (see O'Donnell, 2008). The 11 core ProVaka features were: (a) educators' sufficient representation in each center's leadership team; (b) a supportive leader who actively improved children's social behavior; (c) regular leadership-team meetings; (d) informing families and soliciting their feedback; (e) defining behavioral expectations; (f) defining specific, situationally appropriate behaviors; (g) teaching children these behaviors; (h) expressing expectations to both educators and children; (i) ensuring that classrooms display appropriate behaviors; (j) educators' use of positive feedback to help children develop social behaviors; and (k) the delivery of token reinforcements for expected social behaviors.

The Papilio intervention's 10 core features were: (a) weekly ToH; (b) pre- and post-ToH discussions with children; (c) crafting group Box-Imp Story characters with children; (d) discussing all four emotion imps with children; (e) displaying pictures of the emotion imps and discussing the imps with children daily; (f) noticing and praising children's positive behaviors; (g) educators' narrating children's actions and emotions; (h)

systematically establishing and following shared rules; (i) systematically suppressing negative behaviors by offering solutions; and (j) informing families about the method.

All fidelity measures were rated as 0, 1, or 2; 0 indicated a lacking feature, 1 indicated that only some aspects of the feature were implemented, and 2 indicated a feature implemented as expected. The maximum scores were 22 points for ProVaka and 20 points for Papilio. Fidelity was measured for both the ProVaka and Papilio groups at the end of the intervention phase by trained and experienced coaches who respectively interviewed the leadership teams or teachers about their centers' implementations. The fidelity assessment focused on how items had been implemented by an entire ECEC team (teachers, nurses, and others, such as assistants); it did not focus solely on ECEC teachers.

Structured direct observations (SDO)

SDO were used to study children's behavior and educator feedback during a 15-minute observation period with an interval duration of 10 seconds (Wirth et al., 2014; Zakszeski et al., 2017). Children's behavior was observed using momentary time sampling method, which enabled simultaneous partial interval observation of educator feedback during the same activity (Ferguson et al., 2018). Each ECEC class was observed over three days during both the pre- and post-assessment.

The observers were all undergraduate students in educational sciences, and they were not informed of the study's conditions. The observers were trained to collect data in 15-minute sessions. If a session lasted less than 15 minutes, they were told to conduct another observation of the same kind of situation, if possible. They used laptops to record 10-second intervals in an Excel datasheet. The observers' three six-hour-per-day training sessions included: (a) an introduction to structured direct observation; (b) operational definitions of child behavior, educator interaction, and feedback; (c) video recordings using the study's datasheet and discussions of appropriate ratings; and (d) how to assess inter-observer reliability. All observers achieved the preselected criterion of 90% inter-observer reliability.

Educators' Feedback. Two kinds of educator feedback were observed: positive and negative. We operationalized positive educator feedback as an educator verbally acknowledging appropriate child performance (during behaviors, communication, or tasks) either individually or as a group. Positive feedback did not include neutral verbal interactions, such as asking a child to sit or answer a question. We operationalized negative educator feedback as an educator verbally addressing challenging or inattentive child behavior, either individually or as a group. Negative feedback included only spoken feedback indicating displeasure, such as a raised voice, discussing poor performance, or reprimands. Education feedback was observed in educator-directed learning and lunchtimes for 921 sessions (duration: $M = 16.2$, $SD = 6.1$); 468 took place during pre-

assessment (duration: M = 16.7, SD = 7.0), and 453 (duration: M = 15.7, SD = 5.0) during post-assessment.

Children's behavior. Child behavior was coded as either appropriate or challenging. We operationalized appropriate behavior as active participation—for example, responding to an educator's question or playing with appropriate objects. Listening, asking for help, and respectful or active social engagement—such as greeting, initiating social interaction, or responding to social interaction by an educator or peer—also qualified as appropriate behavior. We defined challenging behavior as aggression toward others or property or refusing to do a task or activity, such as disturbing talk, humming, or inappropriate movement (for example, a child wandering around). Six randomly assigned children were observed during each observation session. Observers rotated between these children throughout each session (Dart et al., 2016). Children's behavior was observed in 885 sessions (duration: M = 16.4, SD = 7.0); 452 occurred during pre-assessment (duration: M = 17.3, SD = 7.8), and 433 (duration: M = 15.4, SD = 5.8) during post-assessment.

Inter-observer agreement (IOA). During inter-observer sessions, two observers simultaneously observed the same session using a headphone splitter to hear an identical voice signal from individual headphones. IOA was calculated by dividing the total number of agreements by the sum of agreements and disagreements. IOA for adult feedback was high throughout the study (M = 95.6%, SD = 4.1%, range = 71.4%–100.0%), across both pre- (M = 94.6%, SD = 4.8%, range = 71.4%–100.0%) and post-assessment (M = 96.5%, SD = 3.0%, range = 86.4%–100.0%). Moreover, IOA for children's behavior was high throughout (M = 98.1%, SD = 2.8%, range = 82.2–100.0%), across pre- (M = 97.8, SD = 3.1, range = 82.2–100.0%) and post-assessment (M = 98.5%, SD = 2.4%, range = 86.7–100.0%). The total 1,608 observer sessions included 406 (25.2%) inter-observer sessions. IOA was achieved on educator feedback for 30.3% of observations and on children's behavior for 31.3%.

ECEC situation questionnaire (ECEC-SQ)

Educators evaluated children's behavior by rating 12 common ECEC activities or situations: working in small groups, playing, participating in adult-led situations, tidying up, dressing up, taking off clothes, nap time, being outdoors, mealtime, toileting, taking trips, and engaging in larger events, such as gatherings for entire ECEC centers. The evaluation focused on how problematic educators found children's behavior in each situation. A four-point Likert-type scale (from 1 = *very problematic* to 4 = *very good*) was used. The corresponding questionnaire was based on the School Situation Questionnaire (SSQ; Barkley, 1987) used to rate an individual child's behavior; it has also been adapted and studied for preschoolers (Pelletier et al., 2006). Unlike the original SSQ, educators have used the ECEC-SQ to rate group-level situational variability in problem behaviors.

Moreover, the items were edited to suit Finnish ECEC practices (e.g., no lessons and emphasis on play).

Social validity

The interventions' social validity was investigated using a modified version of the Intervention Rating Profile (IRP-15; Lane et al., 2009). Educators rated their perceptions of an intervention after post-assessment through nine statements (e.g., "It is acceptable to use this intervention in the ECEC," "I would be willing to use this intervention in the ECEC setting," and "This intervention would not result in negative side-effects for children"). The questionnaire used a Likert-type scale (from 1 = *totally disagree* to 6 = *totally agree*).

Variables and statistical methods

Variables

To create outcome variables, observations of each class from similar sessions over three days were combined into one observation score variable for pre- and post-assessment. Multivariate analysis of variance was used to examine whether these observations differed. Since no significant differences arose on consecutive days, the overall observation score variables were calculated for pre- and post-intervention observations and used for further analysis. When several observations addressed educator-directed learning, playtime, or lunch during the day, the results for the same situations were combined.

To create outcome variables from the ECEC-SQ answers, the ECEC-SQ's overall sum variable was calculated by adding the 12 situation-specific variables' mean values. However, exploratory factor analysis (SPSS Statistics 26.0) suggested the ECEC-SQ measures had relevant subdimensions, so confirmatory factor analysis was also conducted (Mplus 8.4) with a weighted least-square mean and variance-adjusted (WLSMV) estimation because the variables were defined as ordinal. This analysis confirmed that a three-factor solution fit the data well (chi-square = 74.40; df = 51; p = 0.02; CFI = 0.99; TLI = 0.99; RMSEA = 0.04; SRMR = 0.05). The three factors used were (1) dressing and undressing, (2) other transitions, and (3) pedagogical situations. They were also deemed to illustrate meaningful dimensions in everyday ECEC life. Further, we calculated Cronbach's alphas for the ECEC-SQ's overall sum variable and subscale variables. The results were 0.82 for the overall sum variable, 0.86 for dressing and undressing, 0.61 for other transitions, and 0.80 for pedagogical situations. The three subscales' correlations ranged from 0.61 to 0.66, suggesting that the individual dimensions and total score were meaningful for this study.

To measure fidelity, fidelity points were transformed into percentages of the maximum score, from 0% to 100%; 70% implementation fidelity has been suggested to be critical for improving educator and child outcomes (McIntosh et al., 2017), and it was used to indicate sufficient fidelity. This approach was chosen instead of using continuous fidelity score as fidelity is not continuous, but rather the assumption is that when fidelity reaches a certain level (e.g., over 70%), the intervention is implemented as it should be. For the social validity scores, a mean variable was constructed based on these scores. Their Cronbach's alphas were 0.91 for ProVaka and 0.94 for Papilio.

For background information, we created variables for children's age distribution (the proportion of six-year-olds in a class), personnel's education level (the proportion of ECEC nurses in a class's staff), the total number of children in a class, the ECEC center's size, and children's gender distribution.

Statistical methods

Since some participants withdrew after randomization (see Figure 1), the baseline equivalence was investigated by calculating the background variables' Hedges's g-values. The intervention groups differed from the control group by children's age distribution and professionals' educational level.

The intervention effects were analyzed via regression models using Mplus 8.7 software with the complex option type in the analysis command. The complex estimation adjusts standard errors in models to avoid biases due to possible clustering effects (Muthén & Muthén, 1998–2012). Since the data were nested in ECEC centers and municipalities, we examined intra-class correlations to determine whether a multi-level model was needed. As a result, we used multi-level models with the ECEC center as a cluster variable. The models' independent variable was membership in an intervention or control group, and each outcome was a dependent variable. In the results, a positive beta coefficient means that an increase in the independent variable (intervention vs. control) is associated with an increase in the dependent variable (measured outcome variables), while a negative beta coefficient means that an increase in the independent variable is associated with a decrease in the dependent variable. All models' baseline values for each outcome were controlled for, and two background variables (children's age distribution and professionals' education level in a class) were included in the analysis as covariates. Because implementation fidelity has been shown to relate to efficacy (Gitlin & Parisi, 2016), we analyzed the interventions' efficacy separately for the classes that achieved sufficient fidelity. Finally, we calculated Cohen's *d*-values to estimate effect sizes (the mean difference within groups divided by the data's pooled standard deviation). We interpreted $0.2 \leq d < 0.5$ as a small effect size, $0.5 \leq d < 0.8$ as a medium effect size, and $0.8 \leq d$ as a large effect size (Cohen, 1988).

Our analysis of social validity and intervention fidelity included calculating means and standard deviations. Finally, we analyzed whether the social validity and intervention fidelity scores correlated with Spearman's correlation coefficients.

Results

All Papilio classes achieved fidelity of over 70% (M = 91.3; SD = 8.4, range = 70%–100%). For the ProVaka intervention, however, fidelity varied: only 17 of 22 classes achieved fidelity over 70% (M = 75.4%, SD = 31.5). Sufficient fidelity indicates that the core components of the interventions are implemented following the given instructions. Both interventions had high social validity (ProVaka: M = 5.30, SD = 0.49; Papilio: M = 4.49, SD = 0.70). No correlation was found between implementation fidelity and social validity (ProVaka: r[22] = .119, p = .599; Papilio: r[24] = .209, p = .338). Our investigation showed that both interventions had both expected and unexpected effects on educator feedback and educator-reported child behavior in some measured situations (see Table 1).

The ProVaka intervention sustained the level of positive educator feedback in mealtime situations, but only when it was implemented with good fidelity (effect d = 1.048, a large effect; see Table 2). Positive educator feedback decreased both in ProVaka group and in control group, but it decreased statistically significantly less when ProVaka was implemented. No statistically significant effects on negative educator feedback arose in the ProVaka classes. However, the Papilio intervention sustained the level of negative educator feedback in educator-led situations compared to control group. Negative educator feedback decreased in both groups, though it fell less for the Papilio group than the control group, whose level fell remarkably. This effect was rather small (d = .447 for educator-led situations). No statistically significant effects on positive feedback were observed in Papilio group.

Based on educators' scores in ECEC-SQ survey, both interventions had statistically significant positive effects on children's behavior compared to control group on some measured areas of children's behavior. In ProVaka classes, children's behavior was maintained at approximately the same level in dressing and undressing situations as well as in other transitions when children's behavior got worse in control group. The effects were rather small (d = .444; d = .484). Further, educators reported that, in Papilio classes, children's behavior was maintained at approximately the same level in other transitions and in pedagogical situations compared to control group, where children's behavior got worse. The effects were also rather small (dressing and undressing d = .255; transitions d = .283; pedagogical situations d = .252). However, similar findings were not obtained for structured direct observations (SDO) of children's behavior since neither the ProVaka nor Papilio interventions affected children's behavior in any observed situations.

TABLE 1 Descriptive statistics on pre-and post-intervention in the ProVaka classes, Papilio classes, ProVaka classes with sufficient fidelity, and control group

Outcome variable	ProVaka						Papilio						ProVaka with sufficient fidelity						Control					
	Pre			Post			Pre			Post			Pre			Post			Pre			Post		
	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N
SDO																								
Positive feedback																								
educator-led	8.028	3.862	24	6.689	3.305	22	6.202	3.188	24	6.707	4.229	24	8.475	4.217	17	6.103	2.424	17	8.316	4.633	33	5.870	4.117	33
mealtime	3.940	2.250	24	3.152	2.249	22	3.436	2.344	24	2.814	1.723	24	4.496	2.384	17	3.726	2.194	17	2.260	1.232	33	1.895	1.134	33
Negative feedback																								
educator-led	3.654	2.748	24	2.138	1.861	22	3.424	2.918	24	2.860	2.294	24	3.171	2.854	17	2.094	1.471	17	3.045	2.472	33	2.156	2.144	33
mealtime	3.029	2.250	24	1.755	1.200	22	3.425	2.693	24	2.176	1.574	24	3.315	2.446	17	1.625	1.177	17	2.748	2.246	33	1.617	1.275	33
Appropriate behavior																								
educator-led	86.443	2.982	24	87.138	2.5318	22	86.385	3.033	24	87.800	2.235	24	86.635	3.098	17	87.227	2.497	17	87.062	2.800	33	87.961	2.208	33
playing	86.784	3.195	24	87.455	3.109	22	87.197	2.095	24	88.208	1.463	22	86.473	3.612	17	87.271	3.413	17	87.564	3.417	33	88.394	2.166	33
Challenging behavior																								
educator-led	1.974	1.710	24	1.533	1.614	22	1.662	1.526	24	1.070	1.288	24	1.885	1.713	17	1.236	1.081	17	1.556	1.979	33	1.192	1.504	33
playing	1.351	1.276	24	1.350	1.691	22	1.499	1.421	24	1.099	1.062	22	1.495	1.275	17	1.221	1.675	17	1.234	1.289	33	1.062	1.728	33
ECEC-SQ																								
total	3.029	.388	55	3.082	.434	51	2.980	.321	60	3.028	.293	46	3.038	.335	40	3.080	.393	39	2.968	.326	83	2.946	.347	81
dressing	3.055	.490	66	3.075	.571	59	2.940	.374	81	2.905	.543	52	3.060	.507	49	3.055	.554	47	3.005	.444	96	2.825	.555	89
other transitions	2.857	.396	66	2.925	.453	59	2.745	.434	81	2.810	.445	52	2.825	.347	49	2.925	.454	47	2.765	.459	96	2.695	.494	88
pedagogical	3.064	.440	55	3.120	.460	51	3.035	.368	60	3.116	.303	46	3.094	.381	40	3.125	.409	39	3.001	.374	83	3.034	.346	82

Note. SDO = structured direct observations, ECEC-SQ = early childhood education and care situation questionnaire

TABLE 2 The intervention effects in intervention groups vs. control group

Outcome variable	ProVaka vs. control		ProVaka with sufficient fidelity vs. control		Papilio vs. control	
	β	p	β	p	β	p
SDO						
Positive feedback						
educator-led situations	.072	.344	.006	.937	.110	.262
mealtime	.184	.173	**.350**	**.012**	.249	.109
Negative feedback						
educator-led situations	.035	.743	.037	.712	**.206**	**.046**
mealtime	.092	.458	.018	.890	.225	.202
Appropriate behavior						
educator-led situations	-139	.266	-.128	.348	-.011	.926
playing situations	-.141	.344	-.164	.295	-.013	.927
Challenging behavior						
educator-led situations	.046	.681	-.029	.814	-.012	.916
playing situations	.047	.780	.016	.925	-.032	.834
ECEC-SQ						
total	.141	.162	.158	.212	**.253**	**.000**
dressing and undressing	**.249**	**.007**	**.256**	**.008**	.199	.056
other transitions	**.256**	**.011**	**.274**	**.012**	**.240**	**.001**
pedagogical situations	.078	.357	.087	.435	**.236**	**.000**

Note. SDO = structured direct observations, ECEC-SQ = early childhood education and care situation questionnaire

Discussion

In this study, we conducted a cluster-randomized controlled trial in Finnish ECEC classes. Two evidence-based SEL interventions, ProVaka and Papilio, were implemented at typical Finnish ECEC centers, and their effectiveness was studied after nine weeks of implementation. Both interventions aimed to help develop children's social-emotional skills and behavior. Additionally, we investigated the interventions' effects on educator feedback and children's behavior, and their implementation fidelity and social validity.

Educators' ratings indicated a slight improvement in children's behavior (see Table 1), though the study's observations did not indicate this improvement. Based on educators' ratings, both interventions positively affected children's behavior during transitions. Moreover, ProVaka positively affected children's behavior when dressing and undressing, while Papilio positively affected and the children's behavior in pedagogical situations. However, such group-level effects were not evident from the study's observations, though children's behavior was observed to have improved (see Table 1).

This study used two different measurements for children's behavior—direct behavior observations and educator ratings—since it aimed to reveal situations in ECEC classes and children's behavior more comprehensively across different situations. This methodological choice also revealed differences between observations and educators' ratings. Our findings support previous studies in which teacher questionnaires have often assessed children's behavior more positively than direct observers (Hemmeter et al., 2016). However, many activities were not observed, and improvements in these situations were evident for educators but not reflected in observations. Furthermore, while implementing SEL interventions, educators (who monitor children throughout the day in various situations) may also become more aware of children's behavior and recognize minor changes before observers.

Our study reveals that both interventions affected amounts of educator feedback maintaining the level of feedback in the intervention groups compared to the control group, where the level of both types of feedback decreased. ProVaka positively affected favorable educator feedback in mealtime when it was implemented with sufficient fidelity (> 70% of a maximum score). Most ProVaka classes also started their development work by focusing on mealtime, so this effect arose in the situations where the intervention was most vigorously applied. ProVaka sustained the level of positive educator feedback compared to control group, and this group's feedback level fell less than the control group's. Although mealtime is a relatively limited type of activity, it is significant in Finland due to being repeated at least three times a day, and most ProVaka classes sought to focus on this situation. Therefore, ProVaka influenced the situations considered challenging to educators and meaningful in the Finnish ECEC context regarding SEL. This finding highlights contextual differences in interventions' implementation, suggesting that implementers must always integrate new models into organizational practices (Nelen et al., 2020; Willemse et al., 2023).

Unexpectedly, Papilio affected similarly to the amount of negative educator feedback in Papilio and control groups: While the level of negative feedback in educator-led situations decreased as expected, it fell even more remarkably in the control group. Our research does not provide direct answers to this unexpected outcome. The Papilio intervention's rapid start and schedule may have made fully adopting all Papilio components difficult. These classes' feedback levels remained high generally (see Table 1). The training's focus on feedback and educators' focus on giving feedback generally may have sustained the level of negative feedback. Moreover, the selected measures may not have successfully reflected a change in positive feedback levels.

In the study, ProVaka received high social validity ratings while the implementation fidelity was low in a few ProVaka classes that did not get properly started with

intervention implementation. This may indicate that the implementation of ProVaka may have been challenging because of the factors related to the organization of development work—for example, those that are not directly attributable to the intervention itself. It seems that establishing the ProVaka leadership team and organizing the working time for the team was an impracticable task with the given timetable in certain ECEC centers. In contrast, the educators in all the Papilio classes that participated to the study performed the given tasks with excellent fidelity (M = 91.3%). Papilio's components, including ready-made lesson plans and materials, were deemed to suit the Finnish ECEC context, and the group achieved good social validity ratings, although some of those ECEC classes randomized into Papilio group also withdrew from the study due to the unsuitable time for training and other practical issues. When Papilio intervention has gotten started, it seems that the tasks' clarity and the ready-made lesson plans supported the implementation's rapid start in this short intervention period. While the high demands of planning the concrete steps of the ProVaka intervention required extensive time, thereby delaying the start.

However, it needs to be noted that the fidelity measures were individually tailored to both interventions, where each intervention's key components were emphasized. The interventions also involved different training amounts and arrangements; Papilio training took place before the implementation, while ProVaka training occurred before and during the implementation. Because of this difference and fundamental dissimilarities between the two interventions (e.g., Papilio's faster start and ready-made lesson plans versus ProVaka's strong autonomy and creation of structures and practices), they are not comparable. Both fidelity and social validity included collective evaluations from all ECEC staff members, including teachers and other professionals, though these measurements may differ internationally. Further, in our study, fidelity and social validity did not correlate, which challenges previous studies' findings that high social validity supports implementation fidelity (Strain et al., 2012).

Limitations

This study faced a few limitations. Its intervention groups were relatively small (24 Papilio classes and 22 ProVaka classes), which limited its power to identify statistically significant changes. Additionally, the project's funding limited both the number of centers that could be recruited and the study's duration; this research focused on interventions' effectiveness during their first nine weeks of implementation. Therefore, given the relatively small number of participating ECEC classes (N = 81), we cannot assume that all our findings may be generalized to all ECEC centers in Finland or elsewhere. However, regular ECEC classes participated in this study, representing a wide geographical area— 12 municipalities and three different regions of Finland. This scope adds to the study's credibility since, in Finnish ECEC, local agreements, curricula, and educators' pedagogical

freedom cause varying interpretations of national ECEC regulations and concomitant pedagogical practices. A longer study duration would have made supporting long-term implementation and following up on interventions' progress and maintenance after nine weeks possible. The study was originally planned to continue with follow-up in spring 2020, including measuring the similar issues as in the pre- and post-intervention measurements approximately three and six months after the start of the implementation. However, lockdowns due to the COVID-19 pandemic precluded this possibility.

Finally, the level of children's problematic behavior was already rather low in the study's pre-assessment (see Table 2). However, the voluntary participation of specific ECEC centers can be seen as a reflection of their need to develop SEL-related practices despite children's observed good behavior overall. Since ProVaka and Papilio are systemic, preventive interventions, extending their research and implementation to entire ECEC centers and all children in classes with good overall behavior was greatly significant. Future research with longer intervention periods is required, allowing the programs time to develop and become an integrated part of ECEC classes' everyday pedagogy. Since implementing and sustaining evidence-based pedagogical practices requires time and professional development (Yeung et al., 2016), longer interventions and follow-ups are important to obtain more reliable evidence on intervention effects and their sustainability.

Conclusions

Our study's results indicate that both ProVaka and Papilio seem well suitable for supporting SEL of children in ECEC. However, in developing interventions, a balance must be found between cooperation with educators to gain high social validity and using direct instructions or lesson plans, at least initially. Seemingly, implementation fidelity—particularly during the start—can be supported by clearer instructions and more time to prepare to initiate a program. Generally, the ProVaka and PBIS approaches require careful planning to fit with other contextual functions. Despite this, in this study, the effects occurred in just nine weeks of implementation. Although more intervention effects could be found with a longer implementation, the initiation phase's effectiveness offers promising, encouraging information for educators who conduct interventions.

Finally, as the changes in educators' behavior are the major driver of changes in children's behavior, the results of this study give a promising view on the potential of the studied interventions. With these kinds of holistic approaches, it is possible to contribute to the creation of an ECEC culture that meets the individual needs of all children. Getting started with the development work required that educators changed their own mindset while supporting children's socio-emotional skills with systematic teaching and acknowledging appropriate social behavior. As children's behavior perceived as challenging causes stress and emotional exhaustion among teachers, interventions to reduce such behavior may

benefit teachers as well children (Embse et al., 2019; Schaack et al., 2020). When an implementation is conducted with good fidelity, changes in children's behavior are quickly noticeable, and an intervention can also affect professionals' own activities, such as providing feedback.

Acknowledgments

References

Barkley, R. A., & Edelbrock, C. (1987). Assessing situational variation in children's problem behaviors: The home and school situations questionnaires. *Advances in Behavioral Assessment of Children and Families, 3*, 157–176.

Blewitt, C., Fuller-Tyzkiewich, M., Nolan, A., Bergmeier, H., Vicary, D., Huang, T., McCabe, P., McKay, T., & Skouteris, H. (2018). Social and emotional learning associated with universal curriculum-based interventions in early childhood education and care centers. A systematic review and meta-analysis. *JAMA Network Open, 1*(8), 1–19. https://doi.org/10.1001/jamanetworkopen.2018.5727

Bradshaw, C., Waasdorp, T., & Leaf, P. (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*(2), 546–557. https://doi.org/10.1037/a0037630

Carr, E., & Horner, R. (2007). The expanding vision of positive behavior support: Research perspectives on happiness, helpfulness, hopefulness. *Journal of Positive Behavior Interventions, 9*(1), 3–14. https://doi.org/10.1177/10983007070090010201

Carter, D., Van Norman, R., & Tredwell, C. (2011). Program-wide positive behavior support in preschool: Lessons for getting started. *Early Childhood Education Journal, 38*(5), 349–355. https://doi.org/10.1007/s10643-010-0406-0

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Collaborative for Academic, Social, and Emotional Learning (CASEL) (2013). *Effective social and emotional learning programs. Preschool and elementary school edition.* Collaborative for Academic, Social, and Emotional Learning.

Cook, B. G., & Odom, S. L. (2013). Evidence-based practices and implementation science in special education. *Exceptional Children, 79*(2), 135–144. https://doi.org/10.1177/001440291307900201

Dart, E. H., Radley, K. C., Briesch, A. M., Furlow, C. M., & Cavell, H. J. (2016). Assessing the accuracy of class-wide direct observation methods: Two analyses using simulated and naturalistic data. *Behavioral Disorders, 41*(3), 148–160. https://doi.org/10.17988/BD-15-49.1

Dufrene, B. A., Parker, K., Menousek, K., Zhou, Q., Harpole, L. L., & Olmi, D. J. (2012). Direct behavioral consultation in head start to increase teacher use of praise and effective instruction delivery. *Journal of Educational and Psychological Consultation, 22*(3), 159–186. https://doi.org/10.1080/10474412.2011.620817

Dunlap, G., Kincaid, D., Horner, R. H., Knoster, T., & Bradshaw, C. (2014). A comment on the term "positive behavior support." *Journal of Positive Behavior Interventions, 16*(3), 133–136. https://doi.org/10.1177/1098300713497099

Early Childhood Education and Care Act (540/2018). Finlex. www.finlex.fi/fi/laki/alkup/2018/20180540

Embse, N., Ryan, S., Gibbs, T., & Mankin, A. (2019). Teacher stress interventions: A systematic review. *Psychology in the Schools*, *56*(8), 1328–1343. https://doi.org/10.1002/pits.22279

Ferguson, T., Briesch, A., Volpe, R., Donaldson, A., & Feinberg, A. (2018). Psychometric considerations for conducting observations using time-sampling procedures. *Assessment for Effective Intervention, 44*(1), 45–57. https://doi.org/10.1177/1534508417747389

Finnish National Agency for Education. (2022). *Core curriculum for early childhood education and care* (Regulation OPH-700-2022)*. Finnish National Agency for Education. https://www.oph.fi/sites/default/files/documents/National%20core%20curriculum%20 for%20ECEC%202022.pdf

Floress, M. T., Beschta, S. L., Meyer, K. L., & Reinke, W. M. (2017). Praise research trends and future directions: Characteristics and teacher training. *Behavioral Disorders, 43*(1), 227–243. https://doi.org/10.1177/0198742917704648

Fox, L., Hemmeter, M., Snyder, P., Perez Binder, T., & 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*(3), 178–192. https://doi.org/10.1177/0271121411404440

Fullerton, E. K., Conroy, M. A., & Correa, V. I. (2009). Early childhood teachers' use of specific praise statements with young children at risk for behavioral disorders. *Behavioral Disorders, 34*(3), 118–135. https://doi.org/10.1177/019874290903400302

Gitlin, L. N., & Parisi, J. M. (2016). Are treatment effects real? The role of fidelity. In L.N. Gitlin & S. J. Czaja (Eds.), *Behavioral intervention research: Designing, evaluating, and implementing* (pp. 213–240). Springer Publishing Company.

Heiskanen, N., Karhu, A., Savolainen, H., & Närhi, V. (2024). Implementing positive behaviour support in Finnish early childhood education and care: Leadership teams' perspectives. *European Journal of Special Needs Education, 39*(2), 265–280*. https://doi.org/10.1080/08856257.2023.2207057

Hemmeter, M. L., Fox, L., Jack, S., & Broyles, L. (2007). A program-wide model of positive behavior support in early childhood settings. *Journal of Early Intervention, 29*(4), 337–355. https://doi.org/10.1177/105381510702900405

Hemmeter, M., Hardy, J., Schnitz, A., Adams, J., & Kinder, K. (2015). Effects of training and coaching with performance feedback on teachers' use of pyramid model practices. *Topics*

in Early Childhood Special Education, 35(3), 144–156.
https://doi.org/10.1177/0271121415594924

Hemmeter, M., Snyder, P., Fox, L., & Algina, J. (2016). Evaluating the implementation of the pyramid model for promoting social–emotional competence in early childhood classrooms. *Topics in Early Childhood Special Education, 36*(3), 133–146. https://doi.org/10.1177/0271121416653386

Horner, R. H., Sugai, G., & Anderson, C. M. (2010). Examining the evidence base for school-wide positive behavior support. *Focus on Exceptional Children, 42*(8), 1–14. https://doi.org/10.17161/foec.v42i8.6906

Karhu, A., Heiskanen, N., & Närhi, V. (2021). Kohti sosioemotionaalisia taitoja ja lasten osallisuutta tukevaa toimintakulttuuria—ProVaka-toimintamallin pilottivaiheen analyysi [Towards a working culture that supports socio-emotional skills and child participation—pilot study analysis of the ProVaka support model]. *Journal of Early Childhood Education Research, 10*(1), 83–113. https://journal.fi/jecer/article/view/114151

Kato, N., Yanagawa, T., Fujiwara, T., & Morawska, A. (2015). Prevalence of children's mental health problems and effectiveness of population-level family interventions. *Journal of Epidemiology, 25*(8), 507–516. https://doi.org/10.2188/jea.JE20140198

Kincaid, D., & Horner, R. (2017). Changing systems to scale up an evidence-based educational intervention. *Evidence Based Communication Assessment and Intervention, 11*(3–4), 99–113. https://doi.org/10.1080/17489539.2017.1376383

Koivula, M., Laakso, M-L., Viitala, R., Neitola, M., Hess, M., & Scheithauer, H. (2020). The adaptation and implementation of the German social-emotional learning programme Papilio in Finland: A pilot study. *International Journal of Psychology, 55*(S1), 60–69. https://doi.org/10.1002/ijop.12615

LaBrot, Z. C., Dufrene, B. A., Whipple, H., McCargo, M., & Pasqua, J. L. (2020). Targeted and intensive consultation for increasing head start and elementary teachers' behavior-specific praise. *Journal of Behavioral Education*, *29*(4), 717–740. https://doi.org/10.1007/s10864-019-09342-9

Lane, K., Kalberg, J., Bruhn, A., Driscoll, S., Wehby, J., & Elliott, S. (2009). Assessing social validity of school-wide positive behavior support plans: Evidence for the reliability and structure of the primary intervention rating scale. *School Psychology Review, 38*(1), 135–144. https://doi.org/10.1080/02796015.2009.12087854

Lawson, G. M., McKenzie, M. E., Becker, K. D., Selby, L., & Hoover, S. A. (2019). The core components of evidence-based social–emotional learning programs. *Prevention Science, 20*(4), 457–467. https://doi.org/10.1007/s11121-018-0953-y

Mahon, D., Gunning, C., Holloway, J., & Lydon, H. (2020). Implementing positive behavior support within preschool settings: Group functional assessment and CW-FIT. *Journal of Positive Behavior Interventions, 22*(3), 167–180. https://doi.org/10.1177/1098300719890065

Marchant, M., Heath, M. A., & Miramontes, N. Y. (2013). Merging empiricism and humanism: Role of social validity in the school-wide positive behavior support model. *Journal of Positive Behavior Interventions, 15*(4), 221–230. https://doi.org/10.1177/1098300712459356

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.* [Papilio: Theory and basics. A programme for day-care centres to prevent behavioural

problems and promote social-emotional skills. A contribution to addiction and violence prevention]. Papilio Verlag.

Mcintosh, K., Massar, M. M., Algozzine, R. F., George, H. P., Horner, R. H., Lewis, T. J., & Swain-Bradway, J. (2017). Technical adequacy of the SWPBIS tiered fidelity inventory. *Journal of Positive Behavior Interventions, 19*(1), 3–13. https://doi.org/10.1177/1098300716637193

McIntosh, K., Mercer, S. H., Nese R. N. T., Strickland-Cohen, M. K., & Hoselton, R. (2016). Predictors of sustained implementation of school-wide positive behavioral interventions and supports. *Journal of Positive Behavior Interventions, 18*(4), 209–218. https://doi.org/10.1177/10983007155997

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide* (7th ed.). Muthén & Muthén.

Määttä, S., Koivula, M., Huttunen, K., Paananen, M., Närhi, V., Savolainen, H., & Laakso, M.-L. (2017). *Lasten sosioemotionaalisten taitojen tukeminen varhaiskasvatuksessa* (Raportit ja selvitykset 2017:17). [Supporting children's social-emotional skills in early childhood education]. Finnish National Agency for Education. http://oph.fi/julkaisut/2017/lasten_sosioemotionaalisten_taitojen_tukeminen_varhaiskasvatuksessa

Nelen, M. J., Willemse, T., Oudheusden, M. A. van, & Goei, S. (2020). Cultural challenges in adapting SWPBIS to a Dutch context. *Journal of Positive Behavior Interventions, 22*(2), 105–115. https://doi.org/10.1177/1098300719876096

O'Donnell, C. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research, 78*(1), 33–84. https://doi.org/10.3102/0034654307313793

Pelletier, J., Collett, B., Gimpel, G., & Crowley, S. (2006). Assessment of disruptive behaviors in preschoolers: Psychometric properties of the disruptive behavior disorders rating scale and school situations questionnaire. *Journal of Psychoeducational Assessment, 24*(1), 3–18. https://doi.org/10.1177/0734282905285235

Privacy Protection Act (2018). https://www.finlex.fi/fi/laki/ajantasa/2018/20181050

Schaack, D., Le, V., & Stedron, J. (2020). When fulfillment is not enough: Early childhood teacher occupational burnout and turnover intentions from a job demands and resources perspective. *Early Education and Development, 31*(7), 1011–1030. https://doi.org/10.1080/10409289.2020.1791648

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) [Promoting social-emotional skills at preschool age: Results of the Augsburg longitudinal study to evaluate the Papilio® primary prevention programme (ALEPP)]. In T. Malti & S. Perren (Eds.), *Soziale kompetenz bei kindern und jugendlichen—entwicklungsprozesse und fördermöglichkeiten. 2. bearb. & erw. Auflage* [Social competence in children and young people—development processes and support options. 2nd revised and expanded edition] (pp. 155–176). Kohlhammer.

Scheithauer, H., Bondü, R., Niebank, K., & Mayer, H. (2007). Prävention von Verhaltensproblemen und Förderung prosozialen Verhaltens bei hoch-und niedrig-Risikokindern im Kindergarten: Erste Ergebnisse der Augsburger Längsschnittstudie zur Evaluation des Programms Papilio [Prevention of behavioural problems and promotion of prosocial behaviour in high- and low-risk children in kindergarten: First results of the Augsburg

longitudinal study to evaluate the Papilio programme]. *Praxis der Rechtspsychologie, 17*(2), 376–391.

Schulte, A., Easton, J., & Parker, J. (2009). Advances in treatment integrity research: Multidisciplinary perspectives on the conceptualization, measurement, and enhancement of treatment integrity. *School Psychology Review, 38*(4), 460–475.

Shepley, C., & Grisham-Brown, J. (2019). Applied behavior analysis in early childhood education: An overview of policies, research, blended practices, and the curriculum framework. *Behavior Analysis in Practice, 12*(1), 235–246. https://doi.org/10.1007/s40617-018-0236-x

Spilt, J. L., Leflot, G., Onghena, P., & Colpin, H. (2016). Use of praise and reprimands as critical ingredients of teacher behavior management: Effects on children's development in the context of a teacher mediated classroom intervention. *Prevention Science, 17*(6), 732–742. https://doi.org/10.1007/s11121-016-0667-y

Stanton-Chapman, T., Walker, V., Voorhees, M., Snell, M., Houchins, D., & Oakes, W. (2016). The evaluation of a three-tier model of positive behavior interventions and supports for preschoolers in Head Start. *Remedial and Special Education, 37*(6), 333–344. https://doi.org/10.1177/0741932516629650

Stormont, M. A., Covington Smith, S., & Lewis, T. J. (2007). Teacher implementation of precorrection and praise statements in head start classrooms as a component of a program-wide system of positive behavior support. *Journal of Behavioral Education, 16*(3), 280–290. https://doi.org/10.1007/s10864-007-9040-3

Strain, P., Barton, E., & Dunlap, G. (2012). Lessons learned about the utility of social validity. *Education & Treatment of Children, 35*(2), 183–200. https://doi.org/10.1353/etc.2012.0007

Suresh, K. P. (2011). An overview of randomization techniques: An unbiased assessment of outcome in clinical research. *Journal of Human Reproductive Sciences, 4*(1), 8-11. https://doi.org/10.4103/0974-1208.82352

Urban, M., Reikerås, E., Eidsvåg, G. M., Guevara, J., Saegø, J., & Semmoloni, C. (2022). *Nordic approaches to evaluation and assessment in early childhood education and care: Final report.* Nordic Council of Ministers.

Vlasov, J., Salminen, J., Repo, L., Karila, K., Kinnunen, S., Mattila, V., Nukarinen, T., Parrila, S., & Sulonen, H. (2018). *Guidelines and recommendations for evaluating the quality of early childhood education and care* (Publications 5:2019). Finnish Education Evaluation Centre.

Voegler-Lee, M. E., & Kupersmidt, J. B. (2011). Intervening in childhood social development. In P. K. Smith & C. H. Hart (Eds.), *The Wiley-Blackwell handbook of childhood social development* (2nd ed.)(pp. 605–626). Wiley-Blackwell.

Weissberg, R. P., Durlak, J. A., Domitrovich, C. E., & Gullotta, T. P. (2017). Social and emotional learning. Past, present, and future. In J. A. Durlak, C. E. Domitrovich, R. P. Weissberg & T. P. Gullotta (Eds.), *Handbook of social and emotional learning. Research and practice* (pp. 3–19). Guilford.

Willemse, T. M., Goei, S. L., Boei, F., & de Bruïne, E. J. (2023). School-wide positive behavior interventions and support in Dutch schools for special education. *European Journal of Special Needs Education, 38*(3), 424–439. https://doi.org/10.1080/08856257.2022.2120331

Wirth, O., Slaven, J., & Taylor, M. (2014). Interval sampling methods and measurement error: A computer simulation. *Journal of Applied Behavior Analysis, 47*(1), 83–100. https://doi.org/10.1002/jaba.93

Yeung, A. S., Craven, R. G., Mooney, M., Tracey, D., Barker, K., Power, A., Dobia, B., Chen, Z., Schofield, J., Whitefield, P., & Lewis, T. J. (2016). Positive behavior interventions: The issue of sustainability of positive effects. *Educational Psychology Review, 28*(1), 145–170. https://doi.org/10.1007/s10648-015-9305-7

Zakszeski, B., Hojnoski, R., & Wood, B. (2017). Considerations for time sampling interval durations in the measurement of young children's classroom engagement. *Topics in Early Childhood Special Education, 37*(1), 42–53. https://doi.org/10.1177/0271121416659054