Predictors of the implementation of Tuning in to Kids in Norwegian Kindergartens

Pia Kolbræk Eikeland and Aurora Norebø Omre

Acknowledgement

Oslo, May 2021
Pia K. Eikeland

Declaration

This thesis is submitted to the University of Oslo in completion of the requirements for a master thesis in psychology. I, Pia Kolbræk Eikeland and I, Aurora Norebø Omre, declare the work in this thesis to be our own, except when acknowledged.

20.05.2021   20.05.2021

Pia Kolbræk Eikeland   Aurora Norebø Omre

Note. The data in this thesis was part of a randomized controlled trial implementing and evaluating the socio-emotional program Tuning in to Kids for Kindergarten Teachers. We have contributed to the project by translation of one the applied questionnaires (Weiner et al., 2017), and by suggesting questions to measure sustainability and penetration (part of the implementation outcomes). We also aided in collecting data at T3 (June 2020), in which the project did a practical assessment of 100 children's emotional competence via an iPad task. We and two other master students delivered and trained kindergarten staff around Oslo in the iPad-task, as teachers had to do the task with the children themselves due to Covid to reduce contact with more people than necessary. The project has delivered the program to the wait list control group, in which we translated the presentations. We also participated in the training. We assisted by calling some of the leaders to identify respondent's unit belonging. In addition to cleansing and matching the measures we used ourselves, we contributed to the project with data cleansing and matching in the CLASS Toddler (waitlist data) and Supervision responses, both of which we did not use ourselves.

Abstract

Authors: Pia Kolbræk Eikeland and Aurora Norebø Omre

Title: Predictors of the implementation of Tuning in to Kids in Norwegian Kindergartens

Supervisors: Egil Nygaard and Sophie Havighurst

Background: Identifying factors that affect the implementation quality and implementation outcomes are important, as better implementation is associated with better effects of an intervention. The aim of the present study was to research which factors predicted implementation outcomes in the implementation of the universal socio-emotional learning program Tuning in to Kids for Kindergarten Teachers in Norway, and whether this relationship was mediated by implementation quality. The predictors examined in this study were the teachers': education level, length of employment in a kindergarten, perception of kindergarten functioning, psychological distress, emotion dysregulation, emotion socialization, and emotional support (observed).

Method: Data were collected in a bigger cluster randomized controlled intervention trial. We used quantitative questionnaires and observational data from 344 teachers and leaders in 22 intervention kindergartens. Predictors were measured at baseline (August 2019), implementation quality part way through the implementation period (March 2020) and implementation outcome at the end of implementation (July 2020). We performed general linear model, linear mixed model and multilevel mediation analyses to account for and examine clustering effects at unit level.

Results: Teachers and units with a better functioning kindergarten, teachers with more supportive emotion socialization and units consisting of teachers with more mean experience, had better implementation outcomes. Better implementation quality mediated the relationship between the predictors: better kindergarten functioning (teacher & unit level), better emotion socialization (teacher & unit level) and better observed emotional support (unit level), and better implementation outcomes. Individuals and units with better implementation quality had significantly better implementation outcome, for all predictor models.

Conclusion: These results indicate that an increased focus on factors that aid in improving the implementation quality may result in better implementation outcomes, and thus better effects from an intervention. Our results showed that the functioning of the kindergarten, factors related to the program techniques and experience of the teachers affected the implementation.

Key words: Implementation, implementation quality, implementation outcome, predictors, mediator

Table of contents

List of abbreviations

ACK= Acknowledge the child's emotions

AIC= Akaike Information Criterion

CCNES= Coping with Children's Negative Emotions Scale Teacher Version

CFA= Confirmatory factor analysis

CFI= Comparative Fit Index

CFIR= Consolidated Framework for Implementation Research

CLASS= Classroom Assessment Scoring System

CLASS pre-k= Classroom Assessment Scoring System pre-k

CLASS toddler= Classroom Assessment Scoring System toddler

CMIN/DF= Minimum discrepancy per degree of freedom

COVID-19= Coronavirus disease of 2019

CTNES= Coping with Toddlers Negative Emotions Scale

CTNES-T= Coping with toddler's negative emotions scale - teacher version

DERS-18= Difficulties with emotion regulation scale 18

DF= Degrees of freedom

DIS= Distracting emotion socialization reactions

DR= Teacher Distress Reactions

DSM-IV= Diagnostic and Statistical Manual of Mental disorders 4th ed.

EC = Emotional competence

EE= Expressive Encouragement

EFA= Exploratory factor analysis

EFR= Emotion Focused Responses

EM= Expectation Maximization

ES= Emotion socialization

GR= Granting Children's Wishes

H1= Hypothesis 1

H2= Hypothesis 2

IQQ-39= Implementation quality questionnaire 39

IQQ-64= Implementation quality questionnaire 64

K6= Kessler 6

M= Mediator

MR= Minimizing Reactions

NSD = Norsk senter for forskningsdata

PFR= Problem Focused Responses

PR= Punitive Reactions

RMSEA = Root mean square error approximation

SES= Socioeconomic status

SD= Standard deviation

SE= Standard error

TIK = Tuning in to Kids

TIK-KT= Tuning in to Kids for Kindergarten Teachers

TSD = Tjeneste for Sikker Datalagring

X= Predictor

χ^2 = Chi-square (χ^2) goodness of fit

Y= Outcome (implementation outcomes)

List of Tables and Figures

Predictors of the implementation of Tuning in to Kids in Norwegian Kindergartens

Implementation of an intervention into real life settings is challenging and have often led to lesser effects of the intervention (Durlak & DuPre, 2008). An intervention should be implemented well to be effective (Proctor et al., 2010), thus examining factors that influence implementation quality and implementation outcomes are important. In this thesis we explored predictors of implementation outcomes and if this relationship was mediated by implementation quality. We examined this in an implementation of the emotion coaching program Tuning in to Kids for Kindergarten Teachers (TIK-KT), which aims at increasing children's emotional competence through teacher emotion coaching. Before the implementation of this program, information about the teachers and the workplace were gathered to see if it predicted implementation outcomes. The predictors examined were the teachers': education level, experience, perception of the kindergarten functioning, psychological distress, emotion dysregulation, emotion socialization and emotional support. Research on mediation and multilevel relationships in the implementation field is relatively new (Lewis et al., 2020). We aim to contribute to this field by increasing the knowledge by exploring if there exist a mediating relationship between predictors (pre), implementation quality (part way), and implementation outcomes (end), and if these relationships are present on both an individual and group level. More knowledge on this may contribute to decrease the challenges of implementation and lead to better implementation outcomes (and thus intervention effect) in the future.

Emotional competence

Developing emotional competence (EC) is an important developmental task in childhood (Raver, 2002; Tetzchner, 2012). EC involves the ability to correctly understand and perceive others and one's own emotions, to situationally and culturally appropriately express emotions, to be able to inhibit and regulate the experience and expression of emotions (Eisenberg, Cumberland, & Spinrad, 1998).

EC is important for children to help them in the realm of emotions and several favorable factors such as social and academic competence have been positively linked to EC (Izard et al., 2001). Stronger emotional intelligence (a similar term to EC) has been linked to better physical health, psychosomatic health, and mental health (Martins, Ramalho, & Morin, 2010). Deficits in components of EC has been linked to higher ratings of aggressive behavior in children (Bohnert, Crnic, & Lim, 2003), and poor emotional understanding has been linked to behavioral problems in children (Cook, Greenberg, & Kusche, 1994). Lower levels of emotional intelligence have been linked to internalizing problems such as depression, anxiety,

social anxiety, somatic complaints, post-traumatic symptoms, and obsession-compulsion (Salavera, Usán, & Teruel, 2019).

The development of a child's emotions starts during the first year of life when the production, understanding and regulation of basic emotions begins (Tetzchner, 2012). As children start to develop their emotions during their first years in life, aiding the development of EC should start this early as well. One significant and modifiable factor that contributes to EC is emotion socialization, specifically, how parents and caregivers react to the child's emotions, how emotions are modelled and how emotions are talked about (Eisenberg et al., 1998; Havighurst, Wilson, Harley, & Prior, 2009; Morris, Silk, Steinberg, Myers, & Robinson, 2007). This socialization can come from different people including parents, siblings, other nonparent adults such as grandparents, peers and teachers (Eisenberg et al., 1998).

Emotion Socialization Parenting Styles

According to Gottman and DeClaire (1997), parents meta-emotion beliefs (how people feel about emotions and their expression) differ, and this affects the way parents respond to their own and their child's emotions. Gottman and colleagues found that parents could be grouped into four parenting styles in terms of how they responded to their children's emotions. With the first style, emotion dismissing, parents do not acknowledge their child's emotion, may ignore, dismiss, or minimize the child's negative emotions. With the second style, emotion disapproving, parents do not approve and can be critical of their child displaying negative emotions. They may also punish or scold their child if they display emotions. The third style, laissez -faire, parents show empathy and acceptance of their child's emotions, however, they fail to guide their child or set limits around their behavior. With the last style, emotion coaching, parents will view their child displaying emotions as a time for intimacy and teaching. They acknowledge, empathize with and validate the child's feelings, show support, allow them to experience their emotions, but also help them name their feelings and give the child guidance on how to problem solve and may set limits around their behavior. They are also not afraid of showing their feelings in front of their child as they value the purpose and power of emotions (Gottman & DeClaire, 1997). Gottman and DeClaire (1997) found that of the four parenting styles, emotion coaching was the most optimal response to children's emotions. Children of the emotion coaching parents were more emotionally intelligent (similar term to EC) (Gottman & DeClaire, 1997).

Tuning in to Kids

The universal socio-emotional learning program *Tuning in to Kids: Emotionally Intelligent Parenting Program* (TIK; Havighurst & Harley, 2007) aims at increasing the emotional competence in children by aiding their parents in developing better emotion socialization. TIK is delivered as a group program spanning over six weeks with 2 hour sessions each week, and optionally, two additional follow-up sessions over the next two months (Havighurst & Harley, 2007). TIK is based on research by Gottman and DeClaire (1997), Gottman, Katz, and Hooven (1996, 1997) and Eisenberg et al. (1998).

TIK targets parent emotion socialization with children in the early years when children's social and behavioral functioning is most malleable (Havighurst et al., 2009). This involves teaching parents to become more aware of and able to regulate one's own emotion, how to react and respond when children are emotional, and to connect empathically and help their child calm down. The parent will then be able to teach the child to recognize different emotions, accept emotions and know how to regulate them, respond to, and handle their emotions. Part of this is teaching parents five steps of emotion coaching; 1) Becoming more aware of the child's feelings, especially those with low intensity, 2) viewing the child's emotion as an opportunity for closeness and learning, 3) communicating the parents understanding and acceptance of the feeling, 4) helping the child to describe their feelings and 5) if necessary, help the child solve problems and/or set limits around their behavior.-The parents learn that all feelings are acceptable, but not all behavior is. Using emotion coaching techniques in about 30- 40% of the opportunities that arises is sufficient for effectful emotion coaching (Havighurst & Harley, 2007).

Studies conducted in Australia (origin country) have found a significant effect of TIK in improving parents' emotion coaching, decreasing parents' emotion dismissiveness and reducing child problem behavior (Havighurst et al., 2009; Havighurst, Wilson, Harley, Prior, & Kehoe, 2010). Cross cultural studies from Iran, Germany, and Hong Kong have found similar results; parents who received a TIK intervention were more emotion coaching and less dismissive (Aghaie Meybodi, Mohammadkhani, Pourshahbaz, Dolatshahi, & Havighurst, 2019; Otterpohl, Buchenau, Havighurst, Stiensmeier-Pelster, & Kehoe, 2020).

Applying Tuning in to Kids in Kindergartens

As children spend a lot of time in the kindergarten, teachers and childcare workers are likely to have considerable influence over their emotional learning. In 2020, 92% of all the 1–5-year-olds in Norway were enrolled in kindergartens. Ninety-six percent of all children attending kindergarten spent 41 hours or more in the kindergarten *(Statistisk Sentralbyrå (SSB), 2021)*. Thus, applying an intervention like TIK to kindergartens could have a

significant impact on children's emotional learning as it allows a much greater reach with many children simultaneously, compared to working with each parent of each child. As children's production, understanding and regulation of emotion start to develop during the first year of life (Tetzchner, 2012), kindergartens are perfectly placed to address this developmental need, and could be more favorable targets than schools.

Universal prevention programs in kindergartens. A meta-analysis on the effects of universal social and emotional programs used in kindergartens and schools found positive effects on social- and emotional skills, behavior, attitudes, and academic performance, compared to control students (Durlak et al., 2011). Durlak et al. (2011) concluded that teachers and school staff were effective in implementation of socio-emotional learning programs. This suggests kindergartens are a fitting arena to apply a program like TIK.

Teacher emotion socialization. Teachers might be an additional source of emotion competence learning for children. A study showed that teachers' emotion socialization positively contributed to children's emotional knowledge by not having punitive reactions (e.g., "tell child to straighten up or they'll have to sit out"). The punitive reaction effect was in accordance with parental studies on emotion socialization. However, the same study showed that more minimizing reactions (e.g., "tell them they are overreacting") to their negative emotions contributed to more emotional knowledge in the children. This were not in accordance with parental findings. The authors suggested that minimizing reaction showed a opposite direction in their study as this reaction in a classroom might communicate that to regulate your emotions at school is important, not that the emotions are not important. However, this effect was not further examined and need more research (Denham, Ferrier, & Bassett, 2020).

TIK in schools and kindergartens. Havighurst and Harley (2013) delivered a six-week professional training in TIK to early childcare workers in a pilot study implementing TIK in Australian kindergartens. The childcare workers reported positive outcomes including increased emotion coaching skills (Havighurst & Harley, 2013). In Norway, a master thesis performing a pilot study with a mixed methods design implemented TIK, delivered as a 2.5 hour training version and a follow up session after eight weeks, for kindergarten teachers (Eriksen, 2018). They found reduced emotion dismissiveness in teachers and an association between teachers reported emotion coaching and closeness in relation to children, but not increased emotion coaching.

More research on TIK in Norwegian kindergartens were needed to establish if these results may be replicated, preferably with a TIK program specifically developed for

kindergartens, such as TIK-KT. TIK-KT is a modified version of TIK, to be delivered to kindergarten teachers with the aim of learning them emotion coaching skills to apply to the children. The modifications included changes in the delivery method (see method section). It was also necessary to evaluate whether possible effects of TIK-KT in kindergartens, were due to factors related to the studies, the intervention, or the process of implementation.

Implementation

Empirical evidence for an intervention does not mean the given intervention will be implemented well (Proctor et al., 2010). Implementation refers to the transition period between an organizations decision to adopt the intervention, and the routine use of the intervention when stakeholders learn, practice, and commit to it (Damschroder et al., 2009; Klein & Sorra, 1996). Implementation research is a transdisciplinary field studying the systematic uptake of evidence-based practices into real-world settings and aims to decrease the "research to practice gap" by focusing on effective implementations and thus improve the quality of services provided (Eccles & Mittman, 2006; Gottfredson & Gottfredson, 2002).

There have been found lesser effects of interventions when they are implemented in real-life settings such as kindergartens or schools compared to more controlled trials (Durlak & DuPre, 2008). For schools, the norm is for implementations to fail (Gottfredson & Gottfredson, 2002; Lyon & Bruns, 2019; Waschbusch, Breaux, & Babinski, 2019). This may apply to kindergartens as well. It is important to measure implementation to know whether the success or failure of the intervention may be attributed to the intervention itself or to the influence of the implementation (Gottfredson & Gottfredson, 2002; Lyon & Bruns, 2019; Waschbusch et al., 2019). Effect sizes from studies which were well implemented were three times higher (Durlak & DuPre, 2008), indicating the importance of focusing on implementation.

Implementation quality

Implementation quality is in this study considered as a broad term including factors influencing a successful implementation. Implementation quality includes the construct implementation determinants, which are used in some studies (in this paper the term implementation quality will be applied). Implementation determinants are the factors (e.g., characteristics of individuals involved) that influence change at the multiple levels of implementation, and directly influence the outcomes of the implementation (Lyon & Bruns, 2019).

The Consolidated Framework for Implementation Research (CFIR) is a synthesized framework based on 19 existing theories on implementation. CFIR identifies factors

influencing implementation on multiple levels (Damschroder et al., 2009), and is used as a theoretical framework to define and measure implementation quality in the present thesis. The framework explains what works where and why, and in multiple contexts. CFIR identified five main domains (levels) that influenced the success of the implementation: the outer setting, the inner setting, characteristics of individuals involved, characteristics of the intervention and the process of implementation (Damschroder et al., 2009).

The outer setting refers to influences outside the organization where the implementation occurs, like political, economic, or cultural contexts (Pettigrew, Woodman, & Cameron, 2001). According to Damschroder et al. (2009), constructs under this domain includes external policies and incentives, and peer pressure (pressure to implement an intervention from competing organizations). This level can be difficult to influence in an implementation (Lyon & Bruns, 2019).

The inner setting includes the factors immediately related to organizational context where the intervention is implemented (Pettigrew, 1985 as cited in Pettigrew et al., 2001). According to Damschroder et al. (2009), constructs in this domain include the organization's culture and implementation climate. The organization's culture refers to an organization's norms, values, and assumptions about the organization (Gershon, Stone, Bakken, & Larson, 2004). Implementation climate refers to the extent to which an implementation will be rewarded within the organization (Damschroder et al., 2009), and the degree to which individuals are receptive for change (Greenhalgh, Robert, Macfarlane, Bate, & Kyriakidou, 2004). Which variables are considered in the inner or outer setting are dependent on the context of the implementation (Damschroder et al., 2009). Variables on the inner level strongly influence the interventions success (Lyon & Bruns, 2019).

Characteristics of individuals involved with the intervention and/or implementation refers to the power individuals have to influence the process or other individuals and thus influence the implementation. Individuals refers to those targeted by the intervention and others that are affected (e.g., leaders and teachers when one is implementing a program into kindergartens). Constructs under this domain include knowledge and beliefs that the individuals hold about the intervention, self-efficacy, and the individual's identification with the organization (Damschroder et al., 2009).

The implementation process refers to all planned or unplanned processes aiming towards an effective implementation. Constructs under this domain include four essential activities in the implementation process: planning the implementation, engaging involved individuals, executing the implementation according to plan, and evaluating and reflecting

upon the implementation process. Individuals from both the inner and outer setting may promote this process (Damschroder et al., 2009).

Characteristics of the intervention refers to the fit of a specific intervention for the organization (Damschroder et al., 2009; Lyon & Bruns, 2019). Constructs under this domain include adaptability (how much the intervention can be adapted to meet local needs), relative advantages of the intervention compared to other alternatives, and perceived difficulty of delivering the intervention (Damschroder et al., 2009). Interventions usually need to be adapted to the new setting to be a good fit. If not, individuals often resist using the intervention. Interventions have "core components" that are the essential components necessary to gain the desired effect of the intervention. These core components should not be altered as that may decrease the effect of the intervention. Non-core components that are not essential to gain the effect of the intervention may be altered to fit local needs (Molloy, Moore, Trail, Van Epps, & Hopfer, 2013). For example, the five emotion coaching steps may be core components, but the delivery of the training may be tailored to local needs without decreasing the effect of the intervention.

Implementation outcomes

Implementation outcomes are influenced by different aspects of the implementation like the implementation quality (i.e., implementation determinants) (Lewis, 2017, as cited by Lyon & Bruns, 2019). Implementation outcomes are defined as the effect of actions performed to start using and integrating a new program or intervention into an organization (Proctor et al., 2010). Implementation outcomes explain what we aim to improve in practice and explain theoretically and are thus one of the most critical factors in implementation science (Lewis et al., 2015). Implementation outcomes precede and impact school/student outcomes and indicate how successful the implementation has been (Lyon & Bruns, 2019; Proctor et al., 2010). School/student outcomes can be defined as the effect of the intervention, such as increased emotion coaching in teachers or increased emotion competence in children after an intervention (Lyon & Bruns, 2019).

Proctor et al. (2010) suggested a model with a core set of eight implementation outcomes organized as a heuristic taxonomy: Acceptability, adoption, appropriateness, feasibility, fidelity, implementation cost, penetration, and sustainability. The present thesis focused on acceptability, appropriateness, feasibility, fidelity, penetration, and sustainability for the implementation outcomes.

Acceptability is defined as the stakeholders' (individuals involved in the implementation of the intervention) perception that the intervention is satisfactory and

agreeable, i.e., the constructs evaluate the personal perception of fit of the intervention (Proctor et al., 2010; Weiner et al., 2017). Acceptability may change throughout the intervention, for instance, rates might be higher at the beginning of the intervention, compared to later when the stakeholder has more experience (Weiner et al., 2017).

Appropriateness is defined as the perceived fit of the intervention to a particular problem, like increasing teachers' emotion coaching skills, practicing how to deliver the intervention or how compatible the intervention is with the organization (Proctor et al., 2010). Thus, the construct evaluates whether the individual thinks the intervention is a good fit to the organization and deals with their problems appropriately (Weiner et al., 2017).

Feasibility is defined as the degree to which an intervention can be carried out successfully in an organization (Karsh, 2004). The construct evaluates if the individual perception of whether the intervention can be performed with the available resources (Weiner et al., 2017). Resources, recruitment issues or participation rates are potential examples of factors that may affect the feasibility (Proctor et al., 2010).

Fidelity is defined as how well an intervention was implemented compared to the original program's design. This includes the dimensions adherence (the extent to which something, like the use of the program, occurred as planned), quality of delivery by those using the intervention (like a teacher), degree of exposure to the intervention and how involved and engaged participants were (Dane & Schneider, 1998; Mihalic, 2004).

Penetration is defined as how well an intervention is integrated into an organization. It may e.g., be calculated by the number of providers using an intervention divided by the number of providers able to use the intervention (Proctor et al., 2010).

Sustainability is defined as how well an intervention is maintained or institutionalized within the organization and its usual operations. The operational indicators of sustainability are maintenance of the program's benefits, institutionalization of the program in a setting and building capacity in the setting (Rabin, Brownson, Haire-Joshu, Kreuter, & Weaver, 2008). Penetration and sustainability may be correlated as higher penetration may lead to long-term sustainability (Proctor et al., 2010).

Research on Predictors of Implementation Quality

Factors related to the intervention, school and individual characteristics are linked to implementation quality (Domitrovich et al., 2008). School-based prevention programs often have low quality of implementation (Gottfredson & Gottfredson, 2002). School climate/organizational level factors have been shown to lead to better implementation quality as measured by the factor's: teacher's beliefs (e.g., about the interventions effectiveness) and

attitude towards the intervention (Beets et al., 2008; Little, Pokhrel, Sussman, & Rohrbach, 2015). Teachers higher in the predictor emotional support were more engaged in intervention training. Teachers previous use of intervention practices, education level and years of teaching experience were not associated with teacher's engagement in intervention training (Wanless, Rimm-Kaufman, Abry, Larsen, & Patton, 2015). Domitrovich et al. (2008) suggested that psychological functioning of teachers, including stress and depression, probably impact implementation quality, especially when the intervention is perceived as an additional burden.

Research on Implementation Outcomes

Measuring Implementation Outcomes. New sciences like implementation science often have poor quality instruments and measurement gaps (Martinez, Lewis, & Weiner, 2014). A review study by Lewis et al. (2015) identified 104 instruments across the eight implementation outcomes constructs. Half assessed acceptability, 19 examined adoption and less than 10 looked at the other outcome constructs. One instrument had psychometric strength on all six evidence-based psychometric assessment criteria, but the majority had no information regarding predictive validity and responsiveness (Lewis et al., 2015). This makes comparisons across different studies challenging.

Predictors. Predictors of implementation outcomes have been examined using different ways to measure implementation. Individuals that were trained in the interventions practice less recently had better implementation dosage (Exner-Cortens, Spiric, Crooks, Syeda, & Wells, 2020). Teachers with less work experience and teachers with already existing teaching methods compatible with the intervention (of a program aimed at reducing substance abuse in schools) implemented more of the program (Rohrbach, Graham, & Hansen, 1993). Contrary to these results, Wanless et al. (2015) found teachers years of work experience and rated use of intervention practices previous to the intervention could not be related to observed fidelity of implementation in schools, nor did education level and observed emotional support affect the observed fidelity.

Relationship between implementation quality factors and outcomes

Individuals with more positive attitudes (implementation quality component) towards the intervention had higher fidelity towards an intervention (Exner-Cortens et al., 2020). Teachers with stronger self-efficacy had better implementation (Rohrbach et al., 1993). However, Wanless et al. (2015) did not find an effect of self-efficacy on observed fidelity.

Several studies have investigated how factors related to organizational factors (like leader support, high quality training, supervision, climate) have influenced implementation.

Studies indicate that better organizational factors are related to better implementation (Exner-Cortens et al., 2020; Gottfredson & Gottfredson, 2002; Malloy et al., 2014; Molloy et al., 2013). School climate also predicted better implementation on a school level (as opposed to individual level) (Kallestad & Olweus, 2003). The authors concluded that teachers were the main agents of change regarding adoption and implementation of the intervention. However, the concepts applied in these studies have been defined and measured in different ways.

Mechanisms in implementation

Research to identify mechanisms in implementation science aims at better understanding how implementation works to more precisely be able to target strategies that may increase implementation (Lewis et al., 2020). A mechanism explains how implementation affect implementation outcomes (Kazdin, 2007; Lewis et al., 2020). A recent review investigated mechanisms of implementation including mediating effects of implementation quality (referred to as implementation determinants in the review) and group level mechanisms (Lewis et al., 2020). A mediator (M) is a variable that intervenes and transmits the effect of the independent variable X on the dependent variable Y (Lewis et al., 2020; MacKinnon, Fairchild, & Fritz, 2007). Twenty-seven studies investigating mediating mechanisms applied an aspect of implementation quality as the mediating variable (M), and aspects of implementation outcomes (like adoption) as the dependent variable (Y). The independent variable (X) was either another aspect of implementation quality, or the applied implementation strategy (like training) (Lewis et al., 2020). Beets et al. (2008) found that the mediator beliefs (about responsibility to teach the intervention) and attitudes towards the intervention mediated the effect between school climate and adherence to the intervention. The mediator perceived effectiveness of interventions and funding to the school mediated the relationship between organizational support and the adoption (Little et al., 2015). Wanless et al. (2015) found indications that teacher's engagement in training of the intervention mediated the effect between observed emotional support and observed fidelity of implementation later.

Few studies have examined multilevel relationships. Of 46 reviewed studies, 12 investigated mechanisms across multiple levels of implementation, such as both individual level and organizational level (Lewis et al., 2020). According to Bronfenbrenner's Ecological Systems Theory, individuals are affected by factors related to themselves and environmental systems they interact with, which also influence each other (Bronfenbrenner, 1979). CFIR suggested factors on different levels affect implementation and interact (Damschroder et al., 2009). Considering this, further examination of individual and group level effects is important to further examine this mechanism in implementation.

The present study

Studies that have explored predictors of implementation used different methods to measure the intervention or the implementation, which make generalizations and comparisons difficult. Studies investigating implementation quality often focus on one or a few aspects of implementation quality, and the concept is used in different ways. Based on theories and previous research of implementation, it is expected that implementation quality predicts implementation outcomes. Few studies have investigated predictors of the whole concept of implementation quality and implementation outcomes and not just subfactors of these. This also applies for the relationship between implementation quality and implementation outcome.

Literature reviews indicates there may be mediating mechanisms in implementation, but there is currently not enough research with high quality design, enough empirical studies, or sound analysis to conclude and find sound evidence (Lewis et al., 2020; Williams, 2016). None of the previous studies had examined mediating effects with a predictor as independent variable. There was also need for examining implementation mechanisms at multiple levels as implementation quality includes factors on multiple levels that affect implementation outcomes. In the current study, we aim to address this gap in knowledge by exploring predictors of both implementation quality and implementation outcomes, and whether the relationship between the predictors and implementation outcomes are mediated by implementation quality.

The Present Study: **Aim and research questions**

This study aimed to explore the implementation of TIK-KT in Norwegian FUS kindergartens (a private kindergarten organization) in terms of implementation quality and implementation outcomes and which factors predict these. Specifically, the main question examined in this paper was: *What factors in terms of the kindergarten environment or teacher background/functioning predict the implementation outcomes of Tuning in to Kids in the FUS kindergartens, and is this relationship mediated by implementation quality?* The factors that are explored as predictors in this study are teachers: education level, length of employment in a kindergarten, the perception of the kindergartens functioning, psychological distress, emotion dysregulation, emotion socialization, and emotional support. An illustration of these relationships can be seen in Figure 1.

Education Level/ Socio-Economic Status

Education level is an indicator of socio-economic status (SES). Research suggests that higher SES is linked to better emotional intelligence in trainee teachers (Kant & Lenka,

2013). Thus, this should make it easier for the teachers to implement the TIK techniques and may result in them having a more positive attitude to the program. However, previous research by Kallestad and Olweus (2003) did not find an effect of education on implementation of an intervention program by teachers in schools. Wanless et al. (2015) did not find an effect of teachers' education level on fidelity. Suggesting that education might not influence the level of implementation. Considering these conflicting results, investigating the relationship between education and/or SES and implementation outcomes is important.

Kindergarten experience

A study conducted by Rohrbach et al. (1993) found that when teachers had less years of teacher experience, this was correlated with a higher degree of implementation, in schools that received a school-based prevention program for substance abuse. The length that teachers have been employed may make it harder to change the way they conduct their work. Thus, the length of teachers' employment in the kindergarten may have resulted in them being more set in their ways with how they interact with children and thus impact their openness to learning new skills taught in the TIK-KT program. However, Yassien (2018) found that teachers with more experience had significantly higher emotional intelligence than those with less experience. Thus, teachers with more experience could have better emotional competence which may make the TIK-KT techniques easier to implement for the teacher with more experience. The effect of experience on implementation was explored in a study by Kallestad and Olweus (2003) who found no significant effect of experience on teachers' implementation of a bullying prevention program in schools. Considering these conflicting results investigating the relationship between teachers' experience and implementation outcomes further is important.

Perception of kindergarten functioning

How the teachers and the leaders perceive the environment, the management and their information flow is expected to be an important indicator for how well the change in the workplace will be completed. To facilitate the implementation of the program it is expected to be important that the management and their information flow is adequate. Studies have found organizational climate in schools to impact teachers' implementation of a social-emotional and character development program (Malloy et al., 2014). Organizational climate has also been shown to significantly predict organizational commitment (Berberoglu, 2018). If the participants are more committed to their workplace, they might be more committed to the intervention. Thus, a better functioning kindergarten is expected to have greater ease in implementing the intervention.

Psychological distress

Zhang et al. (2016) found a significant negative relationship between emotional intelligence and psychological distress. Teachers with more psychological distress may struggle more with implementation as they may have lower levels of emotional intelligence. Their struggles might influence the liking, acceptance, and implementation of the program. Domitrovich et al. (2008) suggested that the psychological functioning of teachers, including stress and depression probably have an impact on implementation quality. Psychologically distressed participants might not have enough extra energy and experience the implementation as a burden. Considering this investigating the relationship between teachers' psychological distress and implementation outcomes is important.

Emotion dysregulation

As previously mentioned, emotion regulation is a part of EC (Eisenberg et al., 1998). Emotion regulation involves the individual's acceptance, understanding and awareness of emotions, impulsive behavior control and ability to behave in accordance with desired goals, and ability to use appropriate situational emotion regulation strategies (Gratz & Roemer, 2004). Poorer teacher emotion regulation (dysregulation) may affect their capacity to emotion coach as they might be flooded by their own emotions and will not be able to behave in accordance with the goal of being emotion coaching. Therefore, emotion dysregulation is expected to be associated with a lower level of implementation quality and implementation outcomes.

Emotion socialization

TIK targets the emotion socialization of the individual. The teachers that already possessed some of the skillset that the intervention required could have greater ease at implementing the program which could influence their liking, accept, and use of the program. Those who do not have good emotion socialization beforehand might find that learning about the intervention creates cognitive dissonance. Cognitive dissonance is a phenomenon that occurs when attitudes, beliefs and behavior is conflicting (Festinger, 1957). Learning about the intervention and its claims about the effect might make those who were worse in emotion socialization feel discomfort regarding their own actions. To reinstate congruence in their attitude, beliefs, and behavior they might reject the TIK-KT information as non-valid. However, it could also motivate them to change their behavior to make it congruent with their attitudes and beliefs. However, Pettigrew et al. (2001) did not find an effect of having some of the practices that the intervention used did not predict fidelity.

Emotional support

Part of the emotion coaching style in TIK-KT is showing support and acceptance of children's emotions. Thus, a higher level of emotional support may make the TIK-KT techniques easier to use and thus a better liking and accept of the program. As with the emotion socialization, lacking emotional support abilities could create a cognitive dissonance regarding the program which could either create resistance towards the program or promote behavior change. However, Wanless et al. (2015) showed that emotional support did not influence fidelity directly, however, emotional support influenced fidelity through engagement in training.

Hypotheses

Hypothesis 1 (H1): We hypothesize that the kindergarten environment and the following teacher factors; education, teaching experience, psychological distress, emotion dysregulation, emotion socialization, and emotional support are related to implementation outcomes.

Hypothesis 2 (H2): We hypothesize that the relationship between the predictors and the implementation outcomes is mediated by implementation quality (Figure 1).

Figure 1
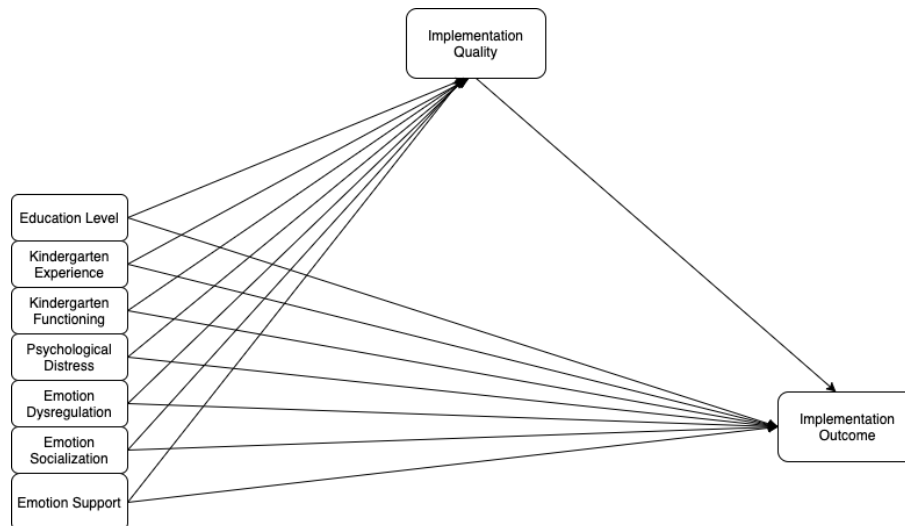
The Expected Relationship Between Predictors, Implementation Quality and Outcomes

Note. Illustration of the expected mediation relationship between the predictors, implementation quality and implementation outcome.

Method

Participants and Design

The present study was part of a bigger cluster randomized controlled intervention trial conducted in 49 Norwegian FUS kindergartens located in Oslo, Bergen, and Trondheim. The

kindergartens were chosen as they were closest to the cities and then grouped into clusters based on proximity and whether they already had a working relationship. Then they randomized the clusters into intervention (22 kindergartens) or waitlist control (27 kindergartens) using a research randomizer. Data were collected from kindergarten leaders, teachers, parents, and children pre- (T1), part way through- (T2) and post (T3) intervention. There were 44 leaders participating in the two-day professional training and 480 teachers that participated in the TIK-KT training (see below). In this thesis, data from teachers and leaders in the intervention kindergartens were included; N=344 (21 leaders and 323 teachers). Information about these participants can be seen in Table 1.

Table 1

Responses, job title (%), gender, education (SD), birthyear (SD), and experience working in a kindergarten (SD) of the participants in the three different questionnaires.

	T1	T2	T3
Responses	287	171	193
Job title:			
Leaders (%)	21 (7)	11 (6)	21 (10)
Teachers (%)	266 (92)	160 (93)	172 (89)
Gender:			
Male	33	17	25
Female	254	98	168
Education M (SD)	4.06 (1.5)	4.14 (1.6)	4.04 (1.6)
Birthyear M (SD)	1981.88 (9.8)	1981.97 (8.95)	1980.62 (9.51)
Experience M (SD)	10.82 (11.03)	10.43 (7.25)	11.28 (7.44)

Note T1= first questionnaire distributed, T2= second questionnaire distributed, T3= third questionnaire distributed, experience=years individuals worked in a kindergarten, M= mean, SD= standard deviation. Education scale: 1 (primary and/or secondary education), 2 (vocational high school level education), 3 (high school), 4 (shorter education/coursing after high school), 5 (higher education of 1-3 years), 6 (higher education of 5-6 years). Gender, age, education, and experience information were based on information from T1.

Drop out analyses

An independent sample t-test and a chi square test were completed to see if there was a difference between the participants who only answered the first questionnaire and those who answered both the first and last questionnaires (see Appendix A). There was a significant difference between the two groups with four of the measures. The participants that completed

both the first and last questionnaires were older, had more teaching experience, scored lower on the non-supportive emotion socialization measure and had less psychological distress.

Procedure

The intervention: TIK-KT

At the start of the study (March 2019), all kindergarten leaders received an initial briefing of the intervention (3 hours). The kindergarten leader and lead pedagogue completed a professional training in the program which spanned over two days (14 hours total) in December 2019. This training gave the leaders and lead pedagogues the skillset to assist the teachers in the use of the program with the children and how to complete supervision sessions with the teachers. This training was provided by the program author Sophie Havighurst.

In January and February 2020, the kindergarten teachers received a 7-hour TIK-KT training in how to use the program and regulate their emotions. This training was provided by Norwegian TIK trainers with an early education master's degree or clinical psychology degree.

The leaders and lead pedagogues that received the professional training were instructed to hold fortnightly supervision sessions with the teachers at their kindergarten. These sessions included discussion about the challenges the teachers met, and review and practice of the TIK-KT techniques using examples provided by the teachers.

In March 2020, the leaders and lead pedagogues received a half day booster session with the Norwegian TIK trainers. This session was used to address challenges and issues they encountered when supporting the teachers in the use of the techniques. The kindergartens received support manuals in online format for all the teachers.

Data Collection

Data were primarily gathered using quantitative questionnaires made in "Nettskjema" (an online questionnaire service) and distributed to participants through e-mails. CLASS observations assessing emotional support were completed in each unit for all kindergartens. Information were gathered at baseline (T1, August 2019), part way (T2, March/April 2020) and at follow-up approximately 9 months later (T3, May/July 2020). Descriptive data and information about the predictors, including CLASS observations, were collected at T1. The implementation quality was measured at T2. Implementation outcomes was measured at T2 (fidelity) and at T3 (sustainability, penetration, feasibility, appropriateness, and acceptability).

Measures

The predictors education level and length of employment data were collected with one question each. Length of employment (i.e., experience) was measured in years worked in a

kindergarten. Education levels answer options were: 1. Primary and secondary school, 2. Vocational school, 3. High school, 4. Shorter education (like courses/ training) after high school, 5. Higher education (1-3 years), 6. Higher education (5-6 years). The rest of the predictors, the mediator and outcomes needed measuring tools or several questions to capture the information.

Perception of Kindergarten Functioning

To assess the teachers' and leaders' perception of the kindergarten functioning nine questions were developed for the purpose of this study. These questions collected information on the quality of the following; their relationship to other employees, their relationships to the parents of the children attending the kindergarten, relationship to the nearest leader, their ability to tend to a child's emotional needs, the diversity of the employees, information from the leader regarding changes, the organizations facilitation for improvement and development, the parents and children's opportunities to affect the kindergartens and the kindergartens strategi for habituating new children. These were answered on a 7-point Likert-type scale ranging from very bad (1) to very good (7). A higher score indicated a perception of better kindergarten functioning; no items were reversed. An exploratory factor analysis (EFA) was completed to assess use of subscales or a mean score. For an overview of the questions in the original language, see Appendix B.

Psychological Distress

Kessler psychological distress scale (K6) was used to examine psychological distress. K6 contains six questions and is a shorter version of the Kessler psychological distress scale (K10). The measure asked how often during the last 30 days the participants felt: nervous, hopeless, restless or fidgety, so depressed nothing could cheer them up, that everything was an effort and worthless. These were answered on a 5-point Likert-type scale ranging from all of the time (1) to none of the time (5). The measure is calculated to a sum score where a higher score indicate less distress. The K6 has been found to have excellent internal consistency reliability (Kessler et al., 2002). It has also shown to discriminate with precision between non-cases and community cases of the DSM-IV disorders (Kessler et al., 2002). The K6 was translated to Norwegian, and back translated to English again for the purpose of this study (see Appendix C for translated version).

Emotion Dysregulation

The Difficulties with Emotion Regulation Scale (DERS-18) was chosen to measure teachers' emotion dysregulation. DERS-18 contains 18 questions and is a shorter version of the original 36-item DERS measure (Gratz & Roemer, 2004; Victor & Klonsky, 2016). The

results from a comparison of DERS-18 and the original DERS demonstrated that this version had similar performance as the original DERS. It showed both excellent validity and reliability (Victor & Klonsky, 2016).

DERS-18 has six domains: "lack of awareness of one's emotions (awareness), lack of clarity about the nature of one's emotions (clarity), lack of acceptance of one's emotions (nonacceptance), lack of access to effective emotion regulation strategies (strategies), lack of ability to engage in goal-directed activities during negative emotions (goals), and lack of ability to manage one's impulses during negative emotions (impulse)" (Victor & Klonsky, 2016, pp. 582-583). The questions were answered on a 5-point Likert-type scale. The answers range from almost never (1) to almost always (5), was calculated to a sum score and a higher score indicated worse emotion regulation (emotion dysregulation) (Victor & Klonsky, 2016).

Emotion Socialization

The Coping with Toddlers Negative Emotions Scale – Teacher Version (CTNES-T) was used to measure the teachers' emotion socialization. CTNES-T is based on *the Coping with Toddlers Negative Emotions Scale* (CTNES; Spinrad, Eisenberg, Kupfer, Gaertner, & Michalik, 2004) and *the Coping with Children's Negative Emotions Scale Teacher Version* (CCNES Teacher version; Fabes, Eisenberg, and Bernzweig (2000). CTNES-T was developed by project leader Sophie Havighurst for the purpose of this study by adding additional kindergarten relevant events; adding two subscales to measure aspects of the TIK-KT intervention; and by modifying one subscale to be consistent with what was taught in the intervention. See Appendix D for an overview of modifications in English, and Appendix E for the questionnaire in Norwegian. CTNES-T was translated to Norwegian and back translated to English.

CTNES-T presented 12 hypothetical situations describing scenarios with children that kindergarten teachers may encounter during their workday, and then eight or nine possible reactions for each situation (nine of the situations had eight reactions, three of the situations had nine reactions). Each reaction represents a subscale of the measure. The reactions presented were: Teacher Distress Reactions (DR) which reflect teachers experience of distress to children's negative affect; Granting Children's Wishes (GWR) refers to the degree to which the teachers do as the child wants, to avoid or remove the problem or negative emotion; Punitive Reactions (PR) refers to the degree to which the teachers react in a way that is punishing and decrease their own need to deal with the children's negative emotions; Minimizing Reactions (MR) refers to the degree to which the teacher minimizes the child's emotions; Expressive Encouragement (EE) refers to the degree that teachers encourage

children to express their negative emotions; Problem Focused Responses (PFR) refers to the degree to which teachers help the child solve the problem behind the negative emotions; Emotion Focused Responses (EFR) refers to the degree which teachers respond with strategies to help the child feel better by affecting the child's negative emotions; Acknowledge the child's emotions (ACK) refers to the degree which the teachers accept and acknowledge the children's feelings; and Distract (DIS) refers to the degree which the teachers use distraction when the child experience negative emotions.

Participants were asked to answer how likely they would be to react in the way presented with answers given on a 7-point Likert-type scale ranging from highly unlikely (1) to very likely (7). A higher score indicated that teachers reacted more in the described way, thus higher score indicated better emotion socialization for EE, EFR, ACK and PFR, and worse emotion socialization for DR, GWR, PR and MR.

CCNES had good construct validity, and both CCNES and CTNES had good internal consistency and good test-retest reliability (Fabes, Poulin, Eisenberg, & Madden-Derdich, 2002; Spinrad et al., 2007). CCNES and CTNES have been applied using the two factors supportive reactions (EE, PFR and EFR) and non-supportive reactions (DR, MR and PR) (Fabes et al., 2002; Spinrad et al., 2007). Spinrad et al. (2007) excluded the two reactions GWR and DR as they did not factor with any of the other subscales, which gave a better internal consistency for the remaining subscales. In the current study, we calculated Cronbach's alpha and did a Confirmatory Factor Analysis (CFA) to assess if the measure should be applied as one, two (supportive and non-supportive reactions) or as nine factors. ACK was included as a supportive reaction and DIS a non-supportive reaction.

Emotional Support

The *Classroom Assessment Scoring System Pre-K* (CLASS Pre-K; Pianta, La Paro, & Hamre, 2008) and *Classroom Assessment Scoring System Toddler* (CLASS Toddler; La Paro, Hamre, & Pianta, 2012) were used to measure observed emotional support in the kindergarten units. CLASS Pre-K has a good internal validity (Li, Liu, & Hunter, 2020), and good predictive validity (Leyva et al., 2015). CLASS Toddler had excellent measurement properties and good criterion validity (Slot, Boom, Verhagen, & Leseman, 2017).

The domain Emotional Support as defined in CLASS assesses the degree to which teachers were able to support social and emotional functioning in the classroom or kindergarten. We used the four dimensions Positive Climate, Negative Climate, Teacher Sensitivity and Regard for Student Perspectives from the domain Emotional Support in CLASS Pre-K to measure emotional support in this study, as these dimensions were also

included in CLASS Toddler. The dimension Positive Climate measured the degree to which teacher and students demonstrated emotional connection, respect, and joy in their interactions by verbal and nonverbal communications. The dimension Negative Climate measured the degree to which teachers and/or students demonstrated expressed negativity like hostility, anger, or aggression in the classroom. The dimension Teacher Sensitivity measured the degree to which teachers demonstrated responsivity and awareness of students emotional and academic concern. The dimension Regard for Student Perspectives measured the degree to which teachers' interactions with students took the students point of view and interests into consideration.

CLASS Toddler was used for children aged 1-2 years and CLASS Pre-K was used for children aged 3-5 years. Each dimension was observed and scored systematically by personnel trained and licensed in CLASS. The CLASS observers performed observations in neighboring kindergartens, not their own, to reduce bias. Each observation consisted of three cycles of 20-minute observations and 10 minutes scoring afterwards, for each unit. The measure was scored ranging from low quality (1) to high quality (7). The dimension Negative Climate was reverse scored. Ratings from the three cycles were averaged. The scores of the dimensions were combined to get a mean score, in which a higher score indicated better emotional support in the unit. For inter-rater reliability five of the 53 observations were performed with a double coder in addition to the original coder.

Implementation Quality

Implementation quality was measured using the *Implementation Quality Questionnaire* (IQQ-39), which is a shorter version of the original *Implementation Quality Questionnaire* (IQQ-64; Bogen (2020) and Liu (to be published)), which evaluated the quality of universal preventive interventions in contexts like schools. See Appendix F for the IQQ-39.

Bogen (2020) and Liu (to be published) developed IQQ-64 based on research where they interviewed 10 participants with experience in implementing health promoting projects in schools with a semi-structured interview guide based on the CFIR, by Damschroder et al. (2009). A thematic analysis of the results identified four main themes in accordance with the CFIR, and the developers then added a fifth theme based on the CFIR; the outer setting (which was the one theme from CFIR not identified in the interviews). Questions to each theme were developed based on the qualitative interviews and CFIR. Thus, IQQ-64 measure five themes: Individuals, intervention, organization, process, and outer context. The theme individuals included the individuals feeling of ownership, self-efficacy, and attitude towards the intervention. The theme intervention included whether the intervention was easy to use in

practice, flexible and useful. The theme organization included cooperation, communication, and norms within the organization. The theme process included questions about the implementation process – if the school and its personnel was involved in developing the intervention, received adequate information about it, wanted and prioritized the intervention. The theme outer context measured if the community or country prioritized the project (Bogen, 2020). IQQ-64 had satisfying reliability levels (Bogen, 2020).

IQQ-39 was modified by Bogen (2020) and Liu (to be published) based on distributional information from their research, their subjective experience of the item's validity and to ensure each of the five main themes from the IQQ-64 were included, with multiple questions with both positive and negative formulations. The shortened measure was then adjusted to be used in kindergartens because it originally was developed to be used in schools. The adjustments consisted of changing the word school to kindergarten and inserting the name of the intervention in places where it was written "intervention". One question was excluded: "Political guidance have been given from the national or regional level to carry out the intervention", as it was not deemed relevant for the current study. Answers were given on a 7-point Likert-type scale ranging from strongly disagree (1) to strongly agree (7). Twenty items were reversed. A low score indicated low implementation quality, and a high score indicated high implementation quality (Bogen, 2020).

As IQQ-39 was a newly developed measure which had not been examined psychometrically we assessed internal reliability and performed an EFA to assess if the measure should be used as one factor, or as several factors.

Implementation Outcomes

To measure the subfactors acceptability, appropriateness, and feasibility previously established questionnaire developed by Weiner et al. (2017) were used: Acceptability of Intervention Measure, Intervention Appropriateness Measure and Feasibility of Intervention Measure, respectively. The factors were each measured with four questions which were answered on a 5-point Likert-type scale ranging from strongly disagree (1) to strongly agree (5). A higher score indicated better implementation outcomes. All three scales have been found to show promising psychometric properties (Weiner et al., 2017). In the current study, we translated the measures to Norwegian, and then back translated it to English by others in the project. The Weiner et al. (2017) measure has open access, thus approval for translation and back-translation were not needed. See Appendix G for the Norwegian translation.

Sustainability, penetration, and fidelity were measured using questions developed for this study. See Appendix G for an overview of the questions for each concept. The theoretical

background for the developed questions were Proctor et al. (2010). We developed a questionnaire to measure sustainability with the same form as the three subscales described above by Weiner et al. (2017), i.e., four questions which were answered on a 5-point Likert-type scale ranging from strongly disagree (1) to strongly agree (5). Penetration was measured with one question: "In what % of the possible situations did you use emotion coaching?". Fidelity consisted of nine questions to evaluate whether the teachers applied emotion coaching techniques during their workday (i.e., applied the intervention as intended). Each question was answered on a 7-point Likert-type scale ranging from strongly disagree (1) to strongly agree (7*)*. For all factors, higher scores indicated better implementation outcomes.

We performed an EFA to assess how to best apply these measures – as one combined measure or as subfactors. In addition, we performed a CFA to compare a one factor model, EFA suggested factor model and theoretically six factor model structure.

Covariates

We controlled for the three covariates: birthyear, number of supervision sessions attended and the impact of Coronavirus disease of 2019 (Covid-19), to explore if these factors influenced the effect of the predictors. Birthyear was measured by asking for the participants birthyear. Number of supervision sessions were measured with one question asking the participants how many times they had attended short meetings in their kindergarten, in which aspects of TIK-KT and emotion coaching was the subject. Covid-19 was measured with five questions developed for this study and included at T3 to assess Covid-19's emotional impact and effect on teacher's ability to emotion coach. These were measured on a 7-point Likert-type scale, ranging from strongly disagree (1) to strongly agree (7). Two questions were reversed. A higher score indicated that the Covid-19 situation led to negative emotional impact and less ability to emotion coach. See Appendix H for the Covid-19 questions.

Ethical Consideration

Participation in this study was not anticipated to lead to distress or other negative consequences for participants, including leaders, teachers, parents, and children. Considering the TIK program has proven benefits for children and parents in previous studies, it is likely that this intervention will be useful for the participants. The intervention was delivered to the control kindergartens after the initial intervention phase in order to ensure that they also got the opportunity to have similar (possible) benefits as those in the intervention condition.

In the questionnaires, adults were asked to give personal information, some of which may be perceived as sensitive (e.g., yearly income and psychological health) with the need to ensure confidentiality of the information collected. The platform TSD (Tjeneste for Sensitive

Data) owned by the University of Oslo, developed and operated by the TSD service group at the same university, IT-department (USIT), was used to store and protect all data.

All participants signed an informed consent form. Participants were informed they could choose to withdraw from the evaluation at any time. For leaders and teachers, withdrawing from the part of the study that consisted of the TIK-KT training, supervisions and use of the intervention was difficult because it was part of professional development by the FUS kindergartens. In Norway, professional development delivered by an employer to the employees is often mandatory and not something one could withdraw from.

The project was ethically approved by the Norwegian Centre for Research Data (NSD) (see Appendix I). Before we joined the study, NSD was applied to and approved as we did not have any data that would not comply with the data protection legislation. As the project was not considered to be health research, an approval from the Regional Committees for Medicine and Health Research Ethics was not necessary.

Data Treatment

General Cleanup

We cleaned the raw data files before starting the analyses. This was necessary as several of the response styles were open-ended for the participants to fill in (i.e., no response options), and thus several of the answers were not optimal for the analyses (answered e.g., 10 years instead of just 10). The ID's needed to be matched in the different questionnaires as the participants created their own ID, which led to typing errors and some not following the ID directions correctly.

Deleting Duplicates

Thirteen participants answered some of the questionnaires twice. Nine of these were from T2 the others were from T3. From the duplicated answers we chose to keep only the first responses that was completed. This were done as most of the other participants answered at the same time as the first responses were registered, to eliminated possible differences in responses due to response time to influence the results.

Missing Data

In the combined dataset, there were 17% missing data from the questionnaire at T1, 11% missing data from the observation measure (CLASS emotional support) at T1, 50% missing at T2 and 44% missing data at T3. The missing data were missing not at random since cases were missing data longitudinally for whole questionnaires as not all participants answered all three questionnaires. When the questionnaires were examined separately, they

did not contain missing data as "Nettskjema" was used to collect the data, and the administration did not allow participants to skip questions.

We used Expectation Maximization (EM) imputation to handle the missing data. The EM algorithm finds maximum likelihood estimates in models with incomplete data using a general-purpose iterative algorithm ("Multiple Imputation and the Expectation-Maximization Algorithm,") EM single imputation was chosen as multiple imputation and maximum likelihood estimation were not accepted by the software used to conduct the multilevel mediation analyses. EM imputation was performed with the kindergarten variable as a predicting variable, as this was the only variable with no missing data and thus the only variable that SPSS accepted. We imputed mean scores for the predictors, covariates, the mediator, and the outcome. Variables with one question and descriptive variables were imputed as they were.

Unit information. Due to missing data, we lacked self-reported information about unit belonging for 121 participants. Different questionnaires were examined to compare answers and identify unit belonging. Nine participants unit were identified through their email address. In two cases where information were conflicting, self-report was used over observations to decide unit, as we assumed self-report was more correct as workers were more likely to be accurate in determining which unit they belong to. For 76 participants who had not self-reported about unit belonging, this was identified through CLASS observations. We called the leaders to identify unit belonging for 20 participants. For 14 participants unit belonging were not identified due to them working in positions not related to one unit.

One kindergarten changed their unit structure during the study. Four of the participants from this kindergarten answered according to the new unit structure in one questionnaire and the old unit structure in the other. As we had information on unit belonging related to the new unit for more participants, we chose to follow the new and existing unit system which consisted of five units compared to two large units in the old system.

Outliers

We used winsorization with cut-of percentiles at 2.5% and 97.5% to handle extreme outliers. Values below 2.5% were altered to the 2.5% value, and values above 97.5% were altered to the 97.5% value. The cut-of percentiles were chosen as they dealt with outliers without altering the variable means.

Statistical Analyses

All statistical analyses were performed using IBM SPSS version 27. CFA was performed using the software IBM ® SPSS ® AMOS Graphics. Multilevel mediator analyses were completed using the SPSS macro MLmed Beta Version 2.0 (Rockwood, 2017).

Reliability and Validity

We performed EFA on the variables: perception of kindergarten functioning, implementation quality, and implementation outcome to examine whether the measures should be treated as one factor or several. Cronbach's alphas and inter-item correlation mean for the measures; perception of kindergarten functioning, CTNES-T, DERS-18, K6, CLASS, implementation quality, implementation outcomes and the covariate Covid-19 were examined to assess internal reliability. We performed a CFA to assess model fit on the implementation outcomes measures and CTNES-T as parts of the measures were established. For CLASS we estimated interrater reliability by calculating intraclass correlation, to assess convergence between raters. All factorial analyses and reliability analyses were performed using data which had not been imputed or winsorized.

Main Analyses

For the main analyses, the relationship between the predictors (education, experience, perception of kindergarten functioning, psychological distress, emotion dysregulation, emotion socialization and emotional supports) and implementation outcomes were examined, and then whether these relationships were mediated by implementation quality. We controlled for the three covariates birthyear, Covid-19 and the number of supervision sessions the teachers received in the main and sensitivity analyses.

A multilevel approach was chosen as we had nested data on individual, unit level and kindergarten level. Multilevel analysis has the advantage of examining outcomes across hierarchical levels and explaining their variance with variables at corresponding levels, which accounts for individual's behavior in a group context (Heck, Thomas, & Tabata, 2013). Main analyses were performed using imputed and winsorized data due to missing data and outliers in the responses, and all data were z-scored. We first ran a null model to investigate if enough of the variance could be explained by the nested structure to include all three level in further analyses. The intraclass correlation for the two levels units and kindergarten was first performed separately. The intraclass correlation for the unit level was 15% for IQQ-39 and 56% for implementation outcome. The intraclass correlation for the kindergarten level was 6% for IQQ-39 and 4% for implementation outcome. The levels intraclass correlation was then estimated in a three-level model. The intraclass correlation was 15% for unit and 0% for

unit*kindergarten for the IQQ-39 measure. The intraclass correlation was 55% for unit and 0% for unit*kindergarten for the implementation outcome. All the intraclass correlation analyses were performed with maximum likelihood estimation. Unit level was chosen as level 2 and kindergarten level was excluded as level 3 based on these results. Fourteen of the 344 participants did not have unit information and were thus not included in the analyses as unit was used as level 2 in the multilevel analyses.

The predictors education, experience, perception of kindergarten functioning, psychological distress, emotion dysregulation and emotion socialization were measured at level one, and emotional support at level two. The mediator and outcomes were measured at level one. The predictors were analyzed separately. The predictors were expected to show effects at the level they were measured. However, due to the strong nesting effect at the unit level we expected to see effects (possibly stronger) at this level even when the predictor was measured at level 1.

To answer hypothesis one (i.e., if there was a significant effect of the predictors on implementation outcomes), we performed general linear model analyses with unit as a factor for the individual level/ within (i.e., individual level effects controlled for unit grouping effect). For group level/ between effects (i.e., group/unit level effects) we performed linear mixed model analyses using group means acquired through aggregating the data. Restricted maximum likelihood was used.

To answer hypothesis two, multilevel mediator analyses were completed using MLmed, as this macro were able to perform the chosen multilevel mediation analysis. For the level 1 predictors a 1-1-1 (all measures at level 1) data design with within and between effects was applied and for the level 2 predictor a 2-1-1 (predictor at level 2, mediator and outcome at level 1) data design with between effects was used. MLmed group-mean centered the predictors and used the group-mean of the predictors as a level-2 predictor to separate between and within effects. The within results showed the level 1 (i.e., individuals) effect controlled for the clustering effect from the units and between results showed the level 2 (i.e., unit) effects. Intercepts were random and slopes were fixed.

Sensitivity Analyses

Sensitivity analyses were run in addition to the main analyses to examine whether the chosen method for handling missing data influenced the results. The sensitivity analyses were run using the same method as the main analyses. The only difference was that the sensitivity analyses used the original data, and the main analyses used the imputed and winsorized data.

We focused on examining effect sizes and directions as significance is influenced by sample size, and the sample size in these analyses were smaller.

Results

Measurement assessment

Mean and range were calculated for all measures after imputation and winsorization. See Table 2.

Internal Reliability

For the measure CTNES-T, we chose to examine the Cronbach's for the subscales, as Spinrad et al (2007) excluded DR and GWR in their study. Both of these did not have a good Cronbach's score in our sample (see Appendix J) and were therefore excluded from the rest of the analyses. The rest of these seven subscales were included in the examination of the supportive (EE, PFR, EFR, ACK) and non-supportive (MR, PR, DIS) factors (see Table 2).

Cronbach's alpha and inter item correlation mean was estimated for the other measures as well. All the measures had a good Cronbach's alpha and can be seen in Table 2.

Table 2

Mean, standard deviation, range, Cronbach's alpha, and Inter-item correlation mean for the predictors, covariates, implementation quality, and implementation outcome.

Measures	M (SD)	Range	Cronbach's alpha	Inter-Item Correlation Mean
Education	4.06 (1.39)	5.00		
Experience	10.41 (6.65)	28.38		
Kindergarten functioning	5.72 (0.56)	2.38	.806	.315
Psychological distress	26.84 (2.26)	9.38	.711	.320
Emotion dysregulation	28.50 (5.54)	24.00	.827	.238
Supportive ES	6.31 (0.45)	1.97	.928	.274
Non-supportive ES	2.11 (0.63)	2.53	.899	.190
CLASS Emotional support	5.52 (0.62)	2.67	.831	.552
Birthyear	1981.97 (8.47)	33.38		
Supervision sessions	3.57 (1.96)	10.00		
Covid-19	2.87 (0.80)	3.80	.640	.250
Implementation Quality	5.51 (0.57)	2.64	.945	.319
Implementation Outcomes	4.11 (0.02)	2.50	.927	.332

Note. Mean and range are after imputation and winsorization was performed. Cronbach's alpha and inter item correlation were estimated using original data. Experience=Years worked

in a kindergarten, ES = emotion socialization reactions, Supervision sessions= Number of supervision sessions in Tuning in to Kids for Kindergarten Teachers the teachers had, Covid-19= Covid 19's emotional and emotion coaching capacity impact on teachers. Education scale: 1 (primary and/or secondary education), 2 (vocational high school level education), 3 (high school), 4 (shorter education/coursing after high school), 5 (higher education of 1-3 years), 6 (higher education of 5-6 years), a higher psychological distress measure indicate less distress.

[a] N=344.

Inter-rater Reliability

As CLASS emotional support was measured through observations and five of these were performed with a double coder (i.e., two observers did the same observation), inter-rater reliability was examined. A one-way random model was used as each unit was rated by different raters (Shrout & Fleiss, 1979). The raters were not completely random as geography was a factor in determining raters, but a one-way random model was the closest fit for the characteristics of this study. The intraclass correlation **(**1,k) with an Absolute agreement definition and 95% confidence interval was 0,72 (0,20-0,91). Thus, emotional support had a moderate interrater reliability, as indicated by its value between 0,50 – 0,75 (Koo & Li, 2016).

Construct Validity

Perception of Kindergarten Functioning. The factorability of the correlation matrix of the nine items was supported by both a statistically significant value of Bartlett's Test of Sphericity and a Keyser-Meyer-Oklin value of .811 exceeding the recommended value of 0.6 (Kaiser, 1974). An EFA was performed to estimate the number of factors to retain. Four criteria were taken into consideration when deciding how many factors to choose: Eigenvalue, Scree-plot, parallel analysis and logic of the structure. First, only two of the factors had an Eigenvalue exceeding 1 (3.58, 1.14), accounting for 52% amount of the total variance. Second, the scree plot showed a clear break after the first factor. Third, only the first component had an eigenvalue exceeding the cut-off value suggested by the parallel analysis (i.e., the highest value that can be expected to occur randomly in a correlation matrix that is based on nine items and 287 respondents). Lastly, one factor makes more sense than two factors for this measure. The measure collects info about the quality of several aspects in the kindergarten but there was no grouping of aspects that would indicate that it should be divided into two factors. Thus, one factor was chosen for this measure.

CTNES-T. CFA was performed to assess goodness of fit for the CTNES-T measure. We estimated a model fit with a nine-factor model and a one-factor model which included all

nine reactions, and a two-factor and a seven-factor model in which GWR and DR was excluded. The two-factor model was run with the two factors; supportive reactions (ACK, EE, PFR, EFR) and non-supportive reactions (DIS, PR, MR). Observations were independent, and the total sum score showed normality. Maximum likelihood estimation was selected. See Table 3 for results. The seven-factor model and the two-factor model showed better model fit compared to the nine-factor model and the one-factor model. The two-factor model and the seven-factor model both showed an acceptable model fit for the CMIN/DF and RMSEA estimates, and a poor model fit for the CFI. AIC values was lowest for the seven-factor model, indicating this model had better predictive value compared to the other models.

Table 3

Results for Confirmatory Factor Analysis of Coping with Toddlers Negative Emotions Scale-Teacher's version (CTNES-T) with a One, Two, Seven and Nine Factor Structure.

Model	χ^2	df	CMIN/DF	CFI	AIC	RMSEA
CTNES-T 1Factor	12619***	4751	2.66	0.39	13217	0.08
CTNES-T 2Factor	6997***	3310	2.11	0.67	7515	0.06
CTNES-T 7Factor	6934***	3299	2.10	0.68	7474	0.06
CTNES-T 9Factor	9482***	4716	2.01	0.63	10147	0.06

Note. $\chi 2$ = Chi-square; df=degrees of freedom; CMIN/DF= Minimum discrepancy per degree of freedom; CFI, Comparative Fit Index; AIC, Akaike Information Criterion; RMSEA, Root-mean-Square Error of Approximation. CTNES-T 1Factor=All reactions combined to one factor, CTNES-T 2Factor= Seven reactions with acceptable internal consistency divided into a supportive (EE, PFR, EFR, ACK) or non-supportive factor (DIS, MR, PR), CTNES-T 7Factor=The reactions EE, PFR, EFR, ACK, DIS, MR, PR. CTNES-T 9Factor=Each reaction represented one latent factor.

*** $p < 0.001$.

We chose to apply the CTNES-T as the two factors non-supportive and supportive reactions for the main analyses as both factors had Cronbach's alpha scores that were satisfactory, the CFA showed satisfying results for this model, and previous research and logical interpretation suggested this was a better fit.

IQQ-39. An EFA was completed for the IQQ-39 measure. The factorability of the correlation matrix of the 39 items was supported by both a statistically significant value of Bartlett's Test of Sphericity and a Keyser-Meyer-Oklin value of .907 exceeding the recommended value of 0.6 (Kaiser, 1974). An EFA was performed to estimate the number of

factors to retain. Four criteria were taken into consideration when deciding how many factors to choose: Eigenvalue, Scree-plot, parallel analysis, and logic of the structure. First, 9 factors had an Eigenvalue exceeding 1 (13.56, 2.89, 2.46, 1.57, 1.29, 1.26, 1.17, 1.10, 1.00), accounting for 68% of the total variance. Second, the scree plot showed a break after the third factor. Third, only the 3 first had an eigenvalue exceeding the cut-of value suggested by the parallel-analysis (i.e., the highest value that can be expected to occur randomly in a correlation matrix that is based on 39 items and 171 respondents). The three factors accounted for 49% of the total variance. An EFA with oblique rotation was then performed on the 39 items with three factors chosen. After examining the three factors model, we concluded that there was no clear logical theme within the three factors. An EFA with oblique rotation was then performed with two factors. The lack of a logical theme within the factors persisted with this factor structure. Both the three and two factor models showed a grouping of the negatively worded questions within one factor even though these not necessarily had the same thematic or direction (i.e., not all were reversed). Since both models lacked a logical structure the one factor model was considered. The Cronbach's alpha for the one factor model was excellent (see table 2). Thus, the one factor model was chosen for the analyses.

Implementation outcomes. In the EFA, the factorability of the correlation matrix of the 26 items was supported by both a statistically significant value of Bartlett's Test of Sphericity and a Keyser-Meyer-Oklin value of .876 exceeding the recommended value of 0.6 (Kaiser, 1974). The four criteria: eigenvalue, Scree-plot, parallel analysis and logic of the structure were taken into consideration when deciding how many factors to choose. First, only 3 of the factors had an Eigenvalue exceeding 1 (11.00, 8.27, 1.01), accounting for 78% amount of the total variance. Second, the scree plot showed a clear break after the second factor. Third, only two components had an eigenvalue exceeding the cut-of value suggested by the parallel-analysis (i.e., the highest value that can be expected to occur randomly in a correlation matrix that is based on 26 items and 287 respondents). An EFA with oblique rotation was then performed on the 26 items with 2 factors chosen. The fidelity questions were the only questions in factor 2 and the rest of the implementation outcome questions was in factor 1. This was likely to be caused by the fidelity questions being collected at a different time and not by an underlying logical structure. Thus, a one factor model was considered.

An CFA was completed in addition to the EFA. A CFA goodness of fit estimation was performed with a one-factor, two-factor (based on results from the EFA), and six-factor model (appropriateness, acceptability, feasibility, fidelity, penetration and sustainability). To compute fit measures with incomplete data the "fit the saturated and independence models"

option was chosen, and Maximum Likelihood estimation. The six-factor model was unidentified. Looking at the output estimates of regression weights and variance, the unidentified factor was the latent factor penetration which included one question and one error variance. Deleting the unidentified factor penetration for the six-factor model gave a five-factor model which was identified (see results in Table 4), but only the estimate CMIN/DF indicated an acceptable model fit. AIC was lower for the five-factor model, which indicated this model had better predictive value than the other two. Thus, compared to results from the one-factor model and the two-factor model, the five-factor model was a better fit.

We chose to apply the implementation outcomes measure as one mean score as the total outcome measure had a satisfactory Cronbach's alpha and Inter-Item Correlation Mean, and the use of the measure as a mean score was logical. The CFA favored the five-factor model over the one-factor model, but as none of the models had good fit, we chose to emphasize the EFA and a logical structure.

Table 4

Results of CFA for Implementation Outcomes with a One-Factor, Two-Factor and Five-Factor Model.

Model	χ^2	df	CMIN/DF	CFI	AIC	RMSEA
Outcomes 1Factor	3997***	302	13.2	0.33	4147	0.21
Outcomes 2Factor	1648***	301	5.48	0.75	1800	0.12
Outcomes 5Factor	879***	267	3.30	0.89	1045	0.09

Note. χ^2 = Chi-square; df=degrees of freedom; CMIN/DF= Minimum discrepancy per degree of freedom; CFI, Comparative Fit Index; AIC, Akaike Information Criterion; RMSEA, Root-mean-Square Error of Approximation. Outcomes 1Factor=All outcome variables as one factor, Outcomes 2Factor=fidelity as one factor, sustainability, penetration, acceptability, appropriateness and feasibility as another factor, Outcomes 5Factor=included the factors sustainability, acceptability, appropriateness, fidelity and feasibility.
***p<0.001.

Main Analyses

General linear models with units as factors (i.e., within effects) and linear mixed models with unit mean scores (i.e., between effects) were completed to answer Hypothesis 1 (i.e., the predictors affect the implementation outcome). Multilevel mediation analyses in MLmed were completed to answer Hypothesis 2 (i.e., implementation quality mediate the effect between predictors and implementation outcome).

Within Effects

Hypothesis 1. See Table 5 for an overview of individual level effects (i.e., within effects), controlled for unit grouping effect, on implementation outcome. Two of the predictor models found a significant effect of the predictors on the implementation outcomes. Teachers' perception of better kindergarten functioning prior to the intervention led to a better implementation outcome. Teachers who had more supportive emotion socialization prior to the intervention had better implementation outcome. The rest of the predictors (education, experience, psychological distress, emotion dysregulation and non-supportive emotion socialization) did not have a significant effect on the implementation outcome on an individual level when controlling for unit grouping effect. Covid-19 was the only covariate that had a significant effect in any of the predictor models. More negative influence of Covid-19 led to worse implementation outcome (see Appendix K for covariate estimates).

Hypothesis 2. See Table 6 for an overview of individual level effects (i.e., within effect results) for the mediating effect of implementation quality between predictors and implementation outcomes. H2 was supported for two predictors: implementation quality significantly mediated both the relationship between perception of kindergarten functioning and implementation outcome and the relationship between supportive emotion socialization and implementation outcome. The rest of the predictor models (education, experience, psychological distress, emotion dysregulation and non-supportive emotion socialization) did not show a significant implementation quality mediation effect on the implementation outcome on an individual level when controlling for unit grouping effect.

For teachers who reported higher implementation quality, the implementation outcomes were also higher (effect of 0.07-0.08) for all predictor models. Covid-19 was the only covariate that had a significant effect in any of the predictor models. More negative influence of Covid-19 led to worse implementation outcome (see Appendix L for covariate estimates).

Between Effects

Hypothesis 1. See Table 5 for an overview of group/unit level effects (i.e., between effect results) on implementation outcomes. Both units with more teacher experience and perception of a better kindergarten functioning significantly predicted a better implementation outcome. The rest of the predictors (education, psychological distress, emotion dysregulation, supportive emotion socialization, non-supportive emotion socialization and emotional support) did not have a significant effect on the implementation outcome on a unit level. In units where teachers had a higher mean number of supervisions, implementation outcomes

were significantly better in all predictor models. Units with a lower mean birthyear (i.e., older age) of the teachers scored higher in implementation outcome in all models expect the experience. There was no effect of Covid-19 on unit level (see Appendix K for covariate estimates).

Hypothesis 2. See Table 6 for an overview of group/unit level effects (i.e., between effect results) for the mediating effect of implementation quality between predictors and implementation outcomes. Implementation quality mediated the relationship between units with perception of better kindergarten functioning and better implementation outcome, units with more supportive emotion socialization and units with better implementation outcome, and more emotional support and better implementation outcome on a unit level.

On a unit level, implementation quality significantly predicted implementation outcomes (effects of 0.17-0.18) in all predictor models. In units where teachers had a higher mean number of supervisions, the implementation quality and implementation outcomes were significantly better in all predictor models. Units with a lower mean birthyear (i.e., older age) of the teachers scored higher in implementation outcome in all models expect the experience (see Appendix L for covariate estimates). Units with teachers that were more affected by Covid-19, scored lower in implementation quality for two predictor models (education and experience).

Table 5

Estimates and standard error (SE) between predictors and implementation outcomes from general linear model (within) and linear mixed model (between) analyses with covid, birthyear and number of supervision sessions as covariates.

Predictor	X to Y (SE)	
	Within	Between
Education	0.02 (0.01)	0.06 (0.05)
Experience	-0.00 (0.02)	0.16** (0.05)
Kindergarten functioning	0.03* (0.01)	0.09* (0.04)
Psychological distress	-0.01 (0.01)	0.00 (0.05)
Emotion dysregulation	0.02 (0.01)	-0.05 (0.04)
Supportive ES	0.04** (0.02)	0.06 (0.04)
Non-supportive ES	-0.02 (0.02)	-0.06 (0.04)
CLASS Emotional support		0.04 (0.03)

Note. Predictors were self-report measures at teacher level except CLASS Emotional Support which was observed at unit level. Data were imputed, winsorized and Z-scored. X= predictor, Y = implementation outcome, ES= Emotion socialization, within= individual level controlled for unit grouping effect, between= unit level effect, Education scale: 1 (primary and/or secondary), 2 (vocational school), 3 (high school), 4 (courses after high school), 5 (higher education 1-3 years), 6 (higher education 5-6 years) a higher psychological distress measure indicate less distress.

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

[a] Within N=330, Between N=108.

Table 6

Estimates and standard error (SE) from multilevel mediation analyses with implementation outcome as dependent variable and implementation quality as mediator, with covid, birthyear and number of supervision sessions as covariates.

Predictors	X to Y (SE)	X to M (SE)	M to Y (SE)	Indirect (SE)
		Within		
Education	0.02(0.01)	0.01(0.06)	0.07***(0.02)	0.00(0.00)
Experience	0.01(0.02)	-0.17(0.09)	0.07***(0.01)	-0.01(0.01)
Kindergarten functioning	0.02(0.01)	0.20**(0.07)	0.07***(0.02)	0.01*(0.01)
Psychological distress	-0.01(0.01)	0.76(0.07)	0.07***(0.02)	0.01(0.01)
Emotion dysregulation	0.03(0.02)	-0.11(0.07)	0.07***(0.02)	-0.01(0.01)
Supportive ES	0.03(0.02)	0.20**(0.07)	0.07***(0.02)	0.01*(0.01)
Non-supportive ES	-0.01(0.02)	-0.13(0.07)	0.07***(0.02)	-0.01(0.01)
		Between		
Education	0.02(0.03)	0.18(0.11)	0.18***(0.03)	0.03(0.02)
Experience	0.11**(0.04)	0.15(0.14)	0.17***(0.03)	0.03(0.32)
Kindergarten functioning	0.04(0.03)	0.28**(0.10)	0.18***(0.03)	0.05**(0.02)
Psychological distress	-0.01(0.04)	0.06(0.12)	0.19***(0.03)	0.01(0.02)
Emotion dysregulation	-0.01(0.03)	-0.16(0.10)	0.18***(0.03)	-0.03(0.02)
Supportive ES	-0.00(0.03)	0.21*(0.10)	0.19***(0.03)	0.04*(0.02)
Non-supportive ES	-0.03(0.03)	-0.14(0.10)	0.18***(0.03)	-0.03(0.02)
CLASS Emotional Support	0.00(0.02)	0.14*(0.06)	0.18***(0.03)	0.03*(0.01)

Note. Predictors were self-report measures at teacher level except CLASS Emotional Support which was observed at unit level. Data were imputed, winsorized and Z-scored. X= predictor,

M= Implementation Quality, Y = implementation outcome, ES= Emotion socialization, within= individual level controlled for unit grouping effect, between= unit level effect, Education scale: 1 (primary and/or secondary), 2 (vocational school), 3 (high school), 4 (courses after high school), 5 (higher education 1-3 years), 6 (higher education 5-6 years). A higher psychological distress measure indicate less distress.

[a] N=330.

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

Sensitivity Analyses

See Appendix M for hypothesis 1 sensitivity analyses. Se appendix N for hypothesis 2 sensitivity analyses.

Within Effects

Hypothesis 1. Individual level effects on implementation outcome from sensitivity analyses were mostly of similar size and direction as the main analyses. The exception was the predictor experience and the covariate birthyear in all predictor model, which were descriptively larger in sensitivity analysis compared to main but had the same direction.

Hypothesis 2. The mediation model individual level effects were mostly of similar direction in main and sensitivity analyses. However, most effect sizes were descriptively larger in sensitivity analyses compared to main analyses.

Between Effects

Hypothesis 1. Unit level effects were mostly of the same direction in the main and sensitivity analyses. Several of the effects were descriptively larger, especially educations effect on implementation outcomes, which were descriptively larger and significant in the sensitivity analyses.

Hypothesis 2. Direction of effect sizes were similar for mediation models on a unit level in both main and sensitivity analyses. Most effect sizes were descriptively larger in the sensitivity analyses compared to the main. Implementation quality had a significant, mediating effect of the relationship between emotion dysregulation and implementation outcomes, and descriptively larger effect sizes in the sensitivity analyses.

Correlation Matrix

A complete correlation matrix was completed including all the measures. See Appendix O.

Discussion

The aim of the present study was to explore the effect of the teachers: education, experience, perception of kindergarten functioning, psychological distress, difficulty with

emotion dysregulation, emotion socialization (measured with supportive and non-supportive emotion socialization) and emotional support on implementation outcomes. We hypothesized that the relationship between the different predictors and implementation outcomes would be mediated by implementation quality.

Hypothesis 1 (i.e., that implementation outcomes would be related to the predictors) was supported for three predictors: Individuals and units that rated higher in perception of kindergarten functioning, individuals with more supportive emotion socialization and units consisting of teachers with more experience working in a kindergarten, scored higher in implementation outcomes. We did not find an effect of the predictors: education, psychological distress, emotion dysregulation, non-supportive emotion socialization and CLASS emotional support on implementation outcomes.

Hypothesis 2 (i.e., implementation quality would mediate the relationship between the predictors and implementation outcomes) was supported for the three predictors. For the predictor perception of kindergarten functioning, the relationship found between the predictor and implementation outcome in hypothesis one was mediated by implementation quality on both an individual and a unit level. Better supportive emotion socialization reactions (individual and unit level) and better observed CLASS emotional support (unit level) led to better implementation outcomes through an indirect effect (mediated by implementation quality), although the relationship between the two predictors and implementation outcomes were not significant in H1. There was no significant mediating effect of implementation quality on the relationship between five of the predictors (education, experience, psychological distress, emotion dysregulation, non-supportive emotion socialization) and implementation outcomes.

In addition, we observed some other interesting findings. Individuals and units with better implementation quality had significantly better implementation outcome, for all predictor models. Units where individuals had more supervision sessions in TIK-KT (covariate) scored better in implementation quality and implementation outcomes. Units with older participants (covariate) scored better in implementation outcomes, for all predictor models except experience. Individuals with more negative impact because of Covid-19, scored lower on implementation quality. All significant effects were descriptively stronger at the unit level compared to the individual level.

Education Level/ SES

The lack of significant results for the effect of educations are in accordance with previous research by Kallestad and Olweus (2003) who did not find an effect of teachers'

education on teachers' implementation of a anti bullying intervention program in schools and show that their results may be generalizable for implementation of programs by teachers in kindergartens. However, SES has been shown to be linked to better emotional intelligence in trainee teachers (Kant & Lenka, 2013), and since education is an indicator of SES we expected that better emotional intelligence in teachers would be associated with an greater ease of implementing TIK. Since all of the employees work in the same organization and several of the employees are employed in similar positions their pay and education could be relatively equal. This may lead to a lack of variability in education and thus make it harder to see whether education had an impact on implementation outcome. However, the sensitivity analysis did show a significant unit level effect of education on implementation outcome. Thus, drawing a clear conclusion regarding educations effect is difficult and further research is need to either confirm the main analysis findings or the sensitivity analysis findings.

Kindergarten Experience

The individual level results did not find an effect of experience on implementation outcomes. This was consistent with findings from a school study by Kallestad and Olweus (2003), which implemented a bullying prevention program, and indicated the generalizability to a kindergarten context.

However, the kindergarten unit level results were not in accordance with Kallestad and Olweus (2003), as kindergarten units with teachers with more experience led to better implementation outcomes in the unit. This was the opposite direction of the expected effect when considering that experience have previously shown to be negatively correlated to the degree of implementation (Rohrbach et al., 1993). However, Yassien (2018) found that teachers with more experience had significantly higher emotional intelligence than those with less experience. Thus, our findings corresponded with the expected effect when considering the research conducted by Yassien (2018). The employees with more experience may be more committed to the job compared to those with less experience. This might have led to them being more committed to the implementation as well. Thus, resulting in more experience leading to a better implementation outcome. The mediation analysis did not show an indirect effect, but the direct effect was still persistent. Indicating that effect of experience directly affected the implementation outcome.

Perception of Kindergarten Functioning

The teachers' perception of kindergarten functioning showed the expected effect on implementation outcome in both the individual and unit level results. The results showed that better (i.e., more positive) perception of kindergarten functioning led to better implementation

outcome which was expected as Malloy et al. (2014) found organizational climate in schools to impact teachers' implementation of a social-emotional and character development program and organizational climate has been shown to significantly predict organizational commitment (Berberoglu, 2018). The commitment of the employees to the organization may have influenced the commitment and liking of the TIK program.

Psychological Distress

The lack of significant effects of psychological distress on implementation outcomes was not expected as previous research have found greater psychological distress to be associated with lower emotional intelligence (Zhang et al., 2016) which we expected to influence the participants use, acceptance and liking of TIK. We also expected an effect of this measure as psychological distressed participants might not have enough extra energy and could experience the implementation as a burden. Our results indicate that an individual's perception of their psychological state does not have an impact on how well they are able to implement a new program in their kindergarten.

There is a possibility that individuals which were affected by their own psychological distress to a degree that would affect their work, were on a sick leave. Nationwide, the second biggest reason for doctor documented absence from work was mental disorders (NAV, 2021). Another explanation may be that individuals that struggle more at work, e.g., with psychological distress, had less energy to answer the questionnaires in this study as this may constitute additional work. This may have led to less variance in our sample due to lack of data from individuals with greater psychological distress.

Emotion Dysregulation

There was no significant effect of emotion dysregulation on implementation outcomes at the individual or unit level in the main analysis. This was not expected as we assumed that emotion dysregulation would lead the teachers' emotions to get in the way of their ability to act according to the goals of using TIK-KT. As with the psychological distress measure, this might be caused by the lack of responses or involvement from those struggling more with emotional dysregulation. Those who struggle to regulate their emotions might either be absent from their work duo to poor mental health or lack time and energy to respond to the questionnaires. However, in the sensitivity analysis there was a significant negative indirect effect for unit level, indicating that less emotion regulation difficulties led to a better implementation outcome through implementation quality which was in accordance with our hypothesis. However, due to the conflicting results from the main and analysis and sensitivity analysis can no clear conclusion be drawn, and more research is needed on this predictor.

Emotion Socialization

The supportive emotion socialization results, and their directions was expected as better emotion socialization should make the program easier and thus influence the acceptability and satisfaction with the program (i.e., the implementation outcomes). This was also expected as those who had less supportive emotion socialization styles could experience cognitive dissonance regarding the TIK-KT techniques and their own behavior and thus reject the program which could further influence the implementation quality and implementation outcome. However, this hypothesis did not translate to the non-supportive emotion socialization as it did not have a significant effect on the implementation outcome.

These results may not be that generalizable to other intervention's implementations due to its strong association to the TIK program, as an important part of the intervention is emotion socialization and implementing a program that teaches emotion socialization may be easier for teachers that already were better at it. Future research should investigate if emotion socialization is a predictor for other interventions as well or if having some of the skillset required by the intervention beforehand increases the implementation outcome.

Emotional Support

We did not find the expected effect of emotional support as we expected to find an effect of unit level observed emotional support on implementation outcomes as we thought units with greater emotional support would have an easier time implementing TIK-KT and less cognitive dissonance as the program required the teachers to show support and acceptance of children's emotions. Our lack of results was in accordance with Wanless et al. (2015) as they found emotional support not to predict fidelity. However, we found an indirect effect of emotional support on implementation outcome through implementation quality. This was also in accordance with Wanless et al. (2015), which also found emotional support to influence fidelity (part of implementation outcome) through a mediator (engagement in training). This suggest that emotional support only affect implementation outcome through a mediating mechanism.

The Mediating Effect of Quality

There was a mediating effect of implementation quality between; better perception of kindergarten functioning (individual and unit level), better supportive emotion socialization (individual and unit level) and better emotional support (unit level), and implementation outcomes. This was expected as implementation quality theoretically should both precede and predict implementation outcomes (Lewis et al., 2015). Results regarding emotional support were consistent with previous research that applied emotional support as predictor,

engagement as mediating factor and fidelity as outcomes factor (Wanless et al., 2015).
Wanless et al. (2015) did not find a mediating effect of teachers' engagement in training
(quality factor) for the relationship between experience with the practices used and
implementation outcomes. We did find a mediating effect of quality on the relationship
between emotion socialization (i.e., previous skill used in the intervention) and
implementation outcomes. Results regarding perception of better kindergarten functioning
was in accordance with previous studies that found the relationship between organizational
support and factors of implementation outcomes to be mediated by factors like beliefs and
attitudes towards the intervention (Beets et al., 2008; Little et al., 2015). These previous
research studies applied different and more narrow constructs as mediators compared to ours
(implementation quality), and they used one component of implementation outcomes as
dependent variable as opposed to six components which we applied. This may explain the
observed differences. A review by Lewis (2020) concluded that existing research on
mediating mechanisms have applied strategies of implementation and implementation quality
(i.e., determinants) as the X variable. Our results indicated that the mediating mechanism still
is present when applying predictors as the X variable and may be generalized to kindergarten
contexts and lend further support to the conclusion that there exists a mediating mechanism in
implementation (Lewis et al., 2020; Williams, 2016).

According to Rucker, Preacher, Tormala, and Petty (2011), it is not necessary to have
a significant effect of X$\rightarrow$Y to have a mediating effect and focusing too much on this aspect
may be too restrictive. They rather recommended focusing on indirect effects and effect sizes.
We chose to follow this logic in our thesis and thus consider our supportive emotion
socialization and emotional support results to be considered as mediating effects.

Implementation Quality and Outcomes

Implementation quality predicted implementation outcomes on both individual and
unit level for all predictors (i.e., supporting part of hypothesis two), which was expected as
this is in accordance with previous research conducted (Exner-Cortens et al., 2020; Malloy et
al., 2014; Molloy et al., 2013). They were conflicting with results from (Wanless et al., 2015),
which found that self-efficacy (one of the factors included in implementation quality) did not
predict fidelity. It may be that differences in measurements or context could explain the
conflicting study. Previous studies have investigated one or a few subfactors of
implementation quality, and one or a few factors of implementation outcomes, whereas our
study applied a measure of all domains included in the concept implementation quality, and
six of the implementation outcomes. Our results thus showed that implementation quality

predicts implementation outcomes when both is applied as broader terms that captures more aspects included in the concepts. Our results and previous research on implementation quality factors support implementation theory which argue that implementation determinants (which implementation quality includes) predict implementation outcomes (Lyon & Bruns, 2019).

According to the theory of planned behavior by Ajzen (1991) behavioral intention predict behavior and is influenced by attitude, subjective norms, and perceived behavioral control. The implementation quality measure includes measures of attitudes towards the intervention (i.e., measure of attitude), perception of resistance toward the intervention and the interventions compliance with the existing norms (i.e., subjective norm towards the intervention) and self-efficacy regarding the use of the intervention (i.e., perceived behavioral control). Thus, individuals that scored higher in implementation quality scored higher in the three constructs that influence behavior intention to apply the intervention and subsequently use of the intervention (behavior). Implementation outcome both include (fidelity) and is closely linked to behavior related to the intervention. Thus, the theory of planned behavior could explain part of the association found between implementation quality and implementation outcomes.

There was some overlap between the quality measure and the outcome measure regarding two themes - attitude towards the intervention and the feasibility of the intervention. Several questions about what the individual thinks or feels about the intervention were similar in the two questionnaires, e.g., "I have sufficient expertise to successfully implement this intervention" from the quality questionnaire and the question "I have enough competence to be able to follow through with this intervention in a good way" from the fidelity subfactor of the implementation outcomes. We expected participants who answered both measures, would have answered similarly in questions with the same theme. Thus, some of the effect from quality on implementation outcomes may be due to this overlapping. However, this may not have been the case if the participant's attitude towards the intervention changed with increased experience using the intervention, as the outcomes measure was answered three months later than the quality measure. There were more themes in the measures that were not overlapping compared to the two themes that were overlapping: In the implementation quality measure there were more questions regarding the intervention itself, the facilitation of the implementation and organizational factors, whilst in the implementation outcomes measure the questions were more themed towards the implementation itself.

Individual and Group Level Effects

We found significant effects on both individual and unit levels for the perception of kindergarten functioning model, and supportive emotion socialization model. These results were expected as Bronfenbrenner's ecological systems theory explains that both individual and group levels affect the individual, and group levels interact and influence with each other's (Bronfenbrenner, 1979). CFIR explains that implementation may be influenced by factors on multiple levels (Damschroder et al., 2009). Our results support this theory and framework. However, some of the predictors significant effect were only present on one of the levels, indicating these predictors effect may be closely linked to the level they showed effect on.

The group level effect sizes were descriptively greater than the individual level effect sizes in several of the predictor model. Teachers may be affected by group conformity. Conformity is an individual's adjustment of their behavior, beliefs and attitudes to the standard of the group (Holt et al., 2015). This effect has been shown to both affect their attitudes publicly and privately, even weeks later (Levitan & Verhulst, 2016). This might explain why several of our analyses showed stronger group level effects. The group's attitudes and thus their answers might have become more similar through the effect of conformity.

One interesting and unexpected finding regarding group level effects were the lack of group level variation between the kindergartens. The abovementioned theories may suggest individuals in the same kindergartens would be more likely respond similarly, especially considering TIK-KT was a group-level intervention were the leader had a lot of responsibilities. Teachers in the same kindergartens probably had more similar number of supervision sessions, organizational climate and support from their leader compared to teachers in other kindergartens. However, according to Social Impact Theory, individuals are more influenced by groups they have closer proximity to (Latan, 1981). As teachers are in much closer proximity to their own unit compared to other units in their kindergarten, they may have been more influenced by factors related to their unit. This may indicate that focusing on implementation in smaller groups with higher proximity within the organization and not just the whole organization might be beneficial.

Limitations

Responses and Missing Data

The length of the questionnaires might have influenced the response rate as several questionnaires were quite long. This may have led to a skewness in the individuals that

participated in the study. Participants who were more passionate about the intervention, or had higher conscientiousness, may have been more likely to answer the questionnaires. The first questionnaire had a larger number of responses (T1: N=287) compared to the rest of the questionnaires (T2: N= 171, T3: N= 193). This may be due to the fact that the participants were discouraged by the length and thus did not want to answer the other questionnaires. Another explanation is that the first questionnaire was pre Covid-19 pandemic and teachers experienced considerable stress during the period of the implementation and evaluation of this intervention. This may have negatively impacted teachers' response rates.

A drop out analysis was performed to examine differences in the individuals who answered both the first and the last questionnaire or just the first. There was a significant difference between the two groups in four of the measures. The respondent that stayed were older, had more experience, scored lower on the non-supportive emotion socialization measure and were less psychologically distressed. It makes sense that both experience and age was significant as these most likely are correlated. Older and more experienced employees might be more involved in the organization which might explain why they completed both questionnaires. It makes sense that people who scored lower on non-supportive emotion socialization were retained as those who scored higher on non-supportive emotion socialization might have struggled or disliked the intervention more. A better implementation quality might be correlated with increased liking and involvement of the implementation which could influence the response rates. Less experienced psychological distress was expected in the group that answered both questionnaires, as they might struggle significantly. Retaining the individuals who were younger, less experienced, had more non supportive emotion socialization and experienced less psychologically distress might have given other nuances in our results. We expect these individuals may have answered lower on implementation quality and implementation outcome as they seem to be less invested as they did not answer the last questionnaire. However, this is only speculations and need further investigation.

Missing data are common in longitudinal studies and may lead to a loss of power and biased estimates, which may be countered by imputing data. In single imputation, the software treats the imputed data as a true value, and only produces one data set (as opposed to in multiple imputation). Variances of single imputed data tend to be underestimated (Engels & Diehr, 2003), thus we may have had less variance in our sample, which may have influenced the results. The sensitivity analyses with raw data indicated stronger effect estimates for several predictors and some additional significance effects compared to the main

analyses which used imputed data. However, single imputation was necessary as the software used (MLmed) did not accept other imputation methods. Another imputation method, less missing data, or a higher sample size may have resulted in larger effects, other significances, or more consistent results between the main and the sensitivity analyses.

There was a possible source of uncertainty regarding unit belonging due to a few participants that belonged to more than one unit. This may be common for e.g., special pedagogues and temporary staffs. As this did not apply to most participants, we do not expect the findings to have been significantly affected by this, however, is a possibility.

We included both teachers and leaders in our analyses. Leaders were assigned their own unit (i.e., one leader from one kindergarten was in a group with only one individual), so they were not mixed with the teachers, which may have had an impact on our results statistically (i.e., groups with one person). This were done as leaders led the whole kindergarten, thus assigning them to one specific unit within the kindergarten would have biased the results from that unit. Another possible issue with including leaders were regarding roles in the kindergarten as leaders had a different role compared to teachers; they were responsible for the implementation and probably had less opportunities to emotion coach the children. This may have influenced their answers, and thus our results. However, excluding the leaders may have led to a loss of information, a smaller sample size and less power, thus we included them in the analyses.

Data processing. We winsorized to handle extreme outliers. This may have led to us missing some essential information from the outlier answers. A drawback of winsorizing is the lack of clear guidelines for cut-off percentiles. However, as we chose a lower and upper limit that were modest (2.5 and 97.5 percentile), we expect this not to have influenced our results greatly.

Measurement Weaknesses

Self-reports. All data except from the emotional support observations were self-reports. This may have led to some of the answers being biased or skewed as they were from the participants perspective. Questionnaires like psychological distress, emotion dysregulation and emotion socialization were more likely to be influenced by biases as teachers might be affected by the social desirability bias. This bias involved the participants want to present themselves in a more socially desirable way (Grimm, 2010). Thus, the responses could be more moderate than what they actually were (e.g., under report their psychological distress). Recent events in the timeframe before the questionnaires might also have influenced the answers from the participants. These could either be related to the intervention (e.g., recent

bad experience with emotion coaching) or an unrelated event that affected the participants mood and cognitive state.

Observation. Emotional support was measured through observations. Teachers in observed units may have been affected by reactivity i.e., acted differently than they normally would (Baum, Forehand, & Zegiob, 1979). They might have tried to present themselves more favorably, such as acting kinder and more empathetically to children. Kindergarten teachers usually worked in proximity with colleagues and thus are used to others being present when they work, however, the additional knowledge that they are being evaluated might have influenced their behavior. Thus, this measure might show better emotional support than what the teachers usually show.

Newly developed measures. The results may be less reliable as some of the measures were newly developed or newly modified. Specifically, the outcomes, the quality measure, the perception of the kindergarten functioning and CTNES-T. As two of these four were some of the most important measures applied (i.e., mediator and outcome), they had a big impact on the study.

The IQQ-39 had not been psychometrically tested before. As the measure was based on a larger questionnaire that had been used in previous research, we expected good estimates from a CFA. This was not the case as none of the models received good model fit, but the five-factor model was much closer to receiving a good model fit compared to the others. This may indicate that the way we applied IQQ-39 was not ideal and could have influenced our results (results may have been more valid and shown more nuances had we applied five factors, and the use of one factor may have skewed our results). However, the Cronbach's alpha and the inter item correlation for the one-factor model received good results which indicate good internal reliability, thus we chose to apply the measure as one factor. IQQ-39 need to be revised further to achieve a good model fit and should be tested in different situations, and with different samples to further investigate its psychometric properties.

There was a difference in the degree to which the questionnaires used to measure the outcomes constructs had been psychometrically tested prior to the study - feasibility, acceptability, and appropriateness were measured by established and psychometrically tested questionnaires (Weiner et al., 2017), while fidelity, penetration, and sustainability were measured by new questionnaires created for this study. As we have tested the psychometric properties of the applied measures and found them acceptable, this indicated that this did not influence our results. However, established measures have usually been used in different situations with different samples, and may thus be viewed as more trustworthy and validated.

We applied one combined factor to measure implementation outcomes but as the measure consisted of several subfactors this choice may have affected our results and given less detailed information about which and to what degree different subfactors of implementation outcomes were affected by the predictors, or if there was a difference between subfactors. However, the mean scores for the combined factor had good reliability and could be justified theoretically.

The perception of kindergarten functioning measure were made for this study. It was not based on any theoretical framework, theory, or research. Applying a more psychometrically sound measure with a string theoretical framework could have given different results which may have been more valid and generalizable. The measure of perception of kindergarten functioning and implementation quality may overlap somewhat. E.g., in quality there were questions about individual's perception of leadership, and in perception of kindergarten functioning there was a question about the individual's relationship with their leader, which may receive similar responses. Organizational support (similar to perception of kindergarten functioning) has been applied in previous studies as both predictors and a quality factor. We applied perception of kindergarten functioning as a predictor because it was measured prior to the intervention.

Emotion socialization was measured with the scale Coping with Toddlers Negative Emotions – Teachers' version, and we used two factors - supportive- and non-supportive reactions. This was consistent with previous research by Fabes et al. (2002) and Spinrad et al. (2007) and made sense theoretically as one factor describes reactions that support emotion coaching, and the other factor describes reactions that do not support emotion coaching. The Cronbach's alpha and inter item correlation mean supported the use of two factors, the model fit for the two-factor and seven-factor model was similar thus we chose the most logical model. However, had we applied all factors as subscales we may have gotten more information about possible differences in the subscales impacts on implementation quality and implementation outcomes. Had we applied one mean score for the entire CTNES-T, the results may also have been different, giving less detailed information and/or stronger effects than when applying two factors, as we did.

Covid-19

Kindergartens had to close for several weeks after March 12 due to the Covid-19 pandemic in Norway and reopening resulted in changes in the number of children and teachers in a unit. This occurred in the middle of the implementation of the intervention. This event may well have influenced employee's emotion coaching skills as they might not have

continued to use the techniques if they were not around children during the lockdown period. This also led some of the supervision session to be held online instead of in person. This limited some of the activities that the supervision sessions usually contain like their ability to practice emotion coaching using roleplay. The kindergarten employees in Norway were instructed not to socially distance to the children when they got back to work. However, being constantly reminded of potentially risking getting the virus in every other part of their life could make their emotion coaching less optimal in this arena too.

We found significant effects of the covariate Covid-19's emotional impact and emotion coaching impact for the individual level effects on the implementation outcome in the main analyses. Thus, showing that Covid-19 affected the implementation in this project. However, we only controlled for five direct questions. Covid-19 might still have influenced the teachers and the project in many other ways which we could not control for, both direct and indirect. The T2 questionnaire was sent out the day before the shutdown and the implementation outcome questionnaire was sent out during the pandemic which might have influenced the response rate.

Leader Responsibility

Part of the implementation strategy for the implementation of TIK-KT into kindergartens gave the kindergarten leaders some of the responsibility for the implementation of TIK-KT in their kindergarten as they were responsible for fortnightly supervision sessions. These were key in helping the teachers solve the challenges they had while emotion coaching. The number of supervisions were controlled for by including it as a covariate. This variable was a significant covariate on all the predictor models except education for the between effects on implementation quality, and all the predictors for the between effects on implementation outcome. Thus, proving that the number of supervisions had an impact on both the quality and outcome of the implementation. Therefore, the leaders could impact the implementation through how many supervisions they had with the teachers. If the leaders held fewer supervision, they therefore could have influenced the implementation negatively. Thus, a focus on increasing supervisions in future studies could lead to a better implementation.

Previous research has linked high quality training and supervision to better implementation (Gottfredson & Gottfredson, 2002). The quality of the supervision was not assessed in this study. This could be an important factor and should be investigated further in future studies. The quality of the help provided by the leaders might vary if the leaders were less or more supportive and motivated. Thus, the input the teachers received might have varied. Other variables like the work capacity of the leader and the overall workload, their

mental and physical health, their understanding of the emotion coaching concepts etc. could have influenced the quality of these supervision sessions which in turn could influence the implementation.

Other Possible Causal Pathways

As this was not an experimental study conducted in a controlled environment, drawing clear causal explanations may be difficult as we do not know if changes were due to the implementation of TIK-KT or other factors. If there was, for example, an increased focus on emotion socialization in the media, if the kindergartens had other programs or interventions affecting the implementation of TIK-KT or if the pandemic had influenced the leaders or teachers, we were not able to capture these additional factors, this may have influenced the results.

Conclusion

The aim of the present study was to explore the effect of the predictors on implementation outcomes (H1). We hypothesized that the relationship between the different predictors and implementation outcomes would be mediated by implementation quality (H2).

Teachers and units with perception of better kindergarten functioning, teachers with more supportive emotion socialization, and units with teachers that had more experience working in a kindergarten, had better implementation outcomes. The predictors education, psychological distress, emotion dysregulation, non-supportive emotion socialization and emotional support did not affect implementation outcomes.

We found a mediating effect of better implementation quality between the three predictors: perception of better kindergarten functioning (unit and individual level), more supportive emotion socialization (unit and individual level), and greater emotional support (unit level), and the dependent variable better implementation outcomes. There was no mediating effect for the five other predictors (education, experience, psychological distress, emotion dysregulation, non-supportive emotion socialization). Implementation quality had a positive significant effect on implementation outcomes in all the predictor models both individual and unit level.

The findings from this study highlight several important factors when implementing a new intervention. Our results highlight that the functioning of the kindergarten is important when implementing a program or intervention. Focusing on improving the kindergarten functioning may improve the implementation of future programs. Our finding from the emotional support and supportive emotion socializations predictors suggests that those who already possessed some of the qualities that the intervention tried to implement would have

better implementation quality and outcomes. Thus, having better fitting candidates for the intervention in future implementations could improve the implementation of the program. However, the need for an implementation would probably be greater when the individuals lack the skillset the intervention focuses on beforehand. The kindergarten experience findings suggest that more experience would lead to a better implementation outcome. Thus, using experienced employees in future implementation may improve the implementation outcomes. Our results further support that implementation quality is significantly associated to implementation outcomes at the individual and unit level. This suggests that focusing on factors that may increase implementation quality is important to increase the implementation outcomes and thus improve the impact of an intervention. Our results also highlight the importance of focusing on implementation on both group and individual levels.

Future research should focus on the development of a psychometrically sound measures of implementation quality and implementation outcome to increase the reliability and validity of results in implementation research, as well as to increase uniformity in the implementation research field. The relationship between implementation quality and implementation outcomes should be duplicated and investigated further to figure out how these results could be transferred and applied theoretically and in practical settings to improve implementation outcomes. In addition, future studies should investigate predictors, implementation quality and implementation outcomes at the group level, since we found descriptively stronger and promising effects on this level.

References

Aghaie Meybodi, F., Mohammadkhani, P., Pourshahbaz, A., Dolatshahi, B., & Havighurst, S. S. (2019). Improving Parent Emotion Socialization Practices: Piloting Tuning in to Kids in Iran for Children With Disruptive Behavior Problems. *Family Relations, 68*(5), 596-607. doi:10.1111/fare.12387

Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior human decision processes, 50*(2), 179-211.

Baum, C. G., Forehand, R., & Zegiob, L. E. (1979). A review of observer reactivity in adult-child interactions. *Journal of Behavioral Assessment, 1*(2), 167-178. doi:10.1007/BF01322022

Beets, M. W., Flay, B. R., Vuchinich, S., Acock, A. C., Li, K.-K., & Allred, C. (2008). School Climate and Teachers Beliefs and Attitudes Associated with Implementation of the Positive Action Program: A Diffusion of Innovations Model. *Prev Sci, 9*(4), 264-275. doi:10.1007/s11121-008-0100-2

Berberoglu, A. (2018). Impact of organizational climate on organizational commitment and perceived organizational performance: empirical evidence from public hospitals. *BMC Health Services Research, 18*(1), 399. doi:10.1186/s12913-018-3149-z

Bogen, A. (2020). Implementering av Robuste barn og unge, Trygg oppvekst og Alle har en psykisk helse: En prosessevaluering av tre psykisk helsefremmende prosjekter. In.

Bohnert, A. M., Crnic, K. A., & Lim, K. G. (2003). Emotional competence and aggressive behavior in school-age children. *Journal of Abnormal Child Psychology volume, 31*(1), 79-91. doi:10.1023/a:1021725400321

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*: Harvard university press.

Cook, E. T., Greenberg, M. T., & Kusche, C. A. (1994). The relations between emotional understanding, intellectual functioning, and disruptive behavior problems in elementary-school-aged children. *Journal of Abnormal Child Psychology volume, 22*(2), 205-219. doi:10.1007/bf02167900

Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement Sci, 4*(1), 50-50. doi:10.1186/1748-5908-4-50

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: are implementation effects out of control? *Clinical psychology review, 18*(1), 23-45.

Denham, S. A., Ferrier, D. E., & Bassett, H. H. (2020). Preschool teachers' socialization of emotion knowledge: Considering socioeconomic Risk. *Journal of Applied Developmental Psychology, 69*, 101160. doi:10.1016/j.appdev.2020.101160

Domitrovich, C. E., Bradshaw, C. P., Poduska, J. M., Hoagwood, K., Buckley, J. A., Olin, S., . . . Ialongo, N. S. (2008). Maximizing the Implementation Quality of Evidence-Based Preventive Interventions in Schools: A Conceptual Framework. *Adv Sch Ment Health Promot, 1*(3), 6-28. doi:10.1080/1754730X.2008.9715730

Durlak, J. A., & DuPre, E. P. (2008). Implementation Matters: A Review of Research on the Influence of Implementation on Program Outcomes and the Factors Affecting Implementation. *Am J Community Psychol, 41*(3-4), 327-350. doi:10.1007/s10464-008-9165-0

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., Schellinger, K. B., & Durlak, J. A. (2011). The impact of enhancing students' social and emotional learning: a meta-analysis of school-based universal interventions. *Child development, 82*(1), 405-432. doi:10.1111/j.1467-8624.2010.01564.x

Eccles, M. P., & Mittman, B. S. (2006). Welcome to Implementation Science. *Implementation Science, 1*(1), 1. doi:10.1186/1748-5908-1-1

Eisenberg, N., Cumberland, A., & Spinrad, T. L. (1998). Parental Socialization of Emotion. *Psychol Inq, 9*(4), 241-273. doi:10.1207/s15327965pli0904_1

Engels, J. M., & Diehr, P. (2003). Imputation of missing longitudinal data: a comparison of methods. *J Clin Epidemiol, 56*(10), 968-976. doi:10.1016/S0895-4356(03)00170-7

Eriksen, C. L. (2018). Emosjonsveiledning i barnehage: En pilotstudie av et opplaeringsprogram i emosjonsveiledning for barnehageansatte. In.

Exner-Cortens, D., Spiric, V., Crooks, C., Syeda, M., & Wells, L. (2020). Predictors of Healthy Youth Relationships Program Implementation in a Sample of Canadian Middle School Teachers. *Canadian Journal of School Psychology, 35*(2), 100-122. doi:10.1177/0829573519857422

Fabes, R. A., Eisenberg, N., & Bernzweig, J. (2000, 25.11.2020). The Coping with Children's Negative Emotions Scale: Teacher Version. Retrieved from https://ccnes.org/ccnes-teacher-responses

Fabes, R. A., Poulin, R. E., Eisenberg, N., & Madden-Derdich, D. A. (2002). The Coping with Children's Negative Emotions Scale (CCNES): Psychometric Properties and Relations with Children's Emotional Competence. *Marriage & family review, 34*(3-4), 285-310. doi:10.1300/J002v34n03_05

Festinger, L. (1957). *A theory of cognitive dissonance*: Stanford University Press.

Gershon, R. R., Stone, P. W., Bakken, S., & Larson, E. J. J. T. J. o. N. A. (2004). Measurement of organizational culture and climate in healthcare. *34*(1), 33-40.

Gottfredson, D. C., & Gottfredson, G. D. (2002). Quality of School-Based Prevention Programs: Results from a National Survey. *The journal of research in crime and delinquency, 39*(1), 3-35. doi:10.1177/002242780203900101

Gottman, J. M., & DeClaire, J. (1997). *The heart of parenting: How to raise an emotionally intelligent child*. New York, NY: Simon & Schuster.

Gottman, J. M., Katz, L. F., & Hooven, C. (1996, 1997). *Meta-emotion : how families communicate emotionally*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Gratz, K., & Roemer, L. (2004). Multidimensional Assessment of Emotion Regulation and Dysregulation: Development, Factor Structure, and Initial Validation of the Difficulties in Emotion Regulation Scale. *Journal of Psychopathology and Behavioral Assessment, 26*, 41-54. doi:10.1007/s10862-008-9102-4

Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., & Kyriakidou, O. J. T. m. q. (2004). Diffusion of innovations in service organizations: systematic review and recommendations. *82*(4), 581-629.

Grimm, P. (2010). Social Desirability Bias. In (Vol. 2).

Havighurst, S. S., & Harley. (2007). *Tuning in to kids: Emotionally intelligent parenting: Program manual*: University of Melbourne.

Havighurst, S. S., & Harley, A. (2013). Tuning in to Kids: emotion coaching for early learning staff. *Belonging early years journal, 2*(1), 22-24.

Havighurst, S. S., Wilson, K. R., Harley, A. E., & Prior, M. R. (2009). Tuning in to kids: an emotion focused parenting program's initial findings from a community trial. *Journal of Community Psychology, 37*(8), 1008-1023. doi:10.1002/jcop.20345

Havighurst, S. S., Wilson, K. R., Harley, A. E., Prior, M. R., & Kehoe, C. (2010). Tuning in to Kids: improving emotion socialization practices in parents of preschool children's findings from a community trial. *J Child Psychol Psychiatry, 51*(12), 1342-1350. doi:10.1111/j.1469-7610.2010.02303.x

Heck, R. H., Thomas, S. L., & Tabata, L. N. (2013). *Multilevel and longitudinal modeling with IBM SPSS*: Routledge.

Holt, N., Bremner, A., Sutherland, E., Vliek, M., Passer, M., & Smith, R. (2015). *Psychology: the science of mind and behaviour* (3rd ed.). Maidenhead: McGraw-Hill Education.

Izard, C., Fine, S., Schultz, D., Mostow, A., Ackerman, B., & Youngstrom, E. (2001). Emotion Knowledge as a Predictor of Social Behavior and Academic Competence in Children at Risk. *Psychological Science, 12*(1), 18-23. doi:10.1111/1467-9280.00304

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*(1), 31-36. doi:10.1007/BF02291575

Kallestad, J. H., & Olweus, D. (2003). Predicting Teachers' and Schools' Implementation of the Olweus Bullying Prevention Program: A Multilevel Study. *Prevention & treatment, 6*(1), No Pagination Specified-No Pagination Specified. doi:10.1037/1522-3736.6.1.621a

Kant, R., & Lenka, S. K. (2013). Does Emotional Intelligence Get Affected from Some Biographical Factors- in Special Reference to Trainee Teachers? *American Journal of Educational Research, 1*(6), 216-220. Retrieved from http://pubs.sciepub.com/education/1/6/8

Karsh, B. T. (2004). Beyond usability: designing effective technology implementation systems to promote patient safety. *Qual Saf Health Care, 13*(5), 388-394. doi:10.1136/qshc.2004.010322

Kazdin, A. E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annu Rev Clin Psychol, 3*(1), 1-27. doi:10.1146/annurev.clinpsy.3.022806.091432

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. L., . . . Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychol Med, 32*(6), 959-976. doi:10.1017/s0033291702006074

Klein, K. J., & Sorra, J. S. (1996). The Challenge of Innovation Implementation. *The Academy of Management review, 21*(4), 1055-1080. doi:10.2307/259164

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med, 15*(2), 155-163. doi:10.1016/j.jcm.2016.02.012

La Paro, K. M., Hamre, B. K., & Pianta, R. C. (2012). *Classroom assessment scoring system (CLASS) manual, toddler*: Paul H. Brookes Publishing Company.

Latan, B. (1981). The psychology of social impact. *The American psychologist, 36*(4), 343-356. doi:10.1037/0003-066X.36.4.343

Levitan, L. C., & Verhulst, B. (2016). Conformity in Groups: The Effects of Others' Views on Expressed Attitudes and Attitude Change. *Political Behavior, 38*(2), 277-315. doi:10.1007/s11109-015-9312-x

Lewis, C. C., Boyd, M. R., Walsh-Bailey, C., Lyon, A. R., Beidas, R., Mittman, B., . . . Chambers, D. A. (2020). A systematic review of empirical studies examining mechanisms of implementation in health. *Implement Sci, 15*(1), 21-21. doi:10.1186/s13012-020-00983-3

Lewis, C. C., Fischer, S., Weiner, B. J., Stanick, C., Kim, M., & Martinez, R. G. (2015). Outcomes for implementation science: an enhanced systematic review of instruments using evidence-based rating criteria. *Implement Sci, 10*(1), 155. doi:10.1186/s13012-015-0342-x

Leyva, D., Weiland, C., Barata, M., Yoshikawa, H., Snow, C., Trevi?o, E., & Rolla, A. (2015). Teacher-Child Interactions in Chile and Their Associations With Prekindergarten Outcomes. *Child Dev, 86*(3), 781-799. doi:10.1111/cdev.12342

Li, H., Liu, J., & Hunter, C. V. (2020). A Meta-Analysis of the Factor Structure of the Classroom Assessment Scoring System (CLASS). *The Journal of experimental education, 88*(2), 265-287. doi:10.1080/00220973.2018.1551184

Little, M. A., Pokhrel, P., Sussman, S., & Rohrbach, L. A. (2015). The Process of Adoption of Evidence-based Tobacco Use Prevention Programs in California Schools. *Prev Sci, 16*(1), 80-89. doi:10.1007/s11121-013-0457-8

Lyon, A. R., & Bruns, E. J. (2019). From Evidence to Impact: Joining Our Best School Mental Health Practices with Our Best Implementation Strategies. *School Ment Health, 11*(1), 106-114. doi:10.1007/s12310-018-09306-w

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. J. A. R. P. (2007). Mediation analysis. *58*, 593-614.

Malloy, M., Acock, A., DuBois, D. L., Vuchinich, S., Silverthorn, N., Ji, P., & Flay, B. R. (2014). Teachers Perceptions of School Organizational Climate as Predictors of Dosage and Quality of Implementation of a Social-Emotional and Character Development Program. *Prev Sci, 16*(8), 1086-1095. doi:10.1007/s11121-014-0534-7

Martinez, R. G., Lewis, C. C., & Weiner, B. J. (2014). Instrumentation issues in implementation science. *Implement Sci, 9*(1), 118-118. doi:10.1186/s13012-014-0118-8

Martins, A., Ramalho, N., & Morin, E. (2010). A comprehensive meta-analysis of the relationship between Emotional Intelligence and health. *Personality and Individual Differences, 49*, 554-564. doi:10.1016/j.paid.2010.05.029

Mihalic, S. J. (2004). The importance of implementation fidelity. *Emotional Behavioral Disorders in Youth, 4*(4), 83-105.

Molloy, L. E., Moore, J. E., Trail, J., Van Epps, J. J., & Hopfer, S. (2013). Understanding Real-World Implementation Quality and Active Ingredients of PBIS. *Prev Sci, 14*(6), 593-605. doi:10.1007/s11121-012-0343-9

Morris, A. S., Silk, J. S., Steinberg, L., Myers, S. S., & Robinson, L. R. (2007). The Role of the Family Context in the Development of Emotion Regulation. *Social development (Oxford, England), 16*(2), 361-388. doi:10.1111/j.1467-9507.2007.00389.x

Multiple Imputation and the Expectation-Maximization Algorithm. In (pp. 511-529). New York, NY: New York, NY: Springer New York.

NAV, A. o. v. (2021). Legemeldt sykefravær etter diagnose. 4.kvartak 2016-2020 (pdf). (xlsx). . *Sykefraværstatistikk.* Retrieved from https://www.nav.no/no/nav-og-samfunn/statistikk/sykefravar-statistikk/sykefravar

Otterpohl, N., Buchenau, K., Havighurst, S., Stiensmeier-Pelster, J., & Kehoe, C. (2020). Tuning in to Kids. *29*(1), 52-60. doi:10.1026/0942-5403/a000300

Pettigrew, A. M., Woodman, R. W., & Cameron, K. S. (2001). STUDYING ORGANIZATIONAL CHANGE AND DEVELOPMENT: CHALLENGES FOR FUTURE RESEARCH. *Academy of Management journal, 44*(4), 697-713. doi:10.2307/3069411

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System™: Manual K-3*. Baltimore, MD, US: Paul H Brookes Publishing.

Proctor, Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., . . . Hensley, M. (2010). Outcomes for Implementation Research: Conceptual Distinctions, Measurement Challenges, and Research Agenda. *Adm Policy Ment Health, 38*(2), 65-76. doi:10.1007/s10488-010-0319-7

Rabin, B. A., Brownson, R. C., Haire-Joshu, D., Kreuter, M. W., & Weaver, N. L. (2008). A Glossary for Dissemination and Implementation Research in Health. *J Public Health Manag Pract, 14*(2), 117-123. doi:10.1097/01.phh.0000311888.06252.bb

Raver, C. C. (2002). Emotions Matter: Making the Case for the Role of Young Children's Emotional Development for Early School Readiness. *Social Policy Report, 16*(3), 1-20. doi:10.1002/j.2379-3988.2002.tb00041.x

Rockwood, N. J. (2017). MLmed: An SPSS macro for fitting multilevel mediation and moderated mediation models. https://njrockwood.com/mlmed.

Rohrbach, L. A., Graham, J. W., & Hansen, W. B. (1993). Diffusion of a School-Based Substance Abuse Prevention Program: Predictors of Program Implementation. *Prev Med, 22*(2), 237-260. doi:10.1006/pmed.1993.1020

Rucker, D. D., Preacher, K. J., Tormala, Z. L., & Petty, R. E. (2011). Mediation Analysis in Social Psychology: Current Practices and New Recommendations: Mediation Analysis in Social Psychology. *Social and personality psychology compass, 5*(6), 359-371. doi:10.1111/j.1751-9004.2011.00355.x

Salavera, C., Usán, P., & Teruel, P. (2019). The relationship of internalizing problems with emotional intelligence and social skills in secondary education students: gender differences. *Psicologia: Reflexão e Crítica, 32*(1), 4. doi:10.1186/s41155-018-0115-y

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological bulletin, 86*(2), 420-428. doi:10.1037/0033-2909.86.2.420

Slot, P. L., Boom, J., Verhagen, J., & Leseman, P. P. J. J. o. A. D. P. (2017). Measurement properties of the CLASS Toddler in ECEC in The Netherlands. *48*, 79-91.

Spinrad, Eisenberg, N., Gaertner, B., Popp, T., Smith, C. L., Kupfer, A., . . . Hofer, C. (2007). Relations of Maternal Socialization and Toddlers' Effortful Control to Children's Adjustment and Social Competence. *Dev Psychol, 43*(5), 1170-1186. doi:10.1037/0012-1649.43.5.1170

Spinrad, Eisenberg, N., Kupfer, A., Gaertner, B., & Michalik, N. (2004). *The coping with toddlers' negative emotions scale.* Paper presented at the Poster presented at the biennial International Conference on Infant Studies, Chicago, IL.

Statistisk Sentralbyrå (SSB). (2021, 02. March). Barnehager. Retrieved from https://www.ssb.no/utdanning/statistikker/barnehager

Tetzchner, S. v. (2012). *Utviklingspsykologi* (2. utg. ed.). Oslo: Gyldendal akademisk.

Victor, S. E., & Klonsky, E. D. (2016). Validation of a Brief Version of the Difficulties in Emotion Regulation Scale (DERS-18) in Five Samples. *Journal of Psychopathology and Behavioral Assessment, 38*(4), 582-589. doi:10.1007/s10862-016-9547-9

Wanless, S. B., Rimm-Kaufman, S. E., Abry, T., Larsen, R. A., & Patton, C. L. (2015). Engagement in Training as a Mechanism to Understanding Fidelity of Implementation of the Responsive Classroom Approach. *Prev Sci, 16*(8), 1107-1116. doi:10.1007/s11121-014-0519-6

Waschbusch, D. A., Breaux, R. P., & Babinski, D. E. J. S. m. h. (2019). School-based interventions for aggression and defiance in youth: A framework for evidence-based practice. *11*(1), 92-105.

Weiner, B. J., Lewis, C. C., Stanick, C., Powell, B. J., Dorsey, C. N., Clary, A. S., . . . Halko, H. (2017). Psychometric assessment of three newly developed implementation outcome measures. *Implement Sci, 12*(1), 108-108. doi:10.1186/s13012-017-0635-3

Williams, N. J. (2016). Multilevel Mechanisms of Implementation Strategies in Mental Health: Integrating Theory, Research, and Practice. *Adm Policy Ment Health, 43*(5), 783-798. doi:10.1007/s10488-015-0693-2

Yassien, B. M. B. (2018). Level of Emotional Intelligence among a Sample of Secondary School Teachers at Al Koura District- Jordan. *Global Journal of Human-Social Science Research*. Retrieved from https://socialscienceresearch.org/index.php/GJHSS/article/view/2722

Zhang, P., Li, C.-Z., Zhao, Y.-N., Xing, F.-M., Chen, C.-X., Tian, X.-F., & Tang, Q.-Q. (2016). The mediating role of emotional intelligence between negative life events and psychological distress among nursing students: A cross-sectional study. *Nurse Education Today, 44*, 121-126. doi:https://doi.org/10.1016/j.nedt.2016.05.025

Appendix A

Drop out analysis

Table A1

Independent sample t-test comparing the age, supportive and non-supportive emotion socialization style, emotion dysregulation, education, psychological distress and teaching experience of those who only answered T1 or answered both T1 and T2.

Measure	N	M	T	DF	P (2-tailed)
Age (1)	95	35.57	-3.15	285	.002
Age (2)	192	39.38			
Supportive ES (1)	95	6.34	0.91	285	.364
Supportive ES (2)	192	6.28			
Non-supportive ES (1)	95	2.25	2.32	285	.021
Non-supportive ES (2)	192	2.05			
Emotion dysregulation (1)	95	1.57	-0.62	285	.534
Emotion dysregulation (2)	192	1.60			
Education (1)	95	4.12	0.42	285	.679
Education (2)	192	4.04			
Psychological distress (1)	95	4.36	-2.56	157.58	.011
Psychological distress (2)	192	4.51			
Experience (1)	94	8.43	-2.10	284	.002
Experience (2)	192	11.28			

Note. 1 = those who only answered at T1. 2= those who answered both at T1 and T2, ES= emotion socialization, Nor Lang= Norwegian language, Nor nat= Norwegian nationality, a higher psychological distress measure indicate less distress.

Table A2

Chi square test comparing the gender, nationality and language of those who only answered T1 or answered both T1 and T2.

Measure	N (1)	N(2)	Df	Chisquare	p
Female	87	168	1	1.29	.256
Male	8	25			
Norwegian nationality	84	176	1	0.79	.755

Other nationality	11	16			
Norwegian language	81	174	1	1.84	0.17
Other language	14	18			

Note. 1 = those who only answered at T1. 2= those who answered both at T1 and T2.

Appendix B

Questions Measuring Perception of Kindergarten Functioning

1. Hvordan anser du kvaliteten på forholdet ditt til de andre ansatte ved barnehagen?

2. Hvordan anser du ditt forhold til foreldrene ved din barnehage?

3. Hvordan anser du kvaliteten på forholdet ditt til din nærmeste leder?

4. Hvor godt tenker du at du er i stand til å svare på barns følelsesmessige behov (særlig sterke følelser)?

5. Hvordan anser du mangfoldet når det gjelder de ansatte i din barnehage?

6. Hvordan opplever du kvaliteten på informasjonen som blir gitt fra din ledelse i forbindelse med endringer?

7. Hvordan opplever du at barnehageorganisasjonen (FUS) tilrettelegger for forbedringer og utvikling?

8. Hvordan opplever du barnas og deres foreldres medvirkningsmuligheter i barnehagehverdagen?

9. Hvordan opplever du din barnehages strategi for tilvenning av nye barn?

Svartype: 7 punkt Likert skala med svar fra 1 (svært dårlig) til 7 (svært godt).

Reverserte svar: Ingen.

Appendix C

Kessler 6 - Norwegian Translation

I løpet av de 4 foregående ukene (28 dager), hvor mye av tiden har du følt…

1. … deg så trist at ingenting kunne muntre deg opp?

2. … deg nervøs?

3. … deg rastløs eller urolig?

4. … deg håpløs?

5. … at alt krevde en innsats?

6. … deg verdiløs?

Svartype: 5-punkt Likert Skala med svar fra 1 (hele tiden) til 5 (ikke i det hele tatt).

Reverserte svar: Ingen.

Appendix D

Modifications made to develop CTNES-T from CCNES and CTNES.

Relevant story stems (one story stem equals one hypothetical situation and its possible reactions) were chosen from both measures if they were appropriate for teachers in the kindergarten setting or adjusted to fit. Five of the reactions (DR, GWR, PR, MM and EE) were kept the same except some minor changes to be relevant for kindergarten teachers. As the CCNES was made for teachers in schools, situations had to be changed to suit a kindergarten, i.e., the words classroom was avoided, student was changed to child and the situations had to be ones that one meets in the kindergarten. The situation "If my student becomes angry because he/she is unable to participate in a classroom social activity (such as a field trip), I would:" from CCNES was adapted to "If a child becomes angry because they cannot wait to have their turn on play equipment, I would:". Two new story stems were added with situations describing the drop off situation in the kindergarten and a child not sharing its toy. In CTNES, the EFR was described as distracting the child to focus on something other than the emotions (Spinrad et al., 2007) In the CTNES-T the EFR was modified so all items were about comforting and helping the child with its emotions. Two new reactions ACK and DIS were added throughout the different situations to reflect common responses to children's emotions.

Appendix E

CTNES-T Norwegian Version

Modifisert versjon for barnehagelærere. Modifisert av Sophie Havighurst ved PSI, UiO.

Oversatt og tilbakeoversatt av Ida Tidemann og Andrea Tøn (begge ved PSI, UiO).

1. Hvis et barn velter noe og lager et stort søl på gulvet, blir opprørt og gråter, ville jeg:
   a. Trøste barnet
   b. Fortelle barnet at han/hun overreagerer eller lager oppstyr for ingenting
   c. Forbli rolig og ikke hisse meg opp
   d. Sende barnet ut av rommet som konsekvens av rotet han/hun skapte
   e. Hjelpe barnet med å finne en måte å rydde opp i rotet
   f. Fortelle barnet at det er i orden å bli opprørt
   g. Anerkjenne barnets følelser
   h. Distrahere barnet

2. Hvis et barn blir sint fordi det ikke kan vente på sin tur i leken, ville jeg:
   a. Sende barnet vekk
   b. La barnet komme først i køen
   c. Distrahere barnet med andre leker
   d. Fortelle barnet at det ikke noen grunn til å bli sint
   e. Fortelle barnet at det er ok å bli sint og ha problemer med å vente
   f. Hjelpe barnet med å finne på noe annet i mellomtiden
   g. Føle meg sint som følge av barnets adferd
   h. Anerkjenne barnets følelser
   i. Trøste barnet og hjelpe det med å roe seg ned

3. Hvis et barn blir opprørt fordi jeg fjernet noe det ikke skulle ha lekt med, ville jeg:
   a. Fortelle barnet at hvis de rører det igjen, vil jeg ikke tillate mer moro
   b. Hjelpe barnet med å finne på noe annet gøy
   c. Selv bli opprørt
   d. Fortelle barnet at det er i orden å bli sint
   e. Distrahere barnet med noe annet interessant
   f. Gi barnet det det vil ha
   g. Ignorere barnets opprørthet og ta tingen fra ham/henne
   h. Anerkjenne barnets følelser
   i. Trøste barnet og hjelpe det med å roe seg ned

4. Hvis et barn leker med et puslespill eller en puttekasse, ikke klarer å plassere brikken rett og blir frustrert og gråter, ville jeg:

    a. Forbli rolig og ikke tillate meg å bli opprørt

    b. Ta leken vekk fra barnet

    c. Trøste barnet og hjelpe det med å roe seg ned

    d. Plassere brikken for barnet

    e. Fortelle barnet at det er i orden å bli frustrert og opprørt

    f. Hjelpe barnet å løse problemet med plasseringen av brikken

    g. Fortelle barnet at det ikke er noe å gråte for

    h. Anerkjenne barnets følelse av frustrasjon

    i. Distrahere barnet fra disse følelsene

5. Hvis et barn har klatret opp på et lekestativ, ikke kommer seg ned igjen, og blir nervøs og begynner å gråte, ville jeg:

    a. Selv bli engstelig og opprørt

    b. Hjelpe barnet med å finne ut hvordan det kan komme seg ned

    c. Løfte barnet ned

    d. Fortelle barnet at han/hun ikke burde klatret opp alene

    e. Fortelle barnet at det ikke er noe å bli opprørt over

    f. Trøste barnet med ord eller med et klapp

    g. Fortelle barnet at det er i orden å bli redd

    h. Anerkjenne barnets følelser

6. Hvis et barn blir sint fordi et annet barn ønsker å dele leken de er opptatt med, ville jeg:

    a. Ta leken vekk

    b. Bli sint på barnet

    c. Finne ut hvordan barna kunne delt leken

    d. Fortelle barnet at det er dumt å være egoistisk og bli sint

    e. Reflektere barnets sinne

    f. Trøste barnet og hjelpe det med å roe seg

    g. Distrahere barna med en annen leke

    h. Anerkjenne barnets følelser

7. Hvis et barn blir levert i barnehagen av sin forelder og blir opprørt og gråter ved avskjeden, ville jeg:

    a. Selv føle ubehag

b. Distrahere barnet

c. Fortelle barnet at foreldre kommer tilbake etterpå

d. Fortelle barnet at det ikke er noe å bli opprørt over

e. Trøste barnet med ord, ved å stryke eller klemme

f. Fortelle barnet at det er i orden å bli trist og savne forelderen sin

g. Fortelle barnet at det må slutte å gråte, ellers vil ingen leke med ham/henne

h. Anerkjenne barnets følelser

8. Hvis et barn faller og får en skramme i forsøk på å nå en favorittleke, ville jeg:

a. Selv bli opprørt

b. Hjelpe barnet med å finne ut hvordan man kan føle seg bedre (finne et plaster)

c. Distrahere barnet med noe annet

d. Fortelle barnet at han/hun burde være mer forsiktig

e. Fortelle barnet at dette ikke er noe å bli opprørt over

f. Fortelle barnet at det er i orden å gråte

g. Anerkjenne barnets følelser

h. Trøste barnet

9. Hvis et barn mister en ting det er glad i og reagerer med tårer, ville jeg:

a. Bli opprørt over ham/henne

b. Fortelle barnet at han/hun overreagerer

c. Hjelpe barnet med å tenke over steder han/hun ikke har lett enda

d. Distrahere barnet ved å snakke om fine ting

e. Fortelle ham/henne at det er i orden å gråte når man er lei seg

f. Fortelle ham/henne hva som skjer når man ikke passer på tingene sine

g. Anerkjenne barnets følelser

h. Trøste barnet

10. Hvis et barn deltar i en gruppeaktivitet med de andre barna og gjør en feil som medfører at han/hun ser flau ut og er på randen av tårer, ville jeg:

a. Trøste barnet

b. Fortelle barnet at han/hun overreagerer

c. Selv føle meg ukomfortabel og flau

d. Fortelle barnet at det må skjerpe seg eller forlate gruppen

e. Oppmuntre barnet til å snakke om han/hennes følelse av flauhet

f. Fortelle barnet at jeg vil hjelpe ham/henne med å øve slik at det går bedre neste gang

g. Anerkjenne barnets følelser

h. Distrahere barnet fra disse følelsene

11. Hvis et barn er ute og leker og ser ut til å være på randen av tårer fordi de andre barna er slemme og ikke vil la ham/henne leke med de, ville jeg:

a. IKKE bli opprørt selv

b. Fortelle barnet at hvis han/hun begynner å gråte vil han/hun bli nødt til å utebli fra leken

c. Fortelle barnet at det er i orden å gråte når man er lei seg

d. Trøste barnet

e. Hjelpe barnet med å finne noe annet å gjøre

f. Fortelle barnet at det ikke må bli opprørt

g. Anerkjenne barnets følelser

h. Distrahere barnet fra disse følelsene

12. Hvis et barn er sjenert og redd rundt fremmede og bestandig blir stille og tilbaketrukket når besøkende kommer inn, ville jeg:

a. Hjelpe barnet med å tenke ut strategier for å gjøre møter med nye mennesker mindre skremmende

b. Fortelle barnet at det er i orden å bli nervøs

c. Trøste barnet og oppmuntre det til å delta når han/hun er klar for det

d. Føle meg opprørt og ukomfortabel på grunn av barnets reaksjoner

e. Fortelle barnet at han/hun må samhandle på passende måte med besøkende

f. Fortelle barnet at han/hun er barnslig eller umoden

g. Anerkjenne barnets følelser

h. Distrahere barnet

Appendix F

The Implementation Quality Questionnaire (IQQ-39) in Norwegian

Om tiltaket og prosjektet «Barnet Først».

Her ber vi deg ta stilling til de følgende utsagnene på en skala fra 1 (svært uenig) til 7 (svært enig).

1. Opplæringen/veiledningen av de ansatte gjennomføres ofte nok.

2. Jeg føler eierskap til «Barnet først».

3. Jeg har tilstrekkelig kompetanse for å kunne gjennomføre dette tiltaket på en god måte.

4. Innholdet i «Barnet Først» er for enkelt.

5. Forholdet mellom de barnehageansatte og ledelsen bør forbedres.

6. Før «Barnet Først» har det vært utfordrende å samarbeide med FUS-ledelsen.

7. Det har kommet for lite informasjon vedrørende «Barnet Først».

8. Det er tydelige indikasjoner på at barnehagen prioriterer dette prosjektet.

9. Prosjektet «Barnet Først» ble implementert på et godt tidspunkt.

10. De ansatte i denne barnehagen ble involvert i planleggingsprosessen.

11. FUS-ledelsen prioriterer ikke tiltak for heving av emosjonell kompetanse.

12. «Barnet først» er svakt forankret i barnehagen.

13. Det er behov for tiltak vedrørende emosjonell kompetanse i norske barnehager i dag.

14. Det er ikke mulig å gjennomføre «Barnet Først» slik det er planlagt.

15. «Barnet Først» krever for mye av de ansatte.

16. «Barnet Først» er dårligere enn andre realistiske alternativer.

17. «Barnet Først» stemmer ikke overens med eksisterende forventninger, normer og behov i barnehagene.

18. Det er god kommunikasjon internt i barnehagen.

19. De som har formidlet informasjonen om at «Barnet Først» skulle gjennomføres har vært entusiastiske og vist at de har hatt tro på intervensjonen.

20. Kun deler av barnehagen er involvert i «Barnet Først».

21. Jeg er positiv til innføringen av «Barnet først» i barnehagen.

22. Innholdet i «Barnet Først» er praktisk og lett anvendbart.

23. «Barnet først» møter denne barnehagens utfordringer.

24. Det er enighet rundt normer og verdier i barnehagen.

25. Målsettingene for «Barnet Først» er tydelig formulert og kommunisert.

26. Det er ikke satt av nok tid til å jobbe med «Barnet Først».

27. Barnehagen var ikke klar for å innføre «Barnet Først».

28. Det har vært rom for å gi tilbakemeldinger underveis.

29. Det er en stor motstand mot innføringen av «Barnet først» i barnehagen.

30. Det har vært lite rom for diskusjon, refleksjon og øving under tiden med «Barnet Først».

31. «Barnet Først» har en svak teoretisk forankring.

32. Det har vært flere utskiftninger i barnehagens bemanning den siste tiden.

33. Innføringen av «Barnet Først» går på bekostning av andre viktige oppgaver.

34. Det er brukt for lite tid på å planlegge og forberede implementeringen.

35. Barnehagens ledelse ble involvert i planleggingsprosessen.

36. Det er en rød tråd mellom de ulike elementene i «Barnet først».

37. «Barnet Først» styrker de ansattes evner til å støtte utviklingen av barnas emosjonelle kompetanse.

38. Det finnes ikke tilstrekkelig kompetanse i FUS-ledelsen for å innføre «Barnet Først» på en god måte.

39. Det er et godt samarbeid med FUS-ledelsen nå.

Appendix G

Implementation Outcomes - Norwegian Translations

Tre implementeringsutfall foreslått av Weiner et al. (2017). Oversatt versjon.

Aksept av Intervensjon Mål (AIM) – Accceptability

1. «Barnet først» får min godkjenning.

2. «Barnet først» appellerer til meg.

3. Jeg liker «barnet først».

4. Jeg ønsker «barnet først» velkommen.

Egnethet av Intervensjon Mål (EIM) – Appropriateness

1. «Barnet først» virker passende.

2. «Barnet først» virker egnet.

3. «Barnet først» virker anvendelig.

4. «Barnet først» virker som en god match.

Gjennomførbarhet av Intervensjon Mål (GIM) – Feasibility

1. «Barnet først» virker mulig å innføre.

2. «Barnet først» virker overkommelig.

3. «Barnet først» virker gjennomførbart.

4. «Barnet først» virker enkel å bruke.

Tre implementeringsutfall laget på norsk for denne studien. Teoretisk grunnlag: Weiner et al. (2017) og Proctor et al. (2010).

Intervensjonens opprettholdelse mål- Sustainability

1. «Barnet først» har blitt lettere å bruke over tid.

2. Jeg har brukt emosjonsveiledning i økende grad siden prosjektet startet

3. «Barnet først» har blitt en naturlig del av vår rutine.

4. Jeg kommer til å fortsette å bruke det jeg har lært via "Barnet først"

Penetration. Basert på Proctor et al (2010).

I hvor stor % av de mulige "barnet først" situasjonene brukte du emosjonell veiledning? (Bruke samme visuelle analoge skala som for livskvalitet: mål fra 0 til 100, men spesifiserer at det er %.)

Fidelity.

«I min arbeidshverdag er jeg i stand til å benytte med av følgende emosjonsveiledningsegenskaper:

Ta stilling til hver av egenskapene ved å krysse av på skalaen fra 1 (svært uenig) til 7 (svært enig). Dersom du er daglig leder og IKKE arbeider med barna, vennligst kryss av for det.»

1. Legge merke til barnets følelser
2. Stoppe og knytte meg til barnet – nærme meg barnets nivå
3. Vise empati
4. Sette navn på følelsen
5. Bli ved (tåle) følelsen, eller roe ned frem til barnets følelser har roet seg
6. Avstå fra å løse problemet / grensesette for tidlig
7. Hjelpe med å løse problemet på en passende måte
8. Sette grenser på en passende måte
9. Håndtere mine egne følelser

Appendix H

Overview of Questions Included in the Covid-19 Measure in Norwegian

Under følger noen utsagn vedrørende din opplevelse av denne perioden.

Vi ber deg indikere ved bruk av skalaen fra 1 (svært uenig) til 7 (svært enig) hvor godt utsagnet passer for deg.

(Hvis du jobbet under perioden med nedstenging ber vi deg ta utgangspunkt i hvordan det føltes når barnehagene åpnet igjen, og alle returnerte til jobb.)

1. Å arbeide i barnehagen i løpet av denne perioden har vært mer stressende enn vanlig.

2. Denne perioden har vært mer stressende enn vanlig for meg personlig.

3. Jeg var komfortabel med å returnere til jobb etter perioden med nedstenging*

4. Covid-19-situasjonen har medført flere muligheter for emosjonsveiledning av barna i barnehagen og kollegaer for meg*

5. Covid-19-situasjonen har redusert min kapasitet til emosjonsveiledning.

*=Reversed items

Appendix I

NSD Ethical Approvement

NSD NORSK SENTER FOR FORSKNINGSDATA

NSD's assessment

Project title

The evaluation of Tuning in to Kids in Norwegian Kindergartens

Reference number

651181

Registered

26.04.2019 av Ida Therese Tidemann - i.t.tidemann@psykologi.uio.no

Data controller (institution responsible for the project)

Universitetet i Oslo / Det samfunnsvitenskapelige fakultet / Psykologisk institutt

Project leader (academic employee/supervisor or PhD candidate)

Sophie Havighurst, s.s.havighurst@psykologi.uio.no, tlf: 22845111

Type of project

Research Project

Project period

01.05.2019 - 31.12.2021

Status

11.06.2019 - Assessed

Assessment (1)

11.06.2019 - Assessed

Our assessment is that the processing of personal data in
this project will comply with data protection legislation, so long as it is
carried out in accordance with what is documented in the Notification Form and
attachments, dated 11.06.2019, as well as in correspondence with NSD. Everything
is in place for the processing to begin. NOTIFY CHANGES If you intend to make
changes to the processing of personal data in this project it may be necessary

to notify NSD. This is done by updating the Notification Form. On our website we explain which changes must be notified. Wait until you receive an answer from us before you carry out the changes. TYPE OF DATA AND DURATION The project will be processing special categories of personal data about health, and general categories of personal data, until 31.12.2021. LEGAL BASIS The project will gain consent from data subjects to process their personal data. We find that consent will meet the necessary requirements under art. 4 (11) and 7, in that it will be a freely given, specific, informed and unambiguous statement or action, which will be documented and can be withdrawn. The legal basis for processing special categories of personal data is therefore explicit consent given by the data subject, cf. the General Data Protection Regulation art. 6.1 a), cf. art. 9.2 a), cf. the Personal Data Act § 10, cf. § 9 (2). PRINCIPLES RELATING TO PROCESSING PERSONAL DATA NSD finds that the planned processing of personal data will be in accordance with the principles under the General Data Protection Regulation regarding: - lawfulness, fairness and transparency (art. 5.1 a), in that data subjects will receive sufficient information about the processing and will give their consent - purpose limitation (art. 5.1 b), in that personal data will be collected for specified, explicit and legitimate purposes, and will not be processed for new, incompatible purposes - data minimisation (art. 5.1 c), in that only personal data which are adequate, relevant and necessary for the purpose of the project will be processed - storage limitation (art. 5.1 e), in that personal data will not be stored for longer than is necessary to fulfil the project's purpose THE RIGHTS OF DATA SUBJECTS Data subjects will have the following rights in this project: transparency (art. 12), information (art. 13), access (art. 15), rectification (art. 16), erasure (art. 17), restriction of processing (art. 18), notification (art. 19), data portability (art. 20). These rights apply so long as the data subject can be identified in the collected data. NSD finds that the information that will be given to data subjects about the processing of their personal data will meet the legal requirements for form and content, cf. art. 12.1 and art. 13. We remind you that if a data subject contacts you about their rights, the data controller has a duty to reply within a month. FOLLOW YOUR INSTITUTION'S GUIDELINES NSD presupposes that the project will meet the requirements of accuracy (art. 5.1 d), integrity and confidentiality (art. 5.1 f) and security (art. 32) when processing personal data. TSD (Tjenester for Sensitive Data) is a data processor for the project. NSD presupposes that the processing of personal data by a data processor meets the requirements under the General Data Protection Regulation arts. 28 and 29. To ensure that these requirements are met you must follow your institution's internal guidelines and/or consult with your institution (i.e. the institution responsible for the project). FOLLOW-UP OF THE PROJECT NSD will follow up the progress of the project underway (every other year) and at the planned end date in order to determine whether the processing of personal data has been concluded/is being carried out in accordance with what is documented. Good luck with the project! Contact person at NSD: Jørgen Wincentsen Data Protection Services for Research: +47 55 58 21 17 (press 1)

Appendix J

Cronbach's alpha and inter item correlation mean (IICM) for the remaining 7 subscales

in CTNES-T

Subscale	Cronbach's alpha	IICM
DR	.493	.110
GWR	.340	141
EFR	.837	.344
MR	.806	.258
PR	.605	.122
PFR	.703	.186
EE	.797	.328
ACK	.942	.592
DIS	.911	.485

Note. DR= Teacher Distress Reactions, GWR= Granting Children's Wishes, EFR = Emotion Focused Responses, MR= Minimizing Reactions, PR= Punitive Reactions, PFR= Problem Focused Responses, EE= Expressive Encouragement, ACK= Acknowledge the child's emotions, DIS= Distract.

[a] N= 287

Appendix K

Main analyses, general linear model (within) and linear mixed mode (between)with covariates

Estimates and standard error (SE) from the Main General Linear Model (within) and Linear Mixed Model (between) Analyses Between Predictors and Implementation Outcomes Controlled for the Three Covariates Covid, Birthyear and Number of Supervision Sessions.

Predictor	X to Y (SE)	C to Y (SE)	B to y	S to Y
		Within		
Education	0.02 (0.01)	-0.07*** (0.02)	-0.02 (0.01)	0.01 (0.02)
Experience	-0.00 (0.02)	-0.07*** (0.02)	-0.02 (0.02)	0.01 (0.02)
Kindergarten functioning	0.03* (0.01)	-0.07*** (0.02)	-0.01 (0.01)	0.01 (0.02)
Psychological distress	-0.01 (0.01)	-0.07*** (0.02)	-0.02 (0.01)	0.01 (0.02)
Emotion dysregulation	0.02 (0.01)	-0.07*** (0.02)	-0.02 (0.01)	0.01 (0.02)
Supportive ES	0.04** (0.02)	-0.07*** (0.02)	-0.01 (0.01)	0.01 (0.02)
Non-supportive ES	-0.02 (0.02)	-0.06*** (0.02)	-0.02 (0.02)	0.01 (0.02)
		Between		
Education	0.06 (0.05)	0.02 (0.03)	-0.18*** (0.04)	0.12*** (0.03)
Experience	0.16** (0.05)	0.00 (0.03)	-0.05 (0.06)	0.12*** (0.03)
Kindergarten functioning	0.09* (0.04)	0.06 (0.03)	-0.19*** (0.04)	0.12*** (0.03)
Psychological distress	0.00 (0.05)	0.03 (0.03)	-0.20*** (0.05)	0.13*** (0.03)
Emotion dysregulation	-0.05 (0.04)	0.03 (0.03)	-0.19*** (0.04)	0.12*** (0.03)
Supportive ES	0.06 (0.04)	0.04 (0.03)	-0.19*** (0.04)	0.13*** (0.03)

Non-supportive ES	-0.06 (0.04)	0.04 (0.03)	-0.18*** (0.04)	0.12*** (0.03)
Emotional support	0.04 (0.03)	0.04 (0.04)	-0.19***(0.04)	0.13***(0.03)

Note. Predictors were self-report measures at teacher level except CLASS Emotional Support which was observed at unit level. Data were imputed, winsorized and Z-scored. X= predictor, Y= implementation outcome, C=Covid-19, B=Birthyear, S=Number of Supervisions, ES= Emotion socialization, within= individual level controlled for unit grouping effect, between= unit level effect, Education scale: 1 (primary and/or secondary), 2 (vocational school), 3 (high school), 4 (courses after high school), 5 (higher education 1-3 years), 6 (higher education 5-6 years). A higher psychological distress measure indicate less distress.

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

[a] Within: N=344 Between: N=108.

Appendix L

Main analyses; multilevel mediation analyses with covariates

Estimates and standard error (SE) from multilevel mediation analyses with implementation outcome as dependent variable and implementation quality as mediator, with covid, birthyear and number of supervision sessions as covariates.

Predictor	X to M (SE)	C to M (SE)	B to M (SE)	S to M (SE)	X to Y (SE)	M to Y (SE)	C to Y (SE)	B to Y (SE)	S to Y (SE)	Indirect (SE)
					Within					
Edu.	0.01(0.06)	-0.09(0.06)	0.01(0.06)	0.02(0.08)	0.02(0.01)	0.07***(0.02)	-0.06*** (0.02)	-0.02(0.01)	-0.01(0.02)	0.00(0.00)
Exp.	-0.17(0.09)	-0.07(0.07)	-0.10(0.08)	0.03(0.08)	0.01(0.02)	0.07***(0.01)	-0.06*** (0.02)	-0.01(0.02)	0.01(0.02)	-0.01(0.01)
Rating	0.20**(0.07)	-0.08(0.07)	0.04(0.06)	0.02(0.08)	0.02(0.01)	0.07***(0.02)	-0.06***(0.02)	-0.02(0.01)	0.01(0.02)	0.01*(0.01)
K6	0.07(0.07)	-0.08(0.07)	0.03(0.07)	0.03(0.08)	-0.01(0.01)	0.07***(0.02)	-0.06***(0.02)	-0.02(0.01)	0.01(0.02)	0.01(0.01)
DERS	-0.11(0.07)	-0.08(0.07)	0.02(0.06)	0.02(0.08)	0.03(0.01)	0.07***(0.02)	-0.06***(0.02)	-0.02(0.01)	0.01(0.02)	-0.01(0.01)
Supp	0.20**(0.07)	-0.09(0.07)	0.04(0.06)	-0.00(0.08)	0.03(0.02)	0.07***(0.02)	-0.06***(0.02)	-0.02(0.01)	0.01(0.02)	0.01*(0.01)
N-supp	-0.13(0.07)	-0.07(0.07)	0.01(0.06)	-0.00(0.08)	-0.01(0.02)	0.07***(0.02)	-0.06***(0.02)	-0.02(0.01)	-0.01(0.02)	-0.01(0.01)
					Between					
Edu.	0.18(0.11)	-0.17*(0.08)	-0.05(0.10)	0.34***(0.08)	0.02(0.03)	0.18***(0.03)	0.02(0.03)	-0.15***(0.03)	0.06*(0.02)	0.03(0.02)
Exp.	0.15(0.14)	-0.18*(0.09)	0.03(0.15)	0.35***(0.08)	0.11**(0.04)	0.17***(0.03)	0.00(0.03)	-0.06(0.04)	0.06*(0.02)	0.03(0.03)
Rating	0.28**(0.10)	-0.07(0.09)	-0.06(0.10)	0.34***(0.07)	0.04(0.03)	0.18***(0.03)	0.03(0.03)	-0.15***(0.03)	0.06*(0.02)	0.05*(0.02)
K6	0.06(0.12)	-0.16(0.09)	-0.06(0.11)	0.36***(0.08)	-0.01(0.04)	0.19***(0.03)	0.02(0.03)	-0.16***(0.04)	0.06*(0.02)	0.01(0.02)
DERS	-0.16(0.10)	-0.15(0.08)	-0.06(0.10)	0.36***(0.08)	-0.01(0.03)	0.18***(0.03)	0.02(0.03)	-0.15***(0.03)	0.06*(0.02)	-0.03(0.02)
Supp.	0.21*(0.10)	-0.13(0.08)	-0.04(0.10)	0.37***(0.07)	-0.00(0.03)	0.19***(0.03)	0.02(0.03)	-0.15***(0.03)	0.06*(0.02)	0.04*(0.02)
N-supp.	-0.14(0.10)	-0.14(0.08)	-0.04(0.11)	0.35***(0.08)	-0.03(0.03)	0.18***(0.03)	0.02(0.03)	-0.15***(0.02)	0.06*(0.02)	-0.03(0.02)
ES	0.14*(0.06)	-0.14(0.08)	-0.06(0.10)	0.40***(0.08)	0.00(0.02)	0.18***(0.03)	0.02(0.03)	-0.15***(0.03)	0.06*(0.02)	0.03*(0.01)

Note. Results from a Multilevel Mediation Analysis with Implementation Quality as Mediator and Implementation Outcome as the Dependent Variable Controlled for Three Covariates with Imputed and Winsorized Data. Estimates presented are z-scores. SE= Standard error, Edu.= Education, Exp.=Experience, Rating=Perception of kindergarten functioning, K6=Kessler6 (measures psychological distress), DERS= Difficulty with emotion regulation 18 (measures emotional regulation), Supp= Supportive emotion socialization reactions, N-supp.= Non-supportive emotion socialization reactions, X= predictor, M= Implementation Quality Questionnaire, C=Covid-19, B= Birthyear, S=Number of Supervisions and Y = implementation outcome. A higher psychological distress measure indicate less distress.

[a] N=330.

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

Appendix M

Sensitivity analyses; general linear model (within) and linear mixed model (between) with covariates

Estimates and standard error (SE) from the general linear model (within) and linear mixed model (between) sensitivity analyses between

predictors and implementation outcomes controlled for the three covariates covid, birthyear and number of supervision sessions.

Predictor	X to Y (SE)	C to Y (SE)	B to y	S to Y
		Within		
Education	0.02(0.08)	-0.20* (0.08)	-0.20 (0.10)	0.12 (0.11)
Experience	-0.30 (0.19)	-0.18* (0.08)	-0.34* (0.13)	0.15 (0.11)
Kindergarten functioning	0.07 (0.08)	-0.18* (0.08)	-0.17 (0.10)	0.11 (0.11)
Psychological distress	-0.06 (0.09)	-0.21* (0.08)	-0.22* (0.10)	0.12 (0.11)
Emotion dysregulation	0.04 (0.06)	-0.21* (0.06)	-0.21 (0.10)	0.12 (0.11)
Supportive ES	0.13 (0.08)	-0.18* (0.08)	-0.18 (0.10)	0.07 (0.11)
Non-supportive ES	-0.07 (0.09)	-0.17 (0.09)	-0.19 (0.10)	0.09 (0.12)
		Between		
Education	0.30**(0.10)	-0.03 (0.09)	-0.39*** (0.09)	0.08 (0.06)
Experience	0.18 (0.17)	-0.02 (0.09)	-0.30* (0.14)	0.11 (0.07)
Kindergarten functioning	0.25* (0.10)	0.06 (0.09)	-0.39*** (0.09)	0.10 (0.06)
Psychological distress	0.02 (0.14)	-0.01 (0.09)	-0.40*** (0.11)	0.12 (0.07)
Emotion dysregulation	-0.18 (0.10)	-0.01 (0.09)	-0.38*** (0.10)	0.11 (0.06)
Supportive ES	0.12 (0.11)	-0.01 (0.09)	-0.39*** (0.10)	0.12 (0.06)

Non-supportive ES	-0.12 (0.10)	0.01 (0.09)	-0.36** (0.10)	0.11 (0.06)
Emotional support	0.07 (0.07)	-0.12 (0.09)	-0.20 (0.11)	0.01(0.09)

Note. Note. Predictors were self-report measures at teacher level except CLASS Emotional Support which was observed at unit level. X= predictor, Y= implementation outcome, C=Covid-19, B=Birthyear, S=Number of Supervisions, ES= Emotion socialization, within= individual level controlled for unit grouping effect, between= unit level effect, Education scale: 1 (primary and/or secondary), 2 (vocational school), 3 (high school), 4 (courses after high school), 5 (higher education 1-3 years), 6 (higher education 5-6 years). A higher psychological distress measure indicate less distress.

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

[a] Within: N=88 Between: N=61 for all except Emotional support N=48.

Appendix N

Sensitivity analyses; multilevel mediation analyses with covariates

Estimates and standard error (SE) from multilevel mediation sensitivity analyses with implementation outcome as dependent variable and implementation quality as mediator, with covid, birthyear and number of supervision sessions as covariates.

Predictor	X to M (SE)	C to M (SE)	B to M (SE)	S to M (SE)	X to Y (SE)	M to Y (SE)	C to Y (SE)	B to Y (SE)	S to Y (SE)	Indirect (SE)
					Within					
Edu.	-0.08(0.12)	-0.37*(0.15)	-0.30(0.15)	0.25(0.18)	0.05(0.06)	0.20**(0.07)	-0.14*(0.07)	-0.07(0.07)	0.08(0.08)	-0.02(0.03)
Exp.	-0.06(0.34)	-0.37*(0.15)	-0.32(0.21)	0.24(0.19)	-0.04(0.15)	0.21**(0.07)	-0.14*(0.07)	-0.10(0.10)	0.10(0.08)	-0.01(0.08)
Rating	0.38**(0.12)	-0.26(0.13)	-0.24(0.14)	0.16(0.16)	0.03(0.07)	0.18*(0.08)	-0.15*(0.07)	-0.08(0.07)	0.09(0.08)	0.07(0.04)
K6	-0.00 (0.14)	-0.38*(0.15)	-0.27(0.16)	0.24(0.18)	-0.02(0.06)	0.21**(0.07)	-0.15*(0.07)	-0.09(0.08)	0.10(0.08)	-0.00(0.03)
DERS	-0.15(0.13)	-0.36*(0.14)	-0.23(0.15)	0.25(0.17)	0.08(0.06)	0.22**(0.07)	-0.16**(0.07)	-0.09(0.07)	0.09(0.08)	-0.03(0.03)
Supp.	0.40**(0.14)	-0.34*(0.14)	-0.23(0.14)	0.09(0.17)	0.05(0.07)	0.19*(0.08)	-0.15*(0.07)	-0.07(0.07)	0.09(0.08)	0.07(0.04)
Nsupp.	-0.26(0.14)	-0.25(0.16)	-0.25(0.15)	0.13(0.18)	-0.02(0.07)	0.19*(0.07)	-0.14(0.07)	-0.08(0.07)	0.09(.09)	-0.05(.04)
					Between					
Edu.	0.37*(0.18)	-0.19(0.14)	-0.27(0.16)	0.15(0.11)	0.14(0.09)	0.30***(0.07)	-0.04(0.07)	-0.33***(0.08)	0.06(0.05)	0.11(0.06)
Exp.	-0.10(0.30)	-0.15(0.15)	-0.36(0.24)	0.20(0.11)	0.18(0.14)	0.33***(0.07)	-0.04(0.07)	-0.23*(0.12)	0.06(0.05)	-0.03(0.10)
Rating	0.46**(0.17)	-0.02(0.15)	-0.27(0.15)	0.17(0.10)	0.07(0.09)	0.31***(0.07)	-0.00(0.07)	-0.34***(0.08)	0.07(0.05)	0.14*(0.06)
K6	0.37 (0.23)	-0.16(0.15)	-0.15(0.18)	0.18(0.11)	-0.10(0.11)	0.34***(0.07)	-0.01(0.07)	-0.37***(0.09)	0.07(0.05)	0.13(0.08)
DERS	-0.40*(0.17)	-0.11(0.14)	-0.23(0.16)	0.18(0.11)	-0.02(0.09)	0.33***(0.07)	-0.03(0.07)	-0.34***(0.08)	0.07(0.05)	-0.13*(0.06)
Supp	0.50**(0.14)	-0.14(0.13)	-0.18(0.14)	0.22*(0.10)	-0.04(0.09)	0.34***(0.07)	-0.02(0.07)	-0.34***(0.08)	0.07(0.05)	0.17**(0.06)
Nsupp	-0.33(0.17)	-0.08(0.15)	-0.16(0.17)	0.15(0.11)	-0.03(0.09)	0.32***(0.07)	-0.01(0.07)	-0.33***(0.08)	0.07(0.05)	-0.11(0.06)
ES	0.13(0.12)	-0.23(0.17)	-0.25(0.22)	0.13(0.17)	0.03(0.05)	0.26***(0.07)	-0.09(0.07)	-0.17(0.09)	0.02(0.07)	0.04(0.03)

Note. Slopes and Indirect Effect for the Within Effect Results for Unit Level from a Multilevel Mediation Analysis with Implementation Quality as Mediator and Implementation Outcome as the Dependent Variable Controlled for Three Covariates with Non-imputed and Non-winsorized Data. Estimates presented are z-scores. Edu.= Education, Exp.=Experience, Rating=Perception of kindergarten functioning, K6=Kessler6/Psychological distress (a higher score indicated less distress), DERS= Difficulty with emotion regulation 18, Supp= Supportive emotion socialization reactions, N-supp.= Non-supportive emotion socialization reactions, ES= emotional support, X= predictor, M= Implementation Quality Questionnaire, C=Covid-19, B= Birthyear, S=Number of Supervisions and Y = implementation outcome.

[a] N=88 for all predictors except ES; N(ES)=74.

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

Appendix O

Complete correlation matrix for all the measures used in the thesis.

Measures	Education	Experience	Rating	K6	DERS	Supp	N-supp	ES	Birthyear	Supervisions	Covid	IQQ	Outcome
Education													
Experience	.083												
Rating	-.010	.095											
K6	0.00	.261***	.189										
DERS	-.011	-.075	-.208***	-.525***									
Supp	.151**	.104	.247***	.212***	-.213***								
N-supp	-.242***	-.311**	.062	-.305***	.306***	-.254***							
ES	.010	.044	.082	.068	-.151**	.057	-.105						
Birthyear	-.055	-.690**	-.147**	-.323***	.066	-.187***	.110*	-.034					
Supervisions	.143**	.117*	-.001	-.006	.000	.025	-.089	-.169**	-0.56				

Covid-19	.083	.101	-.165**	-.113*	.081	.030	.048	-.052	-.048	.127*		
IQQ	.075	-.005	.234***	.063	-.125*	.184**	-.145**	.080	-.022	.217***	-.037	
Outcome	.155**	.230**	.177**	.084	-.045	.140**	-.186**	.048	-.215**	.247**	.038	.436***

Note. Rating=Perception of kindergarten functioning, K6= psychological distress (a higher score indicated less distress), DERS= difficulties with emotion regulation 18, Supp= Supportive emotion socialization reactions, N-supp.= Non-supportive emotion socialization reactions, ES= emotion support, IQQ= implementation quality, Outcome= implementation outcome,

[a] N=344.

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