

**Validation of the Classroom
Assessment Scoring System (CLASS)
Pre-K and Toddler in Norwegian
kindergartens**

Susanne M. Karlsen & Nina Lode



Submitted as Cand. Psychol Thesis
Department of Psychology

UNIVERSITY OF OSLO

April 2021

Validation of the Classroom Assessment Scoring System (CLASS) Pre-K and Toddler in Norwegian kindergartens

By Susanne M. Karlsen & Nina Lode

© Susanne M. Karlsen & Nina Lode

2021

Validation of the Classroom Assessment Scoring System (CLASS) Pre-K and Toddler in Norwegian kindergartens.

Susanne M. Karlsen & Nina Lode

<http://www.duo.uio.no>

Trykk: Reprosentralen, Universitetet i Oslo

Abstract

Authors: Susanne M. Karlsen and Nina Lode

Title: Validation of the Classroom Assessment Scoring System (CLASS) Pre-K and Toddler in Norwegian kindergartens

Supervisors: Egil Nygaard and Sophie Havighurst

Background: Caregivers and other adults play significant roles in shaping children's development, both cognitive, social and emotional. Caregivers are often mentioned as the primary socializers of emotion, but recent research also points to the influence of other adults, such as teachers in schools and kindergartens. In addition, kindergarten quality in itself has been shown to affect outcomes in children, in that good quality can enhance both social, cognitive and emotional development. Kindergarten quality is admittedly difficult to measure, and process quality in particular. Process quality concerns children's experiences in the classroom with a particular focus on their interaction with teachers, peers and materials and activities. The Classroom Assessment Scoring System (CLASS) is an observational tool developed to assess process quality, and more specifically to evaluate the teacher-child relationship. Neither the Pre-K (4-6 years) or Toddler (1-3 years) versions of CLASS have been validated in a Norwegian sample. The main goal of the present study is to validate the CLASS Pre-K and Toddler in a Norwegian kindergarten sample, by investigating the construct and criterion-related validity.

Method: The sample used in the present study was drawn from 27 Norwegian kindergartens that were the control group of a larger Tuning in to Kids intervention trial (N = 49). All kindergartens in the sample were observed using CLASS. In addition, 221 teachers and 187 caregivers of children responded to self- and other-report measures (CTNES for teachers, ERC and BPM for caregivers). The construct validity of CLASS was investigated by performing factor analyses separately for Pre-K and Toddler versions of CLASS. Criterion-related validity was investigated by performing mixed models for CTNES, ERC, and BPM, using the Emotional Support (Pre-K) and Emotional and Behavioural Support (Toddler) domains from CLASS.

Results: The factor analyses supported the three-factor structure for Pre-K and the two-factor structure for Toddler, in accordance with the original factor structures. A positive relationship was found between the CLASS domains Emotional Support (Pre-K) and Emotional and Behavioural Support (Toddler) and unsupportive CTNES. No other significant relationships were found between CLASS and the outcome variables.

Conclusion: The study concluded adequate validity for CLASS Pre-K, but the validity for the Toddler version was less clear. No evidence of good concurrent- or predictive validity was found. More research on both Pre-K and Toddler is needed in Norway, as this study had several limitations. As the limitations were mostly related to the criterion-related validity, the greatest emphasis should be put on the findings on construct validity. Norway has strict governmental guidelines regarding structural quality in kindergartens such as teacher-child ratio and teachers education level. Process quality, on the other hand, is not regulated. The need for a valid instrument of process quality still remains – this would enhance research on quality in Norwegian kindergartens, and could help guide future areas of improvement in quality work.

Acknowledgements

First and foremost, we would like to thank our main supervisor Egil – for spending countless hours teaching us about quantitative methods, SPSS syntax, thesis writing, and for always cheering us on. You have lifted our thesis and our spirits many times.

A big thanks also goes out to our co-supervisor Sophie for providing invaluable feedback and for sharpening our language and thesis.

Thanks also to the N-TIK-KT project for allowing us to use their data material in our thesis, and for encouraging us to complete CLASS and TIK training.

Nina would like to thank her daughter, Josefine, for filling her days with laughter and joy, forcing her to take breaks from writing this thesis, and for making kindergarten research especially interesting. She would also like to thank her husband, Jone, for giving her his love, patience, and care.

Susanne would like to thank Matias for providing endless support, warm hugs and much needed walks around the neighborhood.

Lastly, we would like to thank each other for great teamwork. Our collaboration initially sprang from a common research interest, but more importantly it resulted in a new friendship.

Table of Contents

1 Introduction	1
1.1 Brief description of thesis	1
1.2 Basic emotions and emotional competence	2
1.2.1 Basic emotions	2
1.2.2 Emotional competence	2
1.3 Emotion socialization	3
1.3.1 Emotion socialization from caregivers	3
1.3.2 Emotion socialization from kindergarten teachers	4
1.4 Kindergartens in Norway	5
1.5 Assessment of kindergarten quality	6
1.6 Process quality in Norwegian kindergartens	6
1.6.1 How does process quality impact children?	6
1.6.2 Do Norwegian kindergartens have high process quality?	7
1.6.3 Knowledge gap	8
1.7 CLASS	8
1.7.1 CLASS Pre-K	9
1.7.2 CLASS Toddler	10
1.7.3 CLASS, process quality and emotion socialization	10
1.8 Validity of CLASS	10
1.8.1 Construct validity	11
1.8.2 Criterion-related validity	12
1.9 The present study	14
2 Method	16
2.1 Procedure	16
2.1.1 Ethical considerations	16
2.2 Participants	17
2.2.1 Kindergarten teachers	17
2.2.2 Caregivers and children	17
2.2.3 Exclusion criteria and imputation	18
2.3 Measures	18
2.3.1 Classroom Assessment Scoring System	18
2.3.2 Coping with Toddler's Negative Emotions Scale	19
2.3.3 Emotion Regulation Checklist	21
2.3.4 Brief Problem Monitor	22
2.4 Statistical analysis	22
2.4.1 Factor structure of CLASS	22
2.4.2 Multilevel modelling	22
2.4.3 Parametric tests	24
3 Results	26
3.1 Research questions	26

3.2 First hypothesis: construct validity of CLASS	26
3.2.1 CLASS Pre-K	26
3.2.1.1 Exploratory factor analysis	26
3.2.1.2 Confirmatory factor analysis	27
3.2.2 CLASS Toddler	29
3.2.2.1 Exploratory factor analysis	29
3.2.2.2 Confirmatory factor analysis	30
3.3 Preliminary preparations	32
3.4 Second hypothesis: concurrent validity of CLASS	33
3.4 Third hypothesis: predictive validity of CLASS	35
3.5 Sensitivity analysis	37
4 Discussion	38
4.1 Purpose of the study	38
4.2 Construct validity	38
4.2.1 Pre-K	38
4.2.2 Toddler	39
4.3 Concurrent validity	41
4.4 Predictive validity	42
4.5 Limitations	44
4.5.1 Study sample and data collection	44
4.5.2 COVID-19	44
4.5.3 Methods/instruments	46
4.5.3.1 CLASS	46
4.5.3.2 ERC, BPM and CTNES	47
4.5.4 Analyses	48
5 Conclusion	49
References	50

1 Introduction

1.1 Brief description of thesis

Significant adults and caregivers play an important role in children's development (World Health Organization, 2004). Most children in Norway today attend kindergarten, and it is known that kindergarten staff and the kindergarten environment influences children's development (Peisner-Feinberg, 2007). There are several ways of assessing kindergarten quality, and the current thesis will focus specifically on process quality. Process quality concerns the relationship between teachers and children, amongst children, and between caregivers and kindergarten (Moa & Wold Olsen, 2014; Sommersel et al., 2013). It also includes how resources are applied and pedagogical competence is utilized. Process quality influences children more directly as it concerns children's interactions with the environment. Previous research has found process quality to be predictive of many positive outcomes in children (Brandlistuen et al., 2015; Løkken et al., 2018; Solheim, 2013; Wang et al., 2014). Having valid instruments to measure process quality could contribute to enhancing quality in Norwegian kindergartens and subsequently improve children's outcomes. One way to assess kindergarten quality, and process quality in particular, is to use an observational assessment tool called the Classroom Assessment Scoring System (CLASS). CLASS has been widely utilized, and research has found support for good psychometric properties in other countries (Bihler et al., 2018; Hamre et al., 2013; Hu et al., 2016; Leyva et al., 2015; Pakarinen et al., 2010; Stuck et al., 2016). The psychometric quality of CLASS in Norwegian kindergarten samples has not been investigated. This study will investigate the validity of CLASS using versions of the measure for two different age groups: Pre-K and Toddler. Firstly, the study will look into the factor structure of CLASS, and whether it replicates the original structure outlined by Pianta et al. (2008). Secondly, this study will focus on the construct validity of the CLASS instruments, by investigating if there is consistency between the observational measure and the kindergarten teachers' self-reports on how they manage children's negative emotions, and by that, the teachers' emotion socialization.

Developmental studies suggest that kindergarten quality can affect children's social and emotional abilities in both positive and negative ways (Peisner-Feinberg, 2007; Peisner-Feinberg et al., 2001). Therefore, this study will also look at the predictive validity of

CLASS, specifically regarding the relationship between the measure's constructs and children's emotion regulation, socialization and problem behaviour.

1.2 Basic emotions and emotional competence

1.2.1 Basic emotions

According to Izard (2010), emotions consist of neural circuits and response systems, and a feeling state or process that motivates and organizes cognition and action. Most researchers agree that there are six basic emotions; happiness, sadness, anger, fear, surprise and disgust. These emotions seem to be recognizable in all human cultures (Matsumoto et al., 1988; Yan et al., 2015), although there are some cultural variations (Elfenbein & Ambady, 2002). In addition to experiencing and feeling emotions, children must develop the ability to understand and regulate their emotions. In their early years, children need help understanding and managing their emotions, and guidance on how and when to express their feelings (Hernes & Larsen, 2014, p. 86).

1.2.2 Emotional competence

Emotional competence is a hypernym that includes emotion regulation, emotional expressiveness and emotion knowledge, and it is crucial to children's ability to interact with and understand others and oneself (Denham et al., 2003). Although separate concepts, these domains also intertwine, meaning for example that children who experience and express intense emotions, often have more difficulties regulating them as well (Denham, 2019, p. 508).

Having emotion knowledge means being able to understand the emotions of oneself and others, and in turn reacting appropriately to those emotions (Denham et al., 2003). This is important for prosocial interactions such as attaining and maintaining friendships. The first step in understanding emotions is recognizing them, both in oneself and others. By the age of 5-7 months, infants can discriminate between various expressions, including sad, happy and angry (Caron et al., 1988; Walker-Andrews & Lennon, 1991). From their first birthday, infants start using social referencing when faced with novel or ambiguous stimuli (Klinnert et al., 1986), by monitoring adults' affective displays to regulate themselves and to understand the surrounding environment. When they are able to recognize emotions in others, and have started developing their language skills, children can also start labeling emotions. By the age of three, most children can label some emotions (Denham, 1986), and this ability keeps on

increasing in the years following.

Emotional expressiveness is important for how a child is perceived by peers and teachers. Children who express more positive than negative emotions are often rated higher on friendliness and assertiveness from teachers, and lower on aggressiveness (Denham et al., 2003). The prosocial effects hold true for peers as well (Denham et al., 1990).

The process of emotion regulation begins at birth, and even infants have some forms of automatic emotion regulation: gaze aversion (Field, 1981) and self-soothing (Heron-Delaney et al., 2015). Other than this, they rely on their caregivers for co-regulation, where the caregiver provides comfort to help the child reduce his or her distress (Stifter & Augustine, 2019, p. 415). As children develop, so does their ability to regulate emotions. This happens gradually, and starts with the child using self-comforting or self-distracting strategies (Planalp & Braungart-Rieker, 2015; Stifter & Braungart, 1995). As the capacity for emotion regulation develops, so does the child's ability to conform to adults' expectations, for example by not lashing out when angry. They become increasingly able to use cognitive strategies to regulate their emotions, instead of behavioural strategies (Stifter & Augustine, 2019, p. 411). Development of emotion regulation is important for the development of social competence and general functioning. Children with good regulatory control usually perform better in school, due to their ability to pay attention and exhibit positive classroom conduct (Duckworth et al., 2011). On the other hand, children who struggle with emotion regulation have increased chances of developing externalizing problems (Doan et al., 2012).

As mentioned, emotional expressiveness, emotion regulation and emotion knowledge all fall under the term *emotional competence*. And while a considerable part of how a child expresses and regulates emotion is biologically determined by their temperament, emotional competence is also influenced by others. One of the important ways emotional competence is developed in children, is by emotion socialization from caregivers and other adults.

1.3 Emotion socialization

1.3.1 Emotion socialization from caregivers

Caregivers socialize their children by their modeling of emotional expressiveness, teaching about emotions and reacting to their children's emotions (Denham, 2019, p. 509). If caregivers are open to their children's emotions, encourage their reactions and have generally positive emotional expressions, this will enhance the emotional competence of the child.

Warm and supportive parenting has been shown to promote empathic responses in children,

and may serve as a model of sympathy towards others (Spinrad & Eisenberg, 2019, p. 352). Supportive reactions to children's emotions can help them develop knowledge and understanding of emotions (Denham & Kochanoff, 2002). In contrast, negative parental reactions can influence and increase externalizing emotional expressions in children (Newland & Crnic, 2011). When a parent fails to express emotions, this can also influence their child, as shown in the Tronick Still-Face Paradigm (Tronick et al., 1978). This study shows how quickly infants become distressed from their mothers' lack of emotional expressions, and how the distress increases with each still-face period.

Entering kindergarten and school with emotional competence skills makes it more likely the child will attain positive relationships with peers and teachers (Garner & Waajid, 2008), thus these are important skills to acquire early in life. Different aspects of emotion socialization are important at different stages in a child's life and development (Denham, 2019, p. 514). The support needed for a three year old might differ from a seven year old, and children in grade school might benefit from a more age appropriate type of support, to gradually learn to manage their own emotions.

Socialization of emotions is, like other socialization processes, influenced by culture. Therefore, many of the findings presented might apply mostly to Western or individualistic cultures. Studies with Japanese (Jin et al., 2017) and Indian (McCord & Raval, 2016) mothers find that some of the maternal responses deemed negative in Western studies (minimizing or not supporting the child's emotions) did not yield the same negative outcomes in Japanese and Indian children, perhaps because of the emphasis placed on harmony, community and modesty in these collectivistic cultures. Further, a comparison of Japanese and American children found that American children expressed more anger and aggressive behaviour, but also had mothers who highly encouraged their emotional expressions (Zahn-Waxler et al., 1996). This was in contrast to the Japanese mothers, who valued psychological discipline and were more likely to discourage aggressive and antisocial behaviour in their children.

Even though the present study concerns children and adults in Norway, it is important to keep in mind that the socialization practices that enhance and inhibit emotional competence in Western children, might not be the same as in other cultures.

1.3.2 Emotion socialization from kindergarten teachers

When starting kindergarten, children are suddenly exposed to many new people. They learn about social codes by interacting with other children and adults, and by observing the behaviour of others. Teachers are, in addition to caregivers, important socializers of

children's emotional development (Denham, 2019, p. 521). They socialize in many of the same ways as caregivers; by modeling, responding to and engaging with children's emotional expressions (Morris et al., 2013). This means they play a pivotal role in promoting positive development in children. A study by Curby and colleagues (2013) showed that consistent emotional support from kindergarten teachers was related to better social outcomes for the children. The quality of emotional support in kindergarten has also been found to moderate the risk of early school failure (Hamre & Pianta, 2005).

Teachers' abilities to help children regulate emotions is partially dependent on the teachers' own awareness of emotions. Ersay (2015) found that teachers with low awareness of their own emotions were less likely to self-report that they would help the children regulate and label their emotions. This could indicate that increasing teachers' own emotional competence can help them better perceive and teach children about their emotions.

1.4 Kindergartens in Norway

Kindergartens play a significant role as a socialization institution. Kindergartens in Norway are universally accessible, regulated, and subsidized (Barnehageloven, 2005, §16), which allows caregivers to return to work, and is a step towards a more equal labour market. Today, 92.2% of all Norwegian children aged 1-5 attend kindergarten (SSB, 2020d), which is a marked increase from 1999, when only 58% of children attended kindergarten. This could be related to the fact that the government decided on full kindergarten coverage in the early 2000s.

High quality kindergartens are related to numerous positive outcomes in children. Havnes and Mogstad (2009) studied the long-term effects of kindergartens on children's development and concluded that kindergartens have great positive effects on children's level of education and labour market affiliation in adult life. A study of 75 Norwegian children found that those who attended kindergarten had better grades when leaving middle school, and were more likely to have started higher education at the age of 20, compared to the children who stayed at home (Hartmann, 1992).

Early learning has a self-reinforcing effect (Heckman, 2006). This means that differences in childhood often become more prominent in adult life. Children who do well in school when they are young, often continue to do well, and those who perform under average rarely bridge the gap. This emphasizes the importance of high-quality kindergartens.

1.5 Assessment of kindergarten quality

Kindergarten quality is mainly divided into two different, but related, forms of quality: structural quality and process quality (Howes et al., 2008; Moa & Wold Olsen, 2014; Sommersel et al., 2013). The division of quality into these categories is somewhat arbitrary as they influence and affect each other indistinguishably.

Structural quality refers to materials, economic resources, teachers' educational levels, teacher-child ratio, working conditions etcetera. These variables affect children indirectly and are easy to quantify and conduct research on. Structural quality is a prerequisite for achieving high quality, but it is not the sole determinant. In Norway, the structural quality indicators are highly regulated by political guidelines. Internationally, Norwegian kindergartens are considered to be of high structural quality (UNICEF, 2008).

Process quality can be defined as the social, emotional, physical and instructional aspects of children's day-to-day experiences in classrooms, with a particular focus on children's interactions with teachers, peers, and materials and activities (Howes et al., 2008). It concerns the activities and learning opportunities children have in kindergarten, effective teaching, and the teacher-child relationship. The focus of this study will be on process quality.

1.6 Process quality in Norwegian kindergartens

1.6.1 How does process quality impact children?

The Norwegian Institute of Public Health reported that the greatest association with children's functioning was found when examining process quality (Wang et al., 2014). A good teacher-child relationship was associated with better language skills, less behavioural problems (both externalizing and internalizing) and better adaptation, also called school readiness. Good teacher-child relationships could explain 41% of the variance in school readiness (Wang et al., 2014). The causal direction was not clear, meaning that children with better functioning could also be better at forming good relationships with their teachers. Løkken et al. (2018) found an association between teacher-child interaction quality and empathy in children, and a marginal association between interaction quality and self-control. A report by Brandlistuen and colleagues (2015) for the Norwegian Institute of Public Health (FHI) found significant relationships between teacher-child relationships and behavioural outcomes in vulnerable children. The study used the self-report measure from teachers to measure the amount of closeness and conflict in relationships to the children. A lack of closeness was related to both externalizing and internalizing behaviour problems in all

children, but the effect was strongest for girls who were born premature, with low birth weight or with delayed motor or social development. High conflict was related to more externalizing behaviour problems for all children. In a doctoral thesis on the impact of kindergartens on children's outcomes, Solheim (2013) found that a positive teacher-child relationship was related to better social skills in first grade of primary school. This effect was only significant for children with low or medium relational risk defined as disorganized attachment.

Taken together, these findings show that process quality in Norwegian kindergartens has an important impact on children's development. A good teacher-child relationship seems to be especially important for vulnerable children, but when the difficulties are too severe, a good teacher-child relationship might not be enough to buffer these effects.

1.6.2 Do Norwegian kindergartens have high process quality?

Ree and Emilson (2020) used video recordings from Norwegian kindergartens to evaluate how teachers and children communicate. The results showed that teachers most prominently used closed questions and controlled what content to communicate with the children. This communication form results in children being passive receivers. More explorative communication forms using open ended questions were less frequently observed. This gives evidence of a low interaction quality, which is an important aspect of process quality. Another observational study also investigated the quality of interactions between teachers and toddlers in Norwegian kindergartens (Bjørnstad et al., 2020). This study found moderate to good basic interaction skills, including sensitive responsiveness, but found low quality for the more educational aspects of interaction. An assignment report for the County Governor (Lekhal & Vigmostad, 2014) found great variability in the relationship between teachers and children, as measured by a teacher-report. Most teachers reported evaluating their relationship with the children as good, but about 5% of the teachers responded that the teacher-child relationship was "not good". Moa and Wold Olsen (2014) found that many teachers had an inner motivation to work with children, which was related to higher process quality.

The results from this sparse collection of studies imply that Norwegian kindergartens have low to medium interaction quality between teachers and children. Teachers are motivated to work, which is related to better process quality. The overall process quality can be considered of moderate level.

1.6.3 Knowledge gap

From the literature it is evident that process quality is difficult to measure, and a variety of instruments are used to examine this aspect of kindergarten functioning. Many use observational methods, which allow observation of the interactions directly. Some standardized observational methods include the Caregiver Interaction Profile (CIP), Early Childhood Environment Rating Scale (ECERS) and Infant/Toddler Environment Rating Scale Revised (ITERS-R). Other methods use qualitative measures such as interviews with children or teachers, or analysis of video recorded sessions. Some use self-reports completed by teachers. Using a multitude of different measures can be beneficial for gaining a wide knowledge base, but it can also make it difficult to compare results across studies. This causes there to be no common operationalization, and studies might measure slightly different aspects of process quality. A prerequisite for using any measure is the need for good reliability and validity. The aforementioned studies, which investigated process quality in Norwegian kindergartens, varied by how well the psychometric properties were described. Some did not report on validity and reliability altogether, and some of the measures were not validated in a Norwegian context. Having a good instrument to measure process quality in Norwegian kindergartens with sound psychometric properties, is desirable. This could improve the generalizability of the results across studies, be used as a quality indicator for kindergartens, and guide efforts to intervene for improvement.

Pianta and colleagues developed an observational method that aims to measure process quality in kindergartens and schools, namely the Classroom Assessment Scoring System (CLASS), that also creates a method for interventions and feedback to kindergartens.

1.7 CLASS

CLASS focuses on the heart of great teaching – the teacher-child interaction (*Teachstone Home*, n.d.) – and is designed to improve this interaction. CLASS has been widely utilized worldwide, and especially in the United States, where it is part of the Head Start monitoring process and a component of the Quality Rating and Improvement System (QRIS).

CLASS was developed in the United States. It originally started out as an observational method to evaluate mother-infant interactions, and was called the Observation Record of Caregiving Environment (ORCE) (Buell et al., 2017). The researchers wanted to include other adult interactions, such as teachers, in their measure. From this, the Classroom Observation System (COS) was invented. ORCE and COS were both child-centered

observational methods. CLASS was developed as the interest in quality of early care and education grew, and the method became teacher-centered as opposed to the previously child-centered methods. Since then, CLASS has continued to evolve. CLASS has developed observational instruments to measure quality all through toddler and pre-school years to the K-12 and high school years. The underlying dimensions that CLASS consists of are almost the same for all grade-levels, but the way these dimensions are expressed in classrooms may change as children mature. Developing an instrument that is attuned to children's developmental state, but is based on the same underlying principles and theories, ensures a common language to discuss quality across grades. CLASS Pre-K (4-6 years) and Toddler (1-3 years) are developed to measure interaction quality in kindergartens.

1.7.1 CLASS Pre-K

CLASS Pre-K has three domains that include 10 different dimensions. The domains are Emotional Support, Classroom Organization and Instructional Support. The first domain, Emotional Support, consists of the dimensions Positive Climate, Negative Climate, Teacher Sensitivity and Regard for Student Perspective. This domain focuses on teachers' abilities to support social and emotional functioning in the classroom. Key elements are the emotional connection between teachers and children, interactions with an emphasis on children's interests, and the awareness and responsiveness teachers have for children's concerns. The second domain, Classroom Organization, includes the dimensions Behaviour Management, Productivity, and Instructional Learning Format. This domain focuses on classroom processes which organizes the children's time, behaviour and attention. Classrooms provide most opportunities for learning when children are well-behaved, interested and engaged, and consistently have things to do. A key aspect of this domain is *active* learning in children. The theoretical underpinnings of Classroom Organization is based on constructivist theories (Pianta et al., 2008). The final domain is Instructional Support, and it includes the dimensions of Concept Learning, Quality of Feedback and Language Modelling. Gaining usable knowledge (learning how facts are interconnected, organized, and conditioned on one another) and metacognitive skills are related to cognitive development. Instructional Support aims at improving these skills in children by promoting higher-order thinking, extending children's learning, and encouraging language.

1.7.2 CLASS Toddler

CLASS Toddler consists of two domains made up of eight dimensions. The two domains are Emotional and Behavioural Support and Engaged Support for Learning. Emotional and Behavioural Support includes the dimensions Positive Climate, Negative Climate, Teacher Sensitivity, Regard for Child Perspectives, and Behaviour Guidance. This domain is equivalent to the Emotional Support domain from Pre-K, but reflects the fact that teachers' social-emotional and behavioural interactions with toddlers are more interconnected than with older children. The second domain, Engaged Support for Learning, consists of the dimensions Facilitation of Learning and Development, Quality of Feedback, and Language Modelling. This is equivalent to the Classroom Organization and Instructional Support domains from Pre-K. The focus of this domain is teachers' support of cognition, engagement and language.

1.7.3 CLASS, process quality and emotion socialization

As previously mentioned, teachers are important role models in children's lives and facilitate learning and development. The emotional environment in kindergarten and the teacher-child relationship impacts children's social and emotional development (Curby et al., 2013; Morris et al., 2013). In kindergarten research, this is included in the concept of process quality. CLASS is an instrument designed to measure classroom processes and is especially developed to evaluate and improve the teacher-child relationship (La Paro et al., 2004). The domains Classroom Organization, Instructional Support, and Engaged Support for Learning are primarily concerned with children's cognitive and language development. The domains Emotional Support (Pre-K) and Emotional and Behavioural Support (Toddler) are predominantly concerned with the emotional environment in kindergartens. These aspects of kindergarten functioning may be related to other concepts such as emotion socialization, knowledge and expressiveness – overall referred to as emotion competence. Knowing that acquiring emotion competence is important for children's outcomes later in life, it might be important to examine the emotional and behavioural dimensions of CLASS further.

1.8 Validity of CLASS

CLASS Pre-K and Toddler has been examined in several research papers, and the findings generally conclude with satisfying reliability and validity. Support for the validity, reliability, and the use of CLASS has been found in the United States (Hamre et al., 2013), Finland

(Pakarinen et al., 2010), Germany (Bihler et al., 2018; Stuck et al., 2016), China (Hu et al., 2016), and Chile (Leyva et al., 2015). A study also found evidence of good validity and inter-rater reliability for CLASS used in Norwegian lower-secondary schools (CLASS-S) (Westergård et al., 2019). Although the studies yielded satisfying levels of validity and reliability, there has been mixed results when assessing the construct validity.

1.8.1 Construct validity

Construct validity is considered the most important type of validity and must be based on a theoretical framework. It evaluates to what extent an instrument measures the intended concept. Construct validity is considered superior to all validity aspects (Goodwin & Leech, 2003), and includes convergent validity and discriminant validity (Carmines & Zeller, 1979). Convergent validity measures to what extent the instrument is related to other constructs that it theoretically should be related to, and divergent validity measures to what extent the instrument is not related to measures it is not theoretically meant to be related to. A common way of evaluating construct validity is by performing factor analysis or structural equation modelling.

For CLASS Pre-K there has been uncertainty about whether the original three-factor model or a bifactor model fit the CLASS data best. Hamre et al. (2014) found that a bifactor model fit the data better and had better theoretical value. A bifactor model has one general factor which loads onto all the observable dimensions, in addition to grouped factors which load onto some of the same dimensions. The general factor and the grouped factors are uncorrelated, in contrast to a two-factor model. Using a bifactor model can therefore avoid the issue of multicollinearity when assessing the model's predictive value. However, the study found significant correlations between the three-factor model and children's outcomes, and emphasized the simplicity of the three-factor model (Hamre et al., 2014). In general, the three-factor model has satisfactory fit to most data, and is better than a one-factor, two-factor or bifactor model. Alternative models would add complexity to the scoring process. Other studies have also investigated the construct validity of the CLASS Pre-K. Hu et al. (2016) found support for the bifactor model in a Chinese kindergarten sample, but concluded that the three-factor structure had the best fit. Bihler et al. (2018) found support for the original three-factor structure in German kindergartens, but no support for a bifactor model. Another German study investigating CLASS Pre-K found that a two-factor model, where Emotional Support and Classroom Organization was combined to a single domain, had a marginally better fit compared to the three-factor structure (Stuck et al., 2016). Pakarinen et al. (2010)

found evidence for the three-factor model in Finnish kindergartens when Negative Climate was excluded. To make the model fit, they made the residuals of Productivity correlate with the residuals of Behaviour Management, and the residuals of Quality of Feedback correlate with the residuals of Concept Development. This study also found high correlations among the domains. Leyva et al. (2015) found support for the three-factor model in Chilean kindergartens when the residuals of several of the dimensions were allowed to correlate.

Few studies were found that investigated the construct validity of CLASS Toddler. A study from the Netherlands found an overall good fit for the original two-factor model for CLASS Toddler (Slot et al., 2016). However, the study concluded that a three-factor model fit the sample better. An American study found support for the two-factor structure which had a better fit compared to a single-factor model (Bichay-Awadalla & Bulotsky-Shearer, 2021).

Taken together the results indicate that, for Pre-K, the three-factor model had the best fit across many different cultures. However, most studies found that the model improved with small adjustments. The construct validity for Toddler was less researched, and the results were inconclusive.

1.8.2 Criterion-related validity

Criterion-related validity is used to evaluate to what extent the instrument can estimate important outcomes that are external to the instrument itself. It is often divided into two types: predictive validity and concurrent validity (Carmines & Zeller, 1979). Predictive validity concerns future outcomes, and concurrent validity concerns outcomes at the same time. Criterion-related validity is estimated based on the correlation between the instrument and the outcome measured, which should be theoretically connected.

When using CLASS as a measure of quality, it is important to investigate whether CLASS scores can predict child outcomes, and if high quality kindergartens impact children in meaningful ways. Perlman et al. (2016) conducted a meta-analysis of all available research up to July 2015, containing associations between CLASS scores and child outcomes for American preschool-aged children. This study found a small, but significant positive correlation between the Instructional Support domain and social skills, as measured by the Social Skills Rating System, and a small positive correlation between Classroom Organization and children's self-regulatory abilities, as measured by the Preschool Self-Regulation Assessment (PSRA)/Pencil Tapping. Hamre and colleagues (2014) was one of the studies included in the meta-analysis which found a positive correlation between Classroom Organization and PSRA. This study also found a positive correlation between

Instructional Support and children's gains in language and literacy, and found that Emotional Support was related to *decrease* in inhibitory control. Leyva et al. (2015) found a significant positive correlation between Instructional Support and children's early writing and executive functioning as measured by Pencil Tapping, but the effect size was small. La Paro et al. (2014) found that the Emotional and Behavioural Support domain from the CLASS Toddler was able to predict differences in child behaviour problems, with higher scores on the domain being related to fewer behaviour problems, however the effect size was small. Pakarinen et al. (2020) found that higher quality of the Emotional Support domain was related to more prosocial behaviour, but the results were marginal. The same study did not find a significant relationship between Emotional Support and antisocial behaviour. Hatfield et al. (2013) investigated 63 kindergartners and found support for greater decline in cortisol levels for children attending kindergartens with higher scores on the Emotional Support domain. This could indicate that a more emotionally supportive environment, as measured by CLASS Pre-K, could reduce stress levels in children. The results were not replicated for Classroom Organization or Instructional Support.

These results show that CLASS is related to numerous child outcomes, strengthening the validity of the measure. Emotional (and Behavioural) Support is related to lower cortisol levels, less behavioural problems, and more prosocial behaviour. Classroom Organization is related to inhibitory control and literacy gains, and Instructional Support is related to better social skills, language skills and executive functioning. The findings are in accordance with the theoretical framework of CLASS. However, the effect sizes were often small, and many studies only found significant results for one of the domains. Children experience rapid developmental change in the first five years of life. Considering that studies using CLASS find correlations to outcomes in children, it underlines the importance of high quality kindergartens for children's development.

Little research was found investigating the concurrent validity of CLASS Pre-K or Toddler. Pakarinen et al. (2010) found evidence of good concurrent validity in Finnish kindergartens. Teachers self-rated affection was positively related to Emotional Support and Classroom Organization, and teacher-reported efficacy was positively related to Emotional Support. Instructional support was not related to neither teacher-reported affection nor efficacy. Another study used a three-domain structure for CLASS Toddler and found a negative relationship between teacher-child ratio and Emotional Support and Engaged Support for Learning (Slot et al., 2016). La Paro et al. (2004) compared an old version of the CLASS Pre-K measure (with two domains and nine dimensions) with two other observational

measures: Early Childhood Rating Scales - Revised Edition (ECERS) and Snapshot. The results showed that CLASS was related to ratings from the ECERS and Snapshot in the expected direction and magnitude, giving evidence to good concurrent validity.

The scarce literature on concurrent validity suggests overall good validity. However, more research would be needed before concluding firmly.

Given the good psychometric properties of CLASS found in different countries, the measure could be a valuable asset to the research on process quality in Norwegian kindergartens. No previous studies were found that investigated the validity of CLASS in kindergartens in Norway.

1.9 The present study

The Norwegian Directorate for Education and Training states that children's well-being and development is the main goal for all quality work in the kindergarten sector (Udir, 2017). Research has shown that the teacher-child relationship, which is an important part of process quality, is important to ensure children's well-being and development, and that teachers are important role models and socializers of children's emotional competence.

Norwegian kindergartens have strict guidelines supervising structural quality indicators, but process quality has a more direct influence on children, and should be given attention when improving kindergarten quality. Previous research from Norway shows great variability in the instruments used to evaluate process quality, with mixed psychometric properties. This study aims to investigate the validity of the CLASS instrument which has been readily utilized around the world. More specifically, CLASS Pre-K and CLASS Toddler will be evaluated by investigating the construct validity and criterion-related validity (concurrent and predictive).

Construct validity will be investigated by conducting factor analyses to see whether the original structure is replicated in the Norwegian sample. Criterion-related validity will be investigated in two ways. The predictive validity will be evaluated by CLASS' ability to predict outcomes in children: emotion regulation and problem behaviour, as measured by the Emotion Regulation Checklist (ERC) and Brief Problem Monitor (BPM) respectively. Both outcome variables will be measures filled out by children's caregivers. Concurrent validity will be evaluated by looking at the relationship between observed process quality (CLASS) and teachers' self-reported ability to respond to children's negative emotions, using an adapted version of the Coping with Toddler's Negative Emotions Scale (CTNES).

If the factor structures are replicated, it would be preferable to only use the domains from CLASS that target emotional support and behaviour specifically; Emotional Support (ES) from Pre-K and Emotional and Behavioural Support (EBS) from Toddler. This is because the CTNES, ERC and BPM are all related to emotional and behavioural outcomes, and these are most closely measured in the ES and EBS domains.

Research question 1: construct validity

Is the factor structure of CLASS, as measured in Norwegian kindergartens, similar to the original factor structure? CLASS measures from two time points (T1 and T2) will be combined to evaluate the factor structure.

Hypothesis: The original CLASS factor structures will be replicated for both CLASS measures (three factors for Pre-K and two factors for Toddler).

Research question 2: concurrent validity

Is CLASS related to how teachers respond to negative emotions in children (supportive responses and unsupportive responses), as measured by the Coping with Toddler's Negative Emotions Scale (CTNES)? CLASS at T1 will be compared to CTNES scores at T1.

Hypothesis: We expect higher CLASS scores to be related to supportive responses from teachers. Lower CLASS scores are expected to be related to unsupportive responses.

Research question 3: predictive validity

Is CLASS able to predict children's emotion regulation and behavioural problems (externalizing and internalizing), as measured by the Emotion Regulation Checklist (ERC) and Brief Problem Monitor (BPM)? CLASS scores at T1 will be compared to ERC and BPM scores at T2.

Hypothesis: Higher CLASS scores are expected to predict higher ERC scores in children. Lower CLASS scores are expected to predict higher BPM scores. The results are expected to yield small effect sizes. This is because CLASS is an observational method, and ERC and BPM are reports filled out by caregivers on behalf of their children. CLASS and ERC/BPM are also measures from different contexts (kindergarten and home), and the children might not behave exactly the same in different environments. Observations or self-reports from the same people in the same context would be expected to yield larger effect sizes.

2 Method

2.1 Procedure

This study is a part of the Norwegian Tuning in to Kids for Kindergarten Teachers (N-TIK-KT) project. The project is a collaboration between the Department of Psychology at the University of Oslo, FUS kindergartens, and Kompetansetjeneste for Tidlig Innsats, with Professor Sophie Havighurst as the Principal Investigator. FUS is a private company that runs over 170 kindergartens in Norway.

The project aims to build teacher skills in responding to children's emotions and enhancing their emotional competence. Tuning in to Kids (TIK) teaches emotion coaching to the kindergarten teachers, a way of responding to emotions that assists children in understanding and regulating their emotions. Baseline measures were taken during August and September of 2019, and repeated in May and June of 2020 after delivering TIK to all intervention kindergartens. The control kindergartens received the TIK intervention in the autumn of 2020 and spring 2021. The current study used data from the control sample (who had not received the intervention) at two time points.

2.1.1 Ethical considerations

The project was granted approval from the Norwegian Center for Research Data (NSD). All participants in this study gave their written informed consent before participating, and had the opportunity to withdraw their consent at any time. All data handling was performed in Services for Sensitive Data (TSD). Access to raw files containing sensitive information was granted as it was necessary to calculate age of participants and connect their self-reports to corresponding CLASS observations. After merging and connecting the original files, new files were made where sensitive information was not included. No data material was extracted from TSD, and the analyses presented in this study include no sensitive information regarding any of the participants.

This study included observations of teachers and children in kindergartens. Especial concern should be applied when conducting research on young children. Caregivers gave consent on behalf of their child. The children were informed about the study and all observations were conducted openly with the observers visible in the kindergarten classrooms.

2.2 Participants

The participants were recruited from the 49 FUS kindergartens in Norway that were a part of the larger N-TIK-KT project. The current study used data from the 27 control group kindergartens, so any analysis would not be affected by the intervention.

2.2.1 Kindergarten teachers

The respondents included 221 kindergarten teachers from the 27 FUS kindergartens. To clarify, the term *teacher* is used for all responding employees, independent of their actual educational background. Most teachers were female (91.7%), ranging from 19 to 68 years ($M: 38.3, SD: 9.54$). The vast majority (90.9%) reported having Norwegian nationality, and 87.4% reported Norwegian to be their native tongue. Average education was minimum one year at college or university level (58.7%), and the length of employment was on average 6.5 years ($SD: 4.58$) in the current kindergarten, and 11.1 years ($SD: 7.14$) in kindergartens overall.

2.2.2 Caregivers and children

From 26 of the kindergartens there were 187 caregivers responding on behalf of their child. The 27th kindergarten in the sample had no responding caregivers. The children were on average 4.06 years ($SD: 1.34$) and 50.8% were boys. Most children (75.4%) had one or more siblings. Among the responding caregivers, 84.4% were female and 98.6% reported being the biological parent of the child. Most respondents (67.8%) were between 30 and 40 years of age ($M: 35.96, SD: 5.14$), and the vast majority (91.9%) reported being cohabitated or married. Average family income per caregiver was between 500-599 999 NOK, and the majority of caregivers (58.3%) had completed one year or more of college or university education. Demographic variables such as educational level, number of siblings and nationality were representative for the Norwegian population in general (SSB, 2020b, 2020a, 2020c).

Table 1

Participants' descriptives

	N	Minimum	Maximum	Mean	SD
Teacher					

Length of employment, current kindergarten (in years)	221	0	23	6.53	4.68
Length of employment, total in kindergartens (in years)	221	0	35	11.14	7.14
Educational level	221	1	6	3.91	1.54
Children/caregivers					
Child age (in months)	183	22.58	77.69	48.78	16.16
Caregiver age (in years)	187	21	52	35.96	5.14
Caregiver average income ^A	187	2	12	7.09	.69
Caregiver average educational level	187	1.5	6	4.54	1.28

Note: Educational level: categorical from 1 (primary school, 10 years or less) to 6 (5-6 years of college/university education). ^ACaregiver average income: categorical from 1 (no income) to 12 (≥ 1 million NOK annually). Presented as income per caregiver, averaged from total household income.

2.2.3 Exclusion criteria and imputation

Participants with incorrect or missing kindergarten units, or missing CLASS scores, were excluded. Two caregivers responded on behalf of the same child, so one of the caregivers was randomly excluded. Kindergarten teachers who did not consent to the study were excluded.

For caregivers entering an incorrect date of birth for their child (i.e. age at observation was a negative number), the date of birth was considered missing and EM imputed based on the kindergarten unit and date of birth variables. EM imputation can be used for several reasons, and one of the advantages of the method is that other variables can be used to give information about the missing data without being included in the actual model estimation (Heck et al., 2013, p. 23). Multiple imputation was also considered, but was not a possibility as SPSS was not able to conduct multilevel analyses on multiple imputed data. The descriptive information presented in table 1 is based on non-imputed data.

After exclusion, the sample consisted of 221 kindergarten teachers and 187 caregivers.

2.3 Measures

2.3.1 Classroom Assessment Scoring System

As previously mentioned, CLASS Pre-K and Toddler share many similarities, but are distinct measures. CLASS Pre-K consists of 10 dimensions and three domains, whereas CLASS Toddler consists of eight dimensions and two domains. Each dimension consists of behavioural markers, and these consist of quality indicators. When assessing the kindergarten

environment, the CLASS observers evaluate the presence and frequency of quality indicators to give each behavioural marker a score from 1 to 7. A score of 1-2 indicates low quality, 3-5 indicates medium quality, and 6-7 indicates high quality. When all behavioural markers have a score, the observer calculates the dimension score using the CLASS manual.

For this study, each kindergarten unit was observed three times, 15 to 20 minutes each time. Each observation aspired to observe the environment in different settings – e.g. free play, meals, transitions and organized play – to get a full impression. Each cycle had a score for every dimension. The average dimension scores across all observations for each kindergarten unit was calculated, and these averages were used to calculate the overall domain scores in accordance with the CLASS manual.

All observations were conducted by kindergarten teachers who had completed CLASS certification training. The teachers observed kindergartens in which they were not employed to minimize bias. To become a certified CLASS observer, they had to complete a two-day course and pass an observational test. An inter-rater reliability of at least 80% was required to pass the test. The reliability was calculated by comparing the teacher's scores to a master coding provided by Teachstone. In addition to gaining an overall inter-rater reliability of 80% or better, a good reliability of each dimension score was required, ensuring that no dimensions were consistently scored above or below the master coding.

2.3.2 Coping with Toddler's Negative Emotions Scale

The Coping with Toddler's Negative Emotions Scale (CTNES) is based on a similar measure, namely the Coping with Children's Negative Emotions Scale (CCNES). CCNES is a self-report instrument that reflects different ways caregivers respond to children's negative emotions (Fabes et al., 1990), and it has received support for both internal reliability and construct validity (Fabes et al., 2002). The CTNES was developed and published in 2004 (Spinrad et al., 2004) to use with toddlers. As with the CCNES, the CTNES consists of 12 hypothetical scenarios, and asks the caregivers to rate the likelihood that they would respond to the scenarios in different ways. CTNES has seven response styles and seven corresponding subscales; Problem-Focused Responses, Emotion-Focused Responses, Expressive Encouragement, Minimization Responses, Punitive Responses, Distress Responses and Granting the Child's Wish. Caregivers respond using a Likert scale ranging from 1 (very unlikely) to 7 (very likely).

When evaluating the CTNES, Spinrad and colleagues (2007) found that Granting the Child's Wish and Distress Reactions did not factor in with any of the other subscales, and

they suggested removing them, leaving the scale with five subscales in total. They then used principal components analysis to divide the remaining five subscales into what they called supportive strategies (Problem-Focused, Emotion-Focused and Expressive Encouragement) and unsupportive strategies (Punitive and Minimization responses).

In the version of the CTNES used in the current study, two subscales were added to the original (containing seven subscales) by the Principal Investigator of the N-TIK-KT study; Professor Sophie Havighurst. This was done to adjust and expand the scale to include other aspects of emotion socialization, adjust it to teachers and make it fit typical situations encountered in the kindergarten. The new scales were called Acknowledgement responses and Distract responses, and are thought to complement the original scales. Acknowledgement reflects whether the teachers acknowledge and accept the child’s feelings, and Distract reflects to what degree the teachers distracts the child when they experience emotions.

In the present study, a factor analysis was conducted that yielded somewhat the same results as Spinrad and colleagues (2007), with two factors emerging (table 2). The first factor included the supportive strategies (Problem-Focused, Emotion-Focused and Expressive Encouragement) as well as the new scale Acknowledgement. The second factor included the unsupportive strategies (Punitive and Minimization) as well as the Distress scale and the new Distract scale. Granting the Child’s Wish was excluded as our modified version of the scale only had four such responses. Based on this factor analysis, the remaining eight subscales were divided into supportive strategies (Emotion-Focused, Problem-Focused, Expressive Encouragement and Acknowledgement) and unsupportive strategies (Distress, Punitive, Minimization and Distract). The two subscales were reliability tested, with a resulting Cronbach’s alpha of .94 for the supportive scale and .91 for the unsupportive scale (table 9).

Table 2

Pattern matrix of CTNES

	Factor 1	Factor 2
Expressive Encouragement	.84	-
Emotion-Focused	.87	-
Problem-Focused	.85	-
Acknowledgement	.77	-
Distress	-	.59
Minimization	-	.84

Punitive	-	.84
Distract	-	.74

Note: Principal components analysis with Direct oblimin rotation.

The CTNES was considered a good choice for evaluating the concurrent validity of CLASS as they both measure some components of emotion socialization in teachers. The Emotional Support (Pre-K) and Emotional and Behavioural Support (Toddler) domains from CLASS uses observation to evaluate the emotional environment in kindergartens. CTNES focuses on how the teachers respond directly to children’s emotions using teachers self-report. A higher supportive CTNES score is thought to indicate positive emotion socialization, and could therefore correlate with a positive emotional environment indicated by CLASS.

2.3.3 Emotion Regulation Checklist

The Emotion Regulation Checklist (ERC) is an other-report instrument developed to evaluate emotion regulation in children (Shields & Cicchetti, 1998). It comprises two scales; Emotion Regulation (ER) and Liability/Negativity (L/N). The caregivers answer 24 statements about their children’s ability to regulate their emotions, on a scale from 1 (rarely/never) to 4 (almost always). ERC has been widely used to investigate different issues regarding emotion regulation in children, for example the possible link between emotion regulation and attachment (Borelli et al., 2010) and emotion regulation in children whose mothers express depressive symptomatology (Blandon et al., 2008). ERC has also been translated to various languages, and both Italian (Molina et al., 2014) and Portuguese/Brazilian (Reis et al., 2016) studies support the two-factor solution suggested in the original ERC. The studies also show acceptable structural and construct validity for the instrument. A Norwegian graduate thesis (Wathne Oseland, 2019) has examined the validity of the Norwegian version of ERC, and her findings did not support the two-factor structure. An explorative analysis failed to reveal alternative solutions with acceptable fit and interpretability, and the thesis concludes that the Norwegian version of ERC should be used with caution.

In the present study, Liability/Negativity scores were reversed and an average score for both scales was calculated. The whole ERC scale was also reliability tested, with a resulting Cronbach’s alpha of .85, generally thought to be a robust or good reliability (Taber, 2018).

2.3.4 Brief Problem Monitor

The Brief Problem Monitor (BPM) is an abbreviated version of the Achenbach System of Empirical Based Assessment (ASEBA), used to assess children's externalizing, internalizing and attentional problems (Backer-Grøndahl & Martinussen, 2018). The present study used the BPM 1 ½-5 years version. It was translated and back-translated by Akasie Språktjenester AS and approved by the original author (Achenbach). This translated 1 ½-5 years version has not previously been used or validated in Norway. The 6-18 years version has been both translated and investigated, however, and preliminary evidence suggests good internal consistency and good construct and content validity (Richter, 2015).

BPM 1 ½-5 years, consists of 18 questions divided into two subscales: externalizing and internalizing behaviour. The caregivers responded by either 0 (not true), 1 (somewhat true) or 2 (very true). An average score for the subscales combined was calculated for each child. A reliability test of the scale yielded a Cronbach's alpha of .84.

2.4 Statistical analysis

All statistical analyses were performed in IBM SPSS, version 27, and IBM SPSS AMOS 26. The significance level was set to 5%.

2.4.1 Factor structure of CLASS

Exploratory factor analysis (EFA) was run on CLASS Pre-K and Toddler data separately. Observations from baseline (T1) and follow-up (T2) for all control kindergartens were used to investigate the factor structure: 110 observations for Pre-K (57 from T1 and 53 from T2) and 93 observations for Toddler (46 from T1 and 47 from T2). The factor structure interpreted from the EFA and the original factor structure suggested by Pianta et al. (2008) were then investigated by performing confirmatory factor analysis (CFA). The results from the factor analyses were used to guide what variables to include when investigating the criterion-related validity.

2.4.2 Multilevel modelling

Intraclass correlations (ICC) are commonly used to measure the reliability of an experimental method, and to assess correlations within a class of data (Liljequist et al., 2019). The ICC scores give an estimate of the total variability between individuals that is attributable to within-group similarities (Field, 2013, chapter 20). In this study, ICC scores were calculated

to decide whether to do a regular regression analysis or a multilevel analysis. The finding of a significant cluster effect in either kindergartens or kindergarten units would imply the need for a multilevel model. A commonly used cut off for ICC is 5% (Heck et al., 2013, p. 90), which was also adopted in this study. To investigate intraclass correlations, null models were used, and ICC was calculated for each level separately.

ERC and BPM obtained low ICC scores: for ERC a Hessian error occurred (interpreted as close to zero) for both kindergarten and unit. For BPM an ICC score of 1.05% for kindergarten and Hessian error for unit was obtained. The ICC scores indicated no cluster effect for ERC or BPM. When investigating the predictive validity (research question 3), a mixed model without a second level was performed using ERC and BPM as dependent variables in two separate models. CLASS was the independent variable. Standard demographic variables were used as covariates (Salkind, 2010, p. 347), in this case average caregiver income, gender and age of the child. Average caregiver income was treated as a continuous variable in all analyses.

Unsupportive CTNES showed substantial group-effects. The ICC score for kindergarten was 10.5%, and for unit it was 15.5%. The effect of unit after removing the effect of kindergarten was 7.85%. For supportive CTNES, the ICC score for kindergarten was 4.94%, which was just below the 5% cut-off value. For kindergarten unit the ICC score was 13.29%. The ICC score for unit after the effect on kindergarten had been removed was 12.12%. Including the effect of unit after removing the effect of kindergarten gave no significant improvement to the model, as measured by difference in -2LL. It was therefore not included in the final model. To avoid performing the same analyses twice (for unit and kindergarten), and thereby increase the probability of a type I error, it was decided to perform the analyses adding only kindergarten as the second level. The choice was based on the fact that CLASS was measured at the unit level, and it was considered preferable to perform mixed models on the superior level, which was kindergartens. The final model investigated the effect CLASS had on CTNES scores when the effect of kindergarten was removed. Unsupportive and supportive CTNES were included as the dependent variables in separate analyses, with CLASS as the independent variable, and gender, education level, and length of employment in the current kindergarten were added as covariates. Education level was treated as a continuous variable in all analyses.

After all analyses were performed, the variables were converted to Z scores before entered into the model, and all analyses were rerun in order to compare estimates between models.

2.4.3 Parametric tests

After completing the mixed model analyses, the final model and the variables were investigated to see if the requirements for a linear model were met. Predictors and residuals were graphed to investigate the distribution using the PP-plot and the residual distribution.

For linear models there are two main sources of bias: outliers and violations of assumptions (Field, 2013, chapter 5). The assumptions include additivity and linearity, normality, homoscedasticity/homogeneity and independence.

To account for outliers all variables were checked for extreme values. This study adopted a definition of outlier as a value of ± 3 standard deviations (MIT Critical Data, 2016, p. 166). For variables where outliers were an issue, winsorizing was used to recode the values. An advantage of winsorization is that it preserves the fact that these cases had among the highest or lowest values in the distribution (Reifman & Garrett, 2010).

For multilevel linear models, the assumption of independence is accounted for by nesting groups together (in this case kindergartens). Multilevel linear models can also model variability in heterogeneity in regression slopes by using random slopes instead of fixed.

When investigating the distributions of the results, using histograms, and skewness and kurtosis, it was discovered that the assumption of normality was violated. The central limit theorem states that the parameter estimate of a population will have a normal distribution provided that the sample is big, regardless of how the shape of the sample is. There is no fixed cut-off for what is considered a big sample, but a widely accepted value is a sample size of 30. For heavily skewed distributions, a sample size of 100 or more is redeemed necessary (Field, 2013, chapter 5). Considering the sample size of this study was about 200 (221 and 187), the lack of normal distributions does not necessarily influence the results too much following the logic of the central limit theorem. However, the skewed distributions were tried modified by ranking values and by log-transforming the data. Using ranked cases or log-transforming did not result in a non-significant “Test of Normality” (Kolmogorov-Smirnov and Shapiro-Wilk) for any variables, except for unsupportive CTNES (significance of .200 and .053 for the tests respectively). A natural log (Ln) transformation was applied to unsupportive CTNES, but not the other variables.

The last assumption states that the outcome variable should be linearly related to the predictors. All analyses were run using the new variables with winsorized outliers, and with log-transformed unsupportive CTNES. Again, the parametric assumptions were investigated. Examining the scatterplot of the dependent variables against CLASS, a linear model showed

poor fit, explaining only a few per mille of the variance. Changing the model to cubic or square did not improve the results. The only exception was for unsupportive CTNES, which showed an overall satisfying fit to a linear model. The assumption of linearity is the most important assumption. Violations of this assumption suggests the model should be considered invalid, as it has not been described correctly.

Since there was no improvement in changing our model to a cubic or squared relationship, it was decided to continue using the multilevel model, even though the underlying assumptions for running this analysis were unfulfilled.

3 Results

3.1 Research questions

Our study sought to test three hypotheses. Firstly, whether we could replicate the original factor structure of CLASS. Secondly, whether CLASS has concurrent validity when compared to teachers' self-report on managing children's emotions (CTNES). The hypothesis predicted that CLASS would be positively related to supportive CTNES and negatively with unsupportive CTNES. And lastly, whether CLASS has predictive validity, in the form of predicting emotion regulation (ERC) and problem behaviour (BPM) in children. The hypothesis predicted that CLASS would be positively related to ERC and negatively with BPM.

3.2 First hypothesis: construct validity of CLASS

3.2.1 CLASS Pre-K

3.2.1.1 Exploratory factor analysis

Pre-K observations from T1 and T2 were combined to investigate the factor structure. The Kaiser-Meyer-Olkin (KMO) and Bartlett's test of Sphericity for Pre-K (.87, and χ^2 : 728.05 Sig. .000) confirmed the benefit of performing a factor analysis.

EFA using Maximum Likelihood and Direct Oblimin on Pre-K was performed. Examining the results using the scree plot and the Kaiser's criterion revealed two latent factors. This factor structure resulted in a significant Goodness-of-fit (χ^2 : 74.73, Sig. .000).

Table 3

Pattern matrix of CLASS Pre-K

	Factor 1	Factor 2
Positive Climate	.93	-
Negative Climate	.23	-
Teacher Support	.93	-
Regard for Student Perspective	.87	-
Behaviour Management	.75	-
Productivity	.66	-

Instructional Learning Format	.58	-
Concept Development	-	.74
Quality of Feedback	-	.91
Language Modelling	-	.73

Note: Exploratory factor analysis using Maximum Likelihood and Direct Oblimin. Factor structure interpreted from the scree plot and Kaiser’s criterion. Negative climate has been reversed.

Using the pattern matrix to identify the dimensions that fit to each factor reveals that the original Emotional Support and Classroom Organization were forced together as a single domain. All factor loadings were above .512, with a sample size of above 100, and were therefore considered significant (Field, 2013, p. 681), except for Negative Climate. Forcing the exploratory factor analysis to use three factors resulted in non-significant Goodness-of-fit test, but also resulted in the theoretical model of three domains. The factor loadings in this model were significant for all dimensions, except Negative Climate and Instructional Learning Format.

3.2.1.2 Confirmatory factor analysis

Confirmatory factor analysis was performed using the Maximum Likelihood on the original Pre-K model, and the model suggested by exploratory factor analysis (with two latent factors).

Table 5

Maximum likelihood estimates for CLASS Pre-K

	Two-factor model	Original model
CMIN	99.91, p < .000	76.12, p < .000
TLI	.88	.91
CFI	.91	.94
RMSEA (90% CI)	.13 (.10, .16)	.11 (.08, .15)
AIC	161.91	142.12
BIC	168.87	149.53

Note: Confirmatory factor analysis using Maximum Likelihood on the two-factor model resulting from the EFA, and the original three-factor model.

Abbreviations: CMIN - Chi-Square value, denoted by χ^2 . TLI - Tucker-Lewis Index. CFI - Comparative Fit Index. RMSEA - Root Mean Square Error of Approximation. AIC - Akaike Information criterion. BIC -

Bayesian Information Criterion.

The table shows that none of the models had good fit. The poor fit was concluded from the significant χ^2 , and the insignificant RMSEA test.

The original model had a better fit compared to the two-factor model; the lower bound for the RMSEA was below .1 which suggests good fit (Kline, 2015, chapter 13). The TLI and CFI for the original model was above .9 which is indicative of adequate, but not good fit (Bentler & Bonett, 1981; Hu & Bentler, 1999). When comparing the models, the AIC and BIC are used to evaluate which model is preferred. Smaller AIC and BIC values are indicative of a better model, regardless of number of parameters (Heck et al., 2013). Raftery (1995) stated that a BIC difference of 6-10 per parameter change is considered in strong favour of the model with lowest scores. Even accounting for the two-parameter difference in models (two-factor model: $df=34$, original model: $df=32$), the original three-factor model was strongly preferred. In conclusion, the original model had a better fit compared to the two-factor model, and was considered preferable for further analyses.

Table 6

Maximum Likelihood Estimate for the original theoretical model of CLASS Pre-K

Parameter	Unstandardized		Standardized	
	Estimate	SE	Estimate	Estimate ²
Emotional Support				
PC	1.00	-	.86	.74
NC	0.12	.05	.25	.06
TS	1.21	.09	.95	.89
RSP	1.04	.09	.83	.69
Classroom Organization				
BM	1.00	-	.84	.70
PD	1.01	.10	.83	.69
ILF	0.94	.10	.79	.63
Instructional Support				
CD	1.00	-	.77	.59
QF	1.17	.14	.83	.69

LM	1.06	.12	.84	.71
	Factor covariance		Factor correlation	
	Estimate	SE	Estimate	
ES <-> CO	.52	.09	.87	
IS <-> CO	.37	.08	.68	
IS <-> ES	.30	.07	.61	

Note: The standardized and unstandardized estimates of the variance in CLASS dimensions explained by the CLASS domains, the standardized estimates squared, and the covariance and correlation between all CLASS domains based on the CFA on the original three-factor structure of CLASS Pre-K. Negative climate has been reversed.

Abbreviations: PC - positive climate. NC - negative climate. TS - teacher support. RSP - regard for student perspective. BM - behaviour management. PD - productivity. ILF - instructional learning format. CD - concept development. QF - quality of feedback. LM - language modelling. ES - emotional support. CO - classroom organization. IS- instructional support.

Investigating table 6, looking at the standard regression weights squared, it was evident that the underlying factors explain more than half of the variance in every indicator, except for Negative Climate, which indicates an overall good fit of the model (Kline, 2015). Negative Climate stands out with only 6% of its variance stemming from its domain, indicating a very poor fit of this dimension. Calculating the average standard regression weights for each underlying factor reveals Average Variance Extracted (AVE) scores above .5 (ES: .59, CO: .67, and IS: .66) indicating good convergent validity on the construct level (Hair et al., 2017). Applying the Fornell-Larcker criterion for assessing discriminant validity between the constructs shows that the square root of AVE for Emotional Support (.77) is less than the correlation between ES and CO (.87). This indicates that the Emotional Support factor shares more variance with Classroom Organization than all its individual associated indicators. This could be explained by the poor fit of Negative Climate on Emotional Support.

The internal consistency for CLASS Pre-K was investigated using Cronbach's alpha. The factor analyses for Pre-K used the combination of observations from T1 and T2 which together obtained an $\alpha = .91$. For research question 2 and 3 only the Pre-K observations from T1 were used, which obtained an $\alpha = .92$.

3.2.2 CLASS Toddler

3.2.2.1 Exploratory factor analysis

Toddler observations from T1 and T2 were combined to investigate the factor structure. KMO (.77) and Bartlett's test of Sphericity (χ^2 : 385.64, Sig. .000) on Toddler confirmed the benefit of running a factor analysis. Factor analysis was run for Toddler (93 unique

observations) using Maximum Likelihood and Direct Oblimin. The scree plot and Kaiser's criterion revealed two latent factors with a significant Goodness-of-fit (χ^2 : 40.06, Sig. .000).

Table 4

Pattern matrix for CLASS Toddler

	Factor 1	Factor 2
Positive Climate	.93	-
Negative Climate	.41	-
Teacher Support	.92	-
Regard for Child Perspectives	.60	-
Behaviour Guidance	.64	-
Facilitation of Learning and Development	-	-1.01
Quality of Feedback	-	-.59
Language Modelling	-	-.52

Note: Exploratory factor analysis using Maximum Likelihood and Direct Oblimin. Factor structure interpreted from the scree plot and Kaiser's criterion. Negative climate has been reversed.

The pattern matrix revealed significant factor loadings for all dimensions, except for Negative Climate. The factor structure from the exploratory factor analysis was in accordance with the original theoretical model.

3.2.2.2 Confirmatory factor analysis

For CLASS Toddler, a confirmatory factor analysis was performed using only the original two-factor model considering that the structure was replicated well in the exploratory factor analysis.

Table 7

Maximum Likelihood Estimates for CLASS Toddler

	Original model
CMIN	63.42, p < .000
TLI	.82
CFI	.88

RMSEA (90% CI)	.16 (.18, .20)
AIC	113.42
BIC	118.84

Note: Confirmatory factor analysis using Maximum Likelihood on the original two-factor model.
Abbreviations: CMIN - Chi-Square value, denoted by χ^2 . TLI - Tucker-Lewis Index. CFI - Comparative Fit Index. RMSEA - Root Mean Square Error of Approximation. AIC - Akaike Information criterion. BIC - Bayesian Information Criterion.

The original model shows poor fit indicated by the significant χ^2 , the insignificant RMSEA and the lower bound exceeding .1, and CFI and TLI values not exceeding .9.

Table 8

Maximum Likelihood Estimate for the original model of CLASS Toddler

Parameter	Unstandardized		Standardized	
	Estimate	SE	Estimate	Estimate ²
Emotional and Behavioural Support				
PC	1.00	-	.79	.62
NC	0.17	.04	.46	.21
TS	1.13	.12	.92	.84
RCP	1.07	.13	.82	.67
BG	0.99	.13	.73	.53
Engaged Support for Learning				
FLD	1.00	-	.81	.66
QF	0.84	.14	.72	.51
LM	1.02	.17	.70	.49
	Factor covariance		Factor correlation	
	Estimate	SE	Estimate	
EBS <-> ESL	.31	.08	.61	

Note: The standardized and unstandardized estimates of the variance in CLASS dimensions explained by the CLASS domains, the standardized estimates squared, and the covariance and correlation between all CLASS domains based on the original two-factor structure of CLASS Toddler. Negative climate has been reversed.
Abbreviations: PC - positive climate. NC - negative climate. TS - teacher support. RCP - regard for children perspectives. BG - behaviour guidance. FLD - facilitation of learning and development. QF - quality of feedback. LM - language modelling. EBS - emotional and behavioural support. ESL- engaged support for learning.

The table shows that the latent factors explain more than 50% of the variance for all indicators except Negative Climate and Language Modelling. More variance in Negative Climate is explained by the domain in Toddler than in Pre-K. However, the Emotional and Behavioural Support domain explains less proportion of the variance in Negative Climate compared to all other domains and dimensions, indicating poor fit of Negative Climate. The AVE values for the two factors exceeded .5 and gave support for good convergent validity (EBS: .58, ESL: .55). Investigating the discriminant validity, it shows that the square root of the AVE values for both factors (.76 and .74 respectively) exceed the correlation between the factors (.61). This indicates that, on average, the factors share more variance with its individual indicators than each other, giving support to good discriminant validity.

The internal consistency for CLASS Toddler was investigated using Cronbach's alpha. The factor analyses for Toddler used the combination of observations from T1 and T2 which together obtained an $\alpha = .86$. For research question 2 and 3, only the Toddler observations from T1 were used, which obtained an $\alpha = .88$.

3.3 Preliminary preparations

For the criterion-related validity analyses, it was preferable to combine CLASS Pre-K and Toddler to a single measure to improve statistical power. The results from the factor analyses supported a three-factor structure for Pre-K and two-factor structure for Toddler. Considering the results were in accordance with the original factor structure, the theoretical background for each domain was transferrable. The theoretical background states that the Emotional Support (ES) domain from Pre-K and the Emotional and Behavioural Support (EBS) domain from Toddler are primarily concerned with the emotional environment in kindergartens. To combine the Pre-K and Toddler (from T1) into one variable, the ES and EBS domains were merged together to represent the emotional aspect of CLASS. The new variable was named CLASS ES/EBS and was subsequently used as the independent variable for all criterion-related validity analyses.

The project also contained self- and other-reports from teachers (CTNES) and caregivers (ERC, BPM) regarding emotional and behavioural constructs, constructs that are thought to overlap with the domains of CLASS ES/EBS. The ES/EBS domains were therefore considered favorable to compare with CTNES, ERC and BPM when investigating the criterion-related validity of CLASS.

3.4 Second hypothesis: concurrent validity of CLASS

CTNES and CLASS at T1 were used to investigate the concurrent validity. Multilevel models were performed using the CTNES (supportive and unsupportive) as the dependent variables, CLASS ES/EBS as the independent variable, and gender, education level, and length of employment in the current kindergarten acting as covariates.

Table 9

Descriptives – teachers

	N	Min.	Max.	Mean	SD	Skewness	Kurtosis	Cronbach's alpha
Education level	221	1	6	3.91	1.54	-.55	-1.30	-
Length of employment	221	0	23	6.53	4.68	.51	-.12	-
Supportive CTNES	221	4.75	7	6.31	.53	-1.09	.56	.94
Unsupportive CTNES	221	1	4.43	1.93	.59	1.16	2.29	.91
CLASS ES/EBS	221	3.13	6.73	5.41	.80	-1.04	.65	-

Note: Descriptive information about teachers, CTNES and CLASS ES/EBS, based on raw data before winsorization or log transformation. Length of employment in years. Education level is categorical from 1 (primary school of 10 years or less) to 6 (5-6 years of university/college education), but treated as a continuous variable in all analyses.

Table 9 shows that teachers reported being highly inclined to use supportive response styles when handling children's negative emotions (M : 6.31). Unsupportive response styles were infrequently used (M : 1.93). The distribution of unsupportive CTNES scores were considered normally distributed after applying Ln transformation. The CLASS ES/EBS mean indicated medium to high quality in the observed kindergartens (M : 5.41).

Table 10

Correlation matrix – teachers

	Gender	Educational level	Length of employment	CLASS ES/EBS	Unsupportive CTNES	Supportive CTNES
Gender	-					
Education level	-.00	-				
Length of employment	-.13	-.27**	-			
CLASS ES/EBS	.05	-.04	-.02	-		
Unsupportive CTNES	.02	-.24**	.15*	.02	-	
Supportive CTNES	-.21**	.03	.18**	.05	-.18**	-

**Correlation is significant at the .01 level (2-tailed).

*Correlation is significant at the .05 level (2-tailed).

Note: Gender: 1 = female, 2 = male. Length of employment in years. Education level is categorical from 1 (primary school of 10 years or less) to 6 (5-6 years of university/college education).

Table 10 shows that CLASS ES/EBS was not correlated with any of the CTNES measures. Several of the covariates and CTNES scores had significant correlations with one another. Linear regression analyses were conducted with supportive and unsupportive CTNES as the dependent variables in two separate analyses, to check for issues of multicollinearity. All other variables (covariates and CLASS) were included in the analyses as independent variables. The Variance Inflation Factor (VIF) for all variables were well below the cut-off score for VIF of 4 (Hair et al., 2017). The tolerance for all numbers showed that at least 88% of the variance was unique to each variable. These results indicate that there was no issue with multicollinearity.

When performing the multilevel analysis, the model was built from a simple model to a more complex model using the Δ -2LL significance as a reference for model improvement (Field, 2013, chapter 20). Using unsupportive CTNES as the dependent variable, including random intercept for kindergarten (Δ -2LL 16.88, sig<.01) improved the model significantly. Including random slope did not improve the model. For supportive CTNES the -2LL change for including kindergarten was just below the p<.05 value of 3.84 (Δ -2LL 3.83). To keep consistency in methodology it was decided to include kindergarten level in the model. Random slopes did not improve the model.

The final multilevel models were performed using the Restricted Maximum Likelihood. Two analyses were performed: Unsupportive CTNES and supportive CTNES were both analyzed with the effect of kindergarten removed. CLASS ES/EBS and the covariates were fixed effects and the intercept was random for all analyses.

Table 11

Multilevel analysis on CLASS – teachers

	Unstandardized regression coefficient			Sig.	95% Confidence Interval ^B	
	Est. Z-score ^A	Est. raw score	SE		Lower Bound	Upper Bound
Supportive CTNES	.02	.01	.05	.84	-.08	.10

Unsupportive CTNES	.16	.06	.03	.03*	.01	.11
--------------------	-----	-----	-----	------	-----	-----

* Significant at the .05 level (2-tailed).

Note: Results are the estimates of fixed effects from multilevel analyses with supportive CTNES and unsupportive CTNES as dependent variables, CLASS ES/EBS as independent variable, and gender, education level, and length of employment in current kindergarten as covariates. Multilevel analyses were performed using the Restricted Maximum Likelihood. Kindergarten was included as a second level as random intercept. ^A Unstandardized regression coefficient where all variables were converted to Z-scores, and thus comparable to standardized regression coefficients. ^B CI based on raw score.

The table shows that CLASS ES/EBS was significantly and positively related to unsupportive CTNES after removing the effect of kindergarten. CLASS ES/EBS was not significantly related to supportive CTNES.

3.4 Third hypothesis: predictive validity of CLASS

Predictive validity was investigated using CLASS at T1 and ERC and BPM at T2. Mixed models on ERC and BPM as dependent variables in two separate analyses with CLASS ES/EBS as the independent variable, and caregiver average income, gender and age of the child as covariates, were performed.

Table 12

Descriptives – children and caregivers

	N	Min.	Max.	Mean	SD	Skewness	Kurtosis	Cronbach's alpha
Age of child (in months)	183	22.58	77.69	48.78	16.16	.13	-1.18	-
Caregiver average income	187	2	12	7.09	.69	.25	.88	-
ERC	187	1.98	3.96	3.37	.34	-.79	1.10	.85
BPM	187	0	1.39	.22	.23	1.85	4.94	.84
CLASS ES/EBS	187	3.13	6.67	5.29	.79	-.78	.017	-

Note: Descriptive information about children, based on raw data before winsorization and EM imputation. Caregiver average income is categorical from 1 (no annual income) to 12 (≥ 1 million NOK annually), but treated as a continuous variable in all analyses.

Table 12 shows that the average ERC score was high and close to the maximum score (M : 3.37). BPM had an average score close to zero (M : .22). The BPM scores were considered non-normally distributed based on both the skewness and kurtosis exceeding 1 (Hair et al., 2017). Ranking or transforming the data did not result in normally distributed data. CLASS ES/EBS had an average score of 5.29, indicating medium to high quality in the observed kindergartens.

Table 13

Correlation matrix – children and caregivers

	Gender of child	Age of child	Caregiver average income	CLASS ES/EBS	ERC	BPM
Gender of child	-					
Age of child	-.04	-				
Caregiver average income	.03	.11	-			
CLASS ES/EBS	-.02	-.20**	.10	-		
ERC	-.12	.07	.09	-.04	-	
BPM	.01	-.02	-.19**	.04	-.76**	-

**Correlation is significant at the 0.01 level (2-tailed).

Note: Gender: 1 = female, 2 = male. Caregiver average income is categorical from 1 (no annual income) to 12 (≥ 1 million NOK annually).

Table 13 shows that CLASS ES/EBS did not correlate with either ERC or BPM. CLASS ES/EBS had a significant negative relationship with children's age. BPM was correlated negatively with caregiver average income and ERC. The VIF for all variables were investigated to look for collinearity with ERC and BPM as dependent variables in two separate analyses, and covariates and CLASS ES/EBS acting as independent variables. The VIF was well below the cut-off value of 4 for all variables, and all variables had a tolerance of at least 93.3%. From these results it was concluded that there was no issue with multicollinearity.

When performing the mixed model on ERC and BPM, the Restricted Maximum Likelihood was used. Adding a second level revealed insignificant improvement of the model (based on the ICC and Δ -2LL), thus only one level was included in the model. All variables were fixed.

Table 14

Mixed model analysis on CLASS – children

	Unstandardized regression coefficient		SE	Sig.	95% Confidence Interval ^B	
	Est. Z-score ^A	Est. raw score			Lower Bound	Upper Bound
BPM	.05	.01	.02	.48	-.03	.05
ERC	-.04	-.02	.03	.62	-.08	.05

Note: Results are the estimates of fixed effects from performing mixed model analyses with BPM and ERC as dependent variables, CLASS ES/EBS as independent variable, and gender of the child, age of the child, and caregiver average income acting as covariates. Mixed models were performed using Restricted Maximum

Likelihood. ^A Unstandardized regression coefficient where all variables were converted to Z-scores, and thus comparable to standardized regression coefficients. ^B CI based on raw score. A significant correlation was found between BPM and caregiver average income in the mixed model analysis. Considering this was not part of the research question no further exploration of this relationship was performed.

The table shows that CLASS ES/EBS did not predict either ERC or BPM results.

3.5 Sensitivity analysis

As a sensitivity analysis, it was investigated whether using raw data would yield different results than using the winsorized variables and the natural log transformed unsupportive CTNES. This was not the case, and it was concluded that the transformed data were highly representative for the actual raw data.

4 Discussion

4.1 Purpose of the study

The intent of this study was to investigate the validity of the Classroom Assessment Scoring System (CLASS) in a Norwegian kindergarten sample. The structural validity of CLASS Pre-K and CLASS Toddler was investigated separately. The results found adequate fit for the original three-factor model for CLASS Pre-K, and mixed results for the original two-factor model for CLASS Toddler. Negative Climate showed poor fit for both Pre-K and Toddler. To investigate the concurrent and predictive validity of CLASS, the domains ES and EBS from Pre-K and Toddler were combined. The Coping with Toddler's Negative Emotions Scale (CTNES) was used to investigate the concurrent validity. The results showed a significant positive relationship between unsupportive CTNES and CLASS ES/EBS when the effect of kindergarten was removed. No significant relationship was found between supportive CTNES and CLASS ES/EBS. The Emotion Regulation Checklist (ERC) and Brief Problem Monitor (BPM) were used to investigate the predictive validity. The results showed that CLASS ES/EBS did not predict either ERC or BPM.

4.2 Construct validity

4.2.1 Pre-K

The results indicated an overall acceptable model fit for CLASS Pre-K. The exploratory factor analysis supported a two-factor model where Emotional Support and Classroom Organization were joined together as a single domain. The structural validity was investigated further using confirmatory factor analysis. This revealed an overall acceptable fit for the original three-factor model, and a poor fit for the two-factor model that resulted from the exploratory factor analysis. The AIC and BIC showed strong preference for the original three-factor model.

When exploring the original three-factor structure, the analyses showed good convergent validity. The underlying domains explained over half of the variance in all dimensions, except for Negative Climate. When investigating the discriminant validity, it became evident that the domains Emotional Support and Classroom Organization shared, on average, more variance with each other compared to the dimensions they consist of. This is consistent with the finding from the exploratory factor analysis joining these two domains

together as one. These findings are consistent with previous research showing high correlation between Emotional Support and Classroom Organization, but still concluding with the three-factor model as the best model (Hamre et al., 2013; Hu et al., 2016; Stuck et al., 2016). Hamre et al. (2013) also pointed out the benefits of using the three-factor model: it is easy to use, and the results can be compared to other studies and generalized.

The dimension Negative Climate showed poor fit to the model. This is consistent with previous research from many different countries finding poor fit and little variance for Negative Climate (Bichay-Awadalla & Bulotsky-Shearer, 2021; Leyva et al., 2015; Pakarinen et al., 2010; Slot et al., 2016; Stuck et al., 2016). The Negative Climate dimension measures the frequency of negative affect, punitive control, sarcasm and disrespect, punitive actions, and severe negativity such as bullying and physical punishment. It is the only dimension containing negative connotations, and it is reversed in the scoring process. One explanation for the inadequate fit could be that there were generally few observations of actions related to Negative Climate. There could possibly be a floor effect, and thus not enough variance for it to be a useful measure. Another explanation could be scoring bias from the kindergarten teachers. Most teachers acknowledge that acting with negative affect or in a punitive or disrespectful manner is less ideal, but that it happens in pressed situations (Skoglund & Åmot, 2020). This study used teachers as CLASS observers and it might be that they were more inclined to emphasize the positive and supportive actions observed.

The CLASS Pre-K original model shows an overall acceptable structural validity, with good convergent validity and an acceptable discriminant validity. Future research should investigate potential changes to improve the model fit. Excluding Negative Climate, as suggested by Pakarinen et al. (2010), could be a good first step.

4.2.2 Toddler

Performing the exploratory factor analysis on CLASS Toddler revealed the same factor structure as the original two-factor model with the Emotional and Behavioural Support (EBS) and Engaged Support for Learning (ESL) domains. This supported the two-factor structure, which was further investigated by performing a confirmatory factor analysis. The CFA showed an overall inadequate fit for the two-factor model. The domains explained over half of the variance for most of the dimensions, but not for Negative Climate or Language Modelling. The convergent validity was considered acceptable as, on average, the underlying domain explained over half of the variance in all its dimensions. The discriminant validity showed that the domains, on average, shared more variance with the dimensions they consist

of than with each other, indicating good discriminant validity.

Few previous studies investigating the validity of CLASS Toddler have been found. A study from the Netherlands found an overall good fit for the original two-factor model (Slot et al., 2016). However, the study concluded that a three-factor model, including Emotional Support (indicated by Positive Climate, Teacher Sensitivity, and Regard for Child Perspectives), Behavioural Support (indicated by Negative Climate and Behaviour Guidance), and Engaged Support for Learning (indicated by Facilitation of Learning and Development, Quality of Feedback, and Language Modeling), fit the sample better. Another study found support for the original two-factor structure in kindergartens with ethnically and linguistically diverse children in the USA (Bichay-Awadalla & Bulotsky-Shearer, 2021). However, this study did not investigate a three-factor structure. A three-factor model was not investigated in the present study either, but future research should examine if the three-factor model would be replicable in Norway, or if other models fit CLASS Toddler better.

Previous studies also found Negative Climate to be a problematic dimension, which is in accordance with the findings from this study (Bichay-Awadalla & Bulotsky-Shearer, 2021; Slot et al., 2016). The poor fit of Negative Climate has already been discussed above, but finding the same problematic fit in Toddler gives evidence to the generally inadequate fit of this dimension. The Language Modelling had just below 50% (48.7%) of its variance from the Engaged Support for Learning domain, but was not considered a major deviation from the other dimensions. Generally, the dimensions from Toddler had less variance explained by the domains compared to Pre-K. This makes the overall model fit poorer. With few studies having investigated the validity of CLASS Toddler, and concluding mixed results, more research should be prioritized.

Generally, the validity and reliability of instruments measuring younger children are recognized as less acceptable than those used with older children (La Paro et al., 2004). This is because younger children's competencies are more situationally dependent and unstable. This could help explain the lower validity of CLASS Toddler in the current study, but it also applies to CLASS Pre-K.

This study used CLASS measures from T1 and T2 when investigating the factor structure, and the effect of this is not clear. A previous research paper found significant temporal validity and moderate correlations between two observations for CLASS Pre-K (Stuck et al., 2016). The temporal validity of CLASS was not investigated in this study, but could have increased the random error.

4.3 Concurrent validity

The results from investigating the concurrent validity of CLASS were unexpected.

Table 10 showed no correlations between CLASS ES/EBS and CTNES. Multilevel analyses did not result in any significant relationship between CLASS ES/EBS and supportive CTNES either. However, from the multilevel analyses a significant positive relationship between CLASS ES/EBS and unsupportive CTNES was found. Kindergartens that had a higher score on CLASS ES/EBS had teachers who were more inclined to use unsupportive response styles in meeting with children's negative emotions.

The results were unexpected because unsupportive response styles are generally considered less optimal strategies when socializing children. The children may experience negative arousal which in turn can inhibit their ability to regulate their own behaviour and develop effective coping strategies (Spinrad et al., 2007). For unsupportive CTNES, examples of response styles when a child is upset at kindergarten drop-off, are “stop crying, or no one will want to play with you” (punitive), “it is nothing to get upset about” (minimization), as well as the teacher getting upset or angry themselves (distress) or distracting the child (distract).

Very few studies were found that have investigated the concurrent validity of CLASS Pre-K or Toddler, but the results from this study does not seem to be in accordance with previous results. Pakarinen et al. (2010) found evidence of good concurrent validity in Finnish kindergartens. The Emotional Support domain from Pre-K was positively related to teachers self-reported affections in the classroom and efficacy beliefs. La Paro et al. (2004) compared an old version of the CLASS Pre-K measure (with two domains and nine dimensions) with two other observational measures: Early Childhood Rating Scales – Revised Edition (ECERS) and Snapshot. The results showed that CLASS was related to ratings from the ECERS and Snapshot in the expected direction and magnitude giving evidence to good concurrent validity. Slot et al. (2016) used a three-domain structure for Toddler and found an expected negative relationship between teacher-child ratio and Emotional Support.

There could be many reasons why the results from this study were unexpected. Firstly, supportive and unsupportive CTNES had average scores that were close to the maximum/minimum score (as shown in table 9). This could imply a ceiling/floor effect, and there was little variation in the measure, making it difficult to detect any meaningful differences. Secondly, there could be methodological explanations for the results. Carmines

and Zeller (1979) states that criterion-related validity is largely dependent on the measurement of the criterion as it is on the quality of the measuring instrument itself. This could imply that CTNES is the measure with inadequate validity, not CLASS. The version of CTNES used in this study was modified and no comprehensive evaluation of the validation of this instrument has been studied. Thirdly, few research papers were found that investigated the concurrent validity. This could be a result of publication bias in which negative results were not published. Lastly, the positive relationship between CLASS ES/EBS and unsupportive CTNES could have a more psychological explanation. Previous research has found that teachers reported feeling like a personal and professional failure after acting in an unsupportive manner (Skoglund & Åmot, 2020). The psychological discomfort from occasionally acting in an unsupportive manner could result in compensation strategies later on, ultimately resulting in a more positive environment. It could be that teachers who are able to express their feelings and frustrations for a short period of time are more likely to remain calm and supportive the rest of the time.

From the 27 kindergartens and 82 kindergarten units where a CLASS observation had been completed, 221 teachers responded to the CTNES form. Although there was a considerable cluster sample size, there was sparse data when considering replies for each cluster. Many kindergartens only had a few responders. Previous research has shown that sparse data (few cases per cluster) could overestimate the between-cluster variance and hence result in a biased ICC score, overestimating the importance of clustering (McNeish, 2014). The same study also found that multilevel analyses could provide unbiased estimates with sparse data, even with as little as two observations per cluster, given a considerable number of clusters. The present study represents very sparse data, and the results should be evaluated with care. Optimally, each CLASS observation would have had several CTNES responders to minimize bias.

Taken together it can be concluded that this study did not find support for good concurrent validity for the CLASS ES/EBS. Little relationship was found between CLASS ES/EBS and CTNES, and several limitations exist. The results should be evaluated with care.

4.4 Predictive validity

Both bivariate correlations and regression analyses using ERC and BPM as dependent variables, CLASS ES/EBS as independent variables, and age, gender and caregiver average income as covariates, resulted in non-significant results.

The results are somewhat consistent with previous research. In a study investigating 63 kindergartners, support was found for greater decline in cortisol level for children attending kindergarten with higher scores on the Emotional Support domain (Hatfield et al., 2013). This could indicate that a more emotionally supportive environment, as measured by CLASS Pre-K, could reduce stress levels in children. Pakarinen et al. (2020) found that higher quality of the Emotional Support domain was marginally related to more prosocial behaviour. On the other hand, the meta-analysis conducted by Perlman et al. (2016) found no significant relationships between Emotional Support and children's outcomes. The study by Hamre et al. (2014) actually found that Emotional Support was related to *decrease* in children's inhibitory control. La Paro et al. (2014) found that the Emotional and Behavioural Support domain from the CLASS Toddler was able to predict differences in child behaviour problems with higher scores on the domain being related to fewer behavioural problems. The study had teachers fill out the Brief Toddler Social Emotional Assessment (BITSEA) for all children. The findings from this study would suggest a negative relationship between CLASS ES/EBS and BPM, which was not found. A comparison between the study from La Paro et al. (2014) and this study points out some differences which could cause the contrasting results. Firstly, BITSEA was filled out by the teachers, as opposed to ERC and BPM, which were filled out by caregivers. Having teachers report on the children's behaviours ensures greater personal and contextual similarities with CLASS. Secondly, as shown in table 12, the BPM score had a positive skew, with a low mean and little variation. This could cause a floor effect, causing the mistaken conclusion that the variable has no effect when it actually has (Cramer & Howitt, 2004, p. 21). The BITSEA Behaviour Problem subscale, on the other hand, had a greater range of possible answers, with a much greater variance in scores. It could be that BPM would have greater variance in a clinical sample. Thirdly, La Paro et al. (2014) used only CLASS Toddler as independent variable, while this study used a combination of Toddler and Pre-K. The findings from La Paro et al. (2014) also found that older children displayed fewer behaviour problems. If this holds true for this study as well, including older children would likely cause lower variation in BPM scores. Lastly for BPM, a combination of internalizing and externalizing behaviour problems was used. Separating them could provide more accurate results. This also holds true for ERC, which combined the Emotion Regulation and Liability/Negativity subscales. Table 12 shows that ERC had a negative skew, with high mean and little variation, possibly causing a ceiling effect.

This study did not find support for good predictive validity of the CLASS ES/EBS. The criterion-related validity was investigated by using only one of the domains from Pre-K

and Toddler. It should be noted that research on the other domains, or single dimensions, could yield different results. This would be interesting for future studies.

4.5 Limitations

4.5.1 Study sample and data collection

A possible limitation could be that FUS kindergartens differed significantly from the overall population or from public kindergartens. As mentioned in the method chapter, the caregiver/child sample appeared representable of the Norwegian population in terms of income, number of siblings and ethnicity. The FUS kindergartens, that are all run privately, also follow the same framework plan as public Norwegian kindergartens, and do not rely on a specific philosophy or pedagogy that makes them significantly different from public kindergartens. FUS' service declaration states that children should evolve through play, feel safe and explorative and take part in a playful community with other children (FUS-barnehagene, 2021), and these statements are also found in the first chapter of the Kindergarten act (Barnehageloven, 2005, §1-2). The study sample should therefore not differ significantly from the overall population or from public kindergartens.

Data at baseline and follow-up were collected at different times of the year. While baseline measures were taken in the autumn of 2019, follow-up was conducted in the spring of 2020. Both time points had circumstances that may have influenced the reliability of CLASS, thus increased random error and decreased the possibility of finding relations. In autumn, this could be the case if observations were done before the children had customized to their unit and routines had been stabilized. In spring, COVID-19 may similarly have influenced observations by increasing random error and decreasing reliability.

Another issue is sample size. The present sample size is smaller than that of similar studies (Bihler et al., 2018; Hamre et al., 2013; Hu et al., 2016; Leyva et al., 2015; Pakarinen et al., 2010), which can make it difficult to generalize to other samples and populations in general, as well as decreasing statistical power (Jones et al., 2003).

4.5.2 COVID-19

The changes in kindergarten caused by the pandemic could have increased the random error for this study. Between the baseline and follow-up measures being administered, the COVID-19 pandemic turned families, living situations and working situations upside down. For several weeks, most children were not allowed to attend kindergarten, except those with

caregivers employed in critical society functions, and children with special needs. According to The Norwegian Directorate for Education and Training, only 9% of children attended kindergarten during the lockdown (Udir, 2020). Another consequence of the pandemic was the increased use of temporary personnel, and 45% of all kindergartens reported hiring temporary personnel in the first week of lockdown because of teachers with respiratory symptoms. Also, almost all kindergartens reported keeping children in smaller groups, usually 3-6 children, during the first six weeks of lockdown (Udir, 2020). All these factors may have impacted the children's and kindergartens' functioning between baseline and follow-up measures. Unfamiliar adults (temporary personnel) could possibly lead to the children feeling less secure when attending kindergarten, and the smaller groups usually meant that each child only had one adult for support. The changes in kindergartens during lockdown could also have negatively affected the quality of the CLASS observations at follow-up, in addition to lowering the motivation to respond to the different self-reports, as many people may have felt preoccupied with handling the pandemic.

A report from the Norwegian Institute of Public Health has collected findings from several international studies on the effects of COVID-19 on children and adolescents (Nøkleby et al., 2021). A couple of American studies found that there had been less reports on children's abuse and fewer contacts with child welfare, although this effect was most marked in the first few months of the pandemic, when lockdown reduced contact between children and authorities (Baron et al., 2020; Rapoport et al., 2020). The authors also suggested that closing schools inhibited reporting of child maltreatment, as the primary source of reporting such maltreatment is school personnel (Baron et al., 2020). This could very well be applicable to kindergartens as well. Most children spent more time at home during the first months of the pandemic, for better or for worse.

Some parents of children with special needs voiced concerns about the long-term effects of the pandemic on their children's development (Neece et al., 2020), fearing stagnation or regression. Some studies specifically investigated toddlers and young children, and found increases in behavioural problems and more difficulties with emotion regulation in young children (Di Giorgio et al., 2020). If applicable to this study, this could have increased variability in ERC and BPM scores, but the results do not indicate such an effect.

4.5.3 Methods/instruments

4.5.3.1 CLASS

Observational methods are sensitive in the way that they can lead to change in behaviour of the people being observed (Ostrov & Hart, 2014, p. 297). Knowing that someone was watching and “evaluating” their work was a possible stress factor for the teachers concerned, and they were aware of the observation being conducted. Moreover, the CLASS observers were kindergarten teachers from other FUS kindergartens in the N-TIK-KT project, who had received specific training to become certified CLASS observers. Being trained in assessing and scoring CLASS, these teachers had more knowledge of what they themselves were being evaluated on. This could possibly have decreased the reliability of the CLASS scores compared to if the observers had all been external.

Given the relatively small sample in this study, it was preferable to combine CLASS Pre-K and Toddler to increase statistical power. As the outcome measures were all related to emotion and behaviour, it was considered beneficial to use the CLASS domains most related to emotional and behavioural kindergarten qualities as well. As noted above, previous research has found relationships between CLASS and children’s outcomes. However, no studies were found that combined one or more of the CLASS grade-levels. The effect of combining Emotional Support (ES) and Emotional and Behavioural Support (EBS) is not clear. However, it makes theoretical sense as they both measure – more or less – the same underlying phenomenon. The biggest difference between ES and EBS is that EBS integrates the emotional and behavioural aspects more closely, as they are thought to be more related to each other in toddlers compared to older children. EBS includes the dimensions Behaviour Guidance, while Behaviour Management (in Pre-K) belongs to the Classroom Organization domain. Previous research has found strong correlations between Behaviour Management and the ES domain (Sandilos & DiPerna, 2014). This could also partially help explain the high correlation between ES and Classroom Organization, as there seem to be a strong link between the emotional and behavioural aspects. It also makes sense theoretically. This could impact the present study in two meaningful ways. Firstly, when the behavioural aspects have been insufficiently accounted for, it could be more difficult to find relationships with child outcomes, especially BPM, which specifically measures children’s behaviour. Secondly, CLASS Toddler might be better at predicting children’s emotional and behavioural outcomes as it is more integrated in EBS. However, when EBS is combined with ES, the effect may diminish.

CLASS is an observational method developed to measure the same qualities from toddlers to high-school years. It would be interesting to see research combining different CLASS grade-levels to evaluate the usefulness of CLASS in a longitudinal perspective. If CLASS is useful for all ages, it should be possible to generalize the findings from one grade-level to another.

4.5.3.2 ERC, BPM and CTNES

First and foremost, the instruments are on different levels, and they measure different individuals. Whilst CLASS looks at kindergarten teachers in a joint context, CTNES reflect each teacher's individual self-report. ERC and BPM are thought to measure each individual child, but through their caregivers' reports. Finding a correlation between observations and various self-reports might be optimistic (Phares et al., 1989; Wilkins et al., 2018). It would still, in future research, be interesting to look at the concurrent and predictive validity of CLASS, although maybe with other instruments than ERC, BPM and CTNES at hand. Children's behaviour could for example be measured directly instead of being caregiver-reported.

Another possible issue is the high correlation between ERC and BPM, confirmed in the table 13. This raised the concern that BPM and ERC might measure some of the same underlying mechanisms in children. However, the analysis yielded no significant correlations between CLASS and ERC/BPM, and the conclusions were therefore not confounded by having two related outcomes.

The use of CTNES is also questionable. The original edition of CTNES was presented at a conference in the US (Spinrad et al., 2004) and was not, in the present study, available for use in its original form. Its basis and inspiration stems from the more available and accessible CCNES (Fabes et al., 1990), but as mentioned earlier, they differ slightly in form and content. One of the differences between the scales is the addition of the subscale Granting the Child's Wish in the toddler version (CTNES). Although a factor analysis (Spinrad et al., 2007) indicates that it does not fit well with the other scales, it is still a part of the original scale. In the present study, two additional subscales were also added to compliment the rest; the previously mentioned Acknowledgement and Distract scales. Using a modified scale that has not been validated could reduce generalizability.

4.5.4 Analyses

Mixed models are conceptually similar to multiple regression, and the dependent variable is linearly predicted from the independent variables or covariates (McNeish & Stapleton, 2016). The prerequisites for performing linear analyses (mixed models) were not fulfilled, except for unsupportive CTNES. In particular, the violation of the linearity assumption invalidates the remaining results, as the linear model trying to describe the relationship does not fit.

Investigating if the relationship between variables could better be explained by a cubed or squared relationship did not result in a better fit. The lack of any significant relationships between CLASS ES/EBS and ERC, BPM and supportive CTNES might be because the premises for running linear analyses were not fulfilled. The relation may also go in the other direction; that the lack of significant relationships between CLASS ES/EBS and ERC, BPM and supportive CTNES was the reason for the absence of linear relation.

5 Conclusion

The results from this study were mixed. The results showed adequate fit for the construct validity of CLASS Pre-K, and mixed results for CLASS Toddler. No evidence of good concurrent or predictive validity was found. The study had several methodological limitations, and only one domain from Pre-K and Toddler was used to evaluate the criterion-related validity. The final evaluation should therefore weigh the findings from the assessment of construct validity most. More research on the validity of CLASS Pre-K and Toddler in Norway is needed.

There were several limitations to this study, so the results should be evaluated with care. Important improvements for future research could be to use other domains or dimensions to evaluate criterion-related validity and increase the sample size. Child outcomes could be measured by people more closely linked to the kindergarten context: by the children themselves if possible, or by kindergarten teachers.

The Norwegian Directorate for Education and Training states that the main goal for all quality work in kindergartens is children's well-being and development (Udir, 2017), and Norway has strict governmental guidelines regarding the structural quality. There are, however, no regulations regarding process quality, most likely due to the fact that process quality is difficult to measure validly. Research on process quality in Norway shows that several different instruments are used when assessing process quality, with varied psychometric support. The need for a valid instrument to measure process quality still stands. This would enhance research on quality in Norwegian kindergartens, and could help guide future areas of improvement in quality work.

References

- Backer-Grøndahl, A., & Martinussen, M. (2018). *Måleegenskapene til den norske versjonen av Brief Problem Monitor (BPM)* [Electronic]. <https://doi.org/10.21337/0057>
- Barnehageloven. (2005). *Lov om barnehager (LOV-2005-06-17-64)*. Lovdata. https://lovdata.no/dokument/NL/lov/2005-06-17-64#KAPITTEL_1
- Baron, E. J., Goldstein, E. G., & Wallace, C. T. (2020). Suffering in silence: How COVID-19 school closures inhibit the reporting of child maltreatment. *Journal of Public Economics*, *190*, 104258. <https://doi.org/10.1016/j.jpubeco.2020.104258>
- Bentler, P. M., & Bonett, D. G. (1981). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*(3), 588. <https://doi.org/10.1037/0033-2909.88.3.588>
- Bichay-Awadalla, K., & Bulotsky-Shearer, R. J. (2021). Examining the Factor Structure of the Classroom Assessment Scoring System Toddler (CLASS-T) in Early Head Start and Subsidized Child Care Classrooms. *Early Education and Development*, *0*(0), 1–17. <https://doi.org/10.1080/10409289.2021.1887700>
- Bihler, L.-M., Agache, A., Kohl, K., Willard, J. A., & Leyendecker, B. (2018). Factor analysis of the Classroom Assessment Scoring System replicates the three domain structure and reveals no support for the bifactor model in German preschools. *Frontiers in Psychology*, *9*, 1232. <https://doi.org/10.3389/fpsyg.2018.01232>
- Bjørnestad, E., Broekhuizen, M. L., Os, E., & Baustad, A. G. (2020). Interaction quality in Norwegian ECEC for toddlers measured with the Caregiver Interaction Profile (CIP) scales. *Scandinavian Journal of Educational Research*, *64*(6), 901–920. <https://doi.org/10.1080/00313831.2019.1639813>
- Blandon, A. Y., Calkins, S. D., Keane, S. P., & O'Brien, M. (2008). Individual differences in trajectories of emotion regulation processes: The effects of maternal depressive symptomatology and children's physiological regulation. *Developmental Psychology*, *44*(4),

1110. <https://doi.org/10.1037/0012-1649.44.4.1110>

Borelli, J. L., Crowley, M. J., David, D. H., Sbarra, D. A., Anderson, G. M., & Mayes, L. C. (2010).

Attachment and emotion in school-aged children. *Emotion, 10*(4), 475.

<https://doi.org/10.1037/a0018490>

Brandlistuen, R. E., Helland, S. S., Evensen, L. A., Schjølberg, S., Tambs, K., Aase, H., & Wang, M.

V. (2015). *Sårbare barn i barnehagen—Betydningen av kvalitet* (Rapport 2015/2). Nasjonalt

Folkehelseinstitutt. <http://hdl.handle.net/11250/278990>

Buell, M., Han, M., & Vukelich, C. (2017). Factors affecting variance in Classroom Assessment

Scoring System scores: Season, context, and classroom composition. *Early Child*

Development and Care, 187(11), 1635–1648. <https://doi.org/10.1080/03004430.2016.1178245>

Carmines, E., & Zeller, R. (1979). *Reliability and Validity Assessment*. SAGE Publications, Inc.

<https://doi.org/10.4135/9781412985642>

Caron, A. J., Caron, R. F., & MacLean, D. J. (1988). Infant discrimination of naturalistic emotional

expressions: The role of face and voice. *Child Development, 59*(3), 604–616.

Cramer, D., & Howitt, D. L. (2004). *The SAGE dictionary of statistics: A practical resource for*

students in the social sciences. SAGE.

Curby, T. W., Brock, L. L., & Hamre, B. K. (2013). Teachers' emotional support consistency predicts

children's achievement gains and social skills. *Early Education and Development, 24*(3),

292–309. <https://doi.org/10.1080/10409289.2012.665760>

Denham, S. A. (1986). Social cognition, prosocial behavior, and emotion in preschoolers: Contextual

validation. *Child Development, 57*(1), 194–201. JSTOR. <https://doi.org/10.2307/1130651>

Denham, S. A. (2019). Emotional competence during childhood and adolescence. In V. LoBue, K.

Pérez-Edgar, & K. A. Buss (Eds.), *Handbook of Emotional Development* (pp. 493–541).

Springer International Publishing. https://doi.org/10.1007/978-3-030-17332-6_20

Denham, S. A., Blair, K. A., DeMulder, E., Levitas, J., Sawyer, K., Auerbach–Major, S., & Queenan,

P. (2003). Preschool emotional competence: Pathway to social competence? *Child*

Development, 74(1), 238–256. <https://doi.org/10.1111/1467-8624.00533>

- Denham, S. A., & Kochanoff, A. T. (2002). Parental contributions to preschoolers' understanding of emotion. *Marriage & Family Review, 34*(3–4), 311–343.
https://doi.org/10.1300/J002v34n03_06
- Denham, S. A., McKinley, M., Couchoud, E. A., & Holt, R. (1990). Emotional and behavioral predictors of preschool peer ratings. *Child Development, 61*(4), 1145–1152.
<https://doi.org/10.2307/1130882>
- Di Giorgio, E., Riso, D., Mioni, G., & Cellini, N. (2020). *The interplay between mothers' and children behavioral and psychological factors during COVID-19: An Italian study*.
<https://doi.org/10.31234/osf.io/dqk7h>
- Doan, S. N., Fuller-Rowell, T. E., & Evans, G. W. (2012). Cumulative risk and adolescent's internalizing and externalizing problems: The mediating roles of maternal responsiveness and self-regulation. *Developmental Psychology, 48*(6), 1529. <https://doi.org/10.1037/a0027815>
- Duckworth, A. L., Quinn, P. D., & Tsukayama, E. (2011). What No Child Left Behind leaves behind: The roles of IQ and self-control in predicting standardized achievement test scores and report card grades. *Journal of Educational Psychology, 104*(2), 439.
<https://doi.org/10.1037/a0026280>
- Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin, 128*(2), 203.
<https://doi.org/10.1037/0033-2909.128.2.203>
- Ersay, E. (2015). Preschool teachers' emotional awareness levels and their responses to children's negative emotions. *Procedia - Social and Behavioral Sciences, 191*, 1833–1837.
<https://doi.org/10.1016/j.sbspro.2015.04.220>
- Fabes, R., Eisenberg, N., & Bernzweig, J. (1990). *The Coping with Children's Negative Emotions Scale: Procedures and scoring*. Available from authors. Arizona State University.
- Fabes, R., 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 and Family Review, 34*(3–4), 285–310.

https://doi.org/10.1300/J002v34n03_05

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. SAGE.

Field, T. M. (1981). Infant gaze aversion and heart rate during face-to-face interactions. *Infant Behavior & Development*, 4(3), 307–315.

[https://doi-org.ezproxy.uio.no/10.1016/S0163-6383\(81\)80032-X](https://doi-org.ezproxy.uio.no/10.1016/S0163-6383(81)80032-X)

FUS-barnehagene. (2021). *Serviceerklæring—Dette kan barn og foresatte forvente i*

FUS-barnehagene. <https://fus.no/files/185/Serviceerkl%C3%A6ring-januar-2021.pdf>

Garner, P. W., & Waajid, B. (2008). The associations of emotion knowledge and teacher–child relationships to preschool children’s school-related developmental competence. *Journal of Applied Developmental Psychology*, 29(2), 89–100.

<https://doi.org/10.1016/j.appdev.2007.12.001>

Goodwin, L. D., & Leech, N. L. (2003). The meaning of validity in the new standards for educational and psychological testing: Implications for measurement courses. *Measurement and Evaluation in Counseling and Development*, 36(3), 181–191.

<https://doi.org/10.1080/07481756.2003.11909741>

Hair, J., Hult, G. T., Ringle, C., & Sarstedt, M. (2017). A primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). *Sage, Second edition*, 390.

Hamre, B. K., Hatfield, B., Pianta, R. C., & Jamil, F. (2014). Evidence for general and domain-specific elements of teacher-child interactions: Associations with preschool children’s development. *Child Development*, 85(3), 1257–1274. <https://doi.org/10.1111/cdev.12184>

Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Development*, 76(5), 949–967. <https://doi.org/10.1111/j.1467-8624.2005.00889.x>

Hamre, B. K., Pianta, R. C., Downer, J. T., DeCoster, J., Mashburn, A. J., Jones, S. M., Brown, J. L., Cappella, E., Atkins, M., Rivers, S. E., Brackett, M. A., & Hamagami, A. (2013). Teaching through interactions: Testing a developmental framework of teacher effectiveness in over 4,000 classrooms. *The Elementary School Journal*, 113(4), 461–487.

<https://doi.org/10.1086/669616>

- Hartmann, E. (1992). Mors og barnehagens betydning for barns skolegang, utdanning og utvikling av selvstendighet og autonomi. *Tidsskrift for Norsk Psykologforening*, 29, 716–730.
- Hatfield, B. E., Hestenes, L. L., Kintner-Duffy, V. L., & O'Brien, M. (2013). Classroom Emotional Support predicts differences in preschool children's cortisol and alpha-amylase levels. *Early Childhood Research Quarterly*, 28(2), 347–356. <https://doi.org/10.1016/j.ecresq.2012.08.001>
- Havnes, T., & Mogstad, M. (2009). *Er barnehage bra for barna?* <http://hdl.handle.net/11250/178423>
- Heck, R. H., Thomas, S. L., & Tabata, L. N. (2013). *Multilevel and longitudinal modeling with IBM SPSS*. Routledge.
- Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, 312(5782), 1900–1902. <https://doi.org/10.1126/science.1128898>
- Hernes, M., & Larsen, K. (2014). *Tidlig utvikling: Barns utvikling 0-3 år: hva kan man forvente, og hva bør gis ekstra oppmerksomhet?* (p. 167). Gyldendal akademisk.
- Heron-Delaney, M., Kenardy, J. A., Brown, E. A., Jardine, C., Bogossian, F., Neuman, L., de Dassel, T., & Pritchard, M. (2015). *Early maternal reflective functioning and infant emotional regulation in a preterm infant sample at 6 months corrected age*. 9.
- Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Ready to learn? Children's pre-academic achievement in pre-kindergarten programs. *Early Childhood Research Quarterly*, 23(1), 27–50. <https://doi.org/10.1016/j.ecresq.2007.05.002>
- Hu, B. Y., Fan, X., Gu, C., & Yang, N. (2016). Applicability of the Classroom Assessment Scoring System in Chinese preschools based on psychometric evidence. *Early Education and Development*, 27(5), 714–734. <https://doi.org/10.1080/10409289.2016.1113069>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Izard, C. E. (2010). The many meanings/aspects of emotion: Definitions, functions, activation, and regulation. *Emotion Review*, 2(4), 363–370. <https://doi.org/10.1177/1754073910374661>

- Jin, Z., Zhang, X., & Han, Z. R. (2017). Parental emotion socialization and child psychological adjustment among Chinese urban families: Mediation through child emotion regulation and moderation through dyadic collaboration. *Frontiers in Psychology, 8*.
<https://doi.org/10.3389/fpsyg.2017.02198>
- Jones, S. R., Carley, S., & Harrison, M. (2003). *An introduction to power and sample size estimation*.
<https://emj-bmj-com.ezproxy.uio.no/content/20/5/453>
- Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). Guilford Publications. <http://ebookcentral.proquest.com/lib/oslo/detail.action?docID=4000663>
- Klennert, M., Emde, R., Butterfield, P., & Campos, J. (1986). Social referencing. The infant's use of emotional signals from a friendly adult with mother present. *Developmental Psychology, 22*, 427–432. <https://doi.org/10.1037/0012-1649.22.4.427>
- La Paro, K. M., Pianta, R. C., & Stuhlman, M. (2004). The Classroom Assessment Scoring System: Findings from the pre-kindergarten year. *The Elementary School Journal, 104*(5), 409–426. JSTOR.
- La Paro, K. M., Williamson, A. C., & Hatfield, B. (2014). Assessing quality in toddler classrooms using the CLASS toddler and the ITERS-R. *Early Education and Development, 25*(6), 875–893. <https://doi.org/10.1080/10409289.2014.883586>
- Lekhal, R., & Vigmostad, I. (2014). *Trivsel og kvalitet i barnehagen* (Oppdragsrapport nr. 6-2014; p. 63). Høgskolen i Hedmark, Senter for praksisrettet utdanningsforskning.
https://brage.inn.no/inn-xmlui/bitstream/handle/11250/194503/opprapp06_14.pdf?sequence=1
- Leyva, D., Weiland, C., Barata, M., Yoshikawa, H., Snow, C., Treviño, E., & Rolla, A. (2015). Teacher–Child Interactions in Chile and Their Associations With Prekindergarten Outcomes. *Child Development, 86*(3), 781–799. <https://doi.org/10.1111/cdev.12342>
- Liljequist, D., Elfving, B., & Roaldsen, K. S. (2019). Intraclass correlation – a discussion and demonstration of basic features. *PLOS ONE, 14*(7), e0219854.
<https://doi.org/10.1371/journal.pone.0219854>
- Løkken, I. M., Broekhuizen, M. L., Barnes, J., Moser, T., & Bjørnstad, E. (2018). Interaction quality

and children's social-emotional competence in Norwegian ECEC. *Journal of Early Childhood Education Research*.

<https://jecer.org/wp-content/uploads/2018/12/Lokken-Broekhuizen-Barnes-Moser-Bjornestad-issue7-2.pdf>

- Matsumoto, D., Kudoh, T., Scherer, K., & Wallbott, H. (1988). Antecedents of and reactions to emotions in the United States and Japan. *Journal of Cross-Cultural Psychology*, *19*(3), 267–286. <https://doi.org/10.1177/0022022188193001>
- McCord, B. L., & Raval, V. V. (2016). Asian Indian immigrant and white American maternal emotion socialization and child socio-emotional functioning. *Journal of Child and Family Studies*, *25*(2), 464–474. <https://doi.org/10.1007/s10826-015-0227-2>
- McNeish, D. M. (2014). Modeling sparsely clustered data: Design-based, model-based, and single-level methods. *Psychological Methods*, *19*(4), 552. <https://doi.org/10.1037/met0000024>
- McNeish, D. M., & Stapleton, L. M. (2016). The effect of small sample size on two-level model estimates: A review and illustration. *Educational Psychology Review*, *28*(2), 295–314. <https://doi.org/10.1007/s10648-014-9287-x>
- MIT Critical Data. (2016). *Secondary analysis of electronic health records*. Springer Nature.
- Moa, V., & Wold Olsen, A. G. (2014). *På hvilken måte påvirker ledelse prosesskvalitet i barnehagen?* Norges teknisk-naturvitenskapelige universitet, NTNU og Dronning Mauds Minne, Høgskole for barnehagelærerutdanning. https://ntnuopen.ntnu.no/ntnu-xmlui/bitstream/handle/11250/270158/762119_FULLTEXT01.pdf?sequence=1&isAllowed=y
- Molina, P., Sala, M. N., Zappulla, C., Bonfigliuoli, C., Cavioni, V., Zanetti, M. A., Baiocco, R., Laghi, F., Pallini, S., De Stasio, S., Raccanello, D., & Cicchetti, D. (2014). The Emotion Regulation Checklist – Italian translation. Validation of parent and teacher versions. *European Journal of Developmental Psychology*, *11*(5), 624–634. <https://doi.org/10.1080/17405629.2014.898581>
- Morris, C. A. S., Denham, S. A., Bassett, H. H., & Curby, T. W. (2013). Relations among teachers' emotion socialization beliefs and practices and preschoolers' emotional competence. *Early*

- Education and Development*, 24(7), 979–999. <https://doi.org/10.1080/10409289.2013.825186>
- Neece, C., McIntyre, L. L., & Fenning, R. (2020). Examining the impact of COVID-19 in ethnically diverse families with young children with intellectual and developmental disabilities. *Journal of Intellectual Disability Research*, 64(10), 739–749. <https://doi.org/10.1111/jir.12769>
- Newland, R. P., & Crnic, K. A. (2011). Mother–child affect and emotion socialization processes across the late preschool period: Predictions of emerging behaviour problems. *Infant and Child Development*, 20(6), 371–388. <https://doi.org/10.1002/icd.729>
- Nøkleby, H., Berg, R. C., Muller, A. E., & Ames, H. M. R. (2021). *Konsekvenser av covid-19 på barn og unges liv og helse*. Folkehelseinstituttet.
- Ostrov, J. M., & Hart, E. J. (2014). Observational Methods. In T. D. Little (Ed.), *The Oxford Handbook of Quantitative Methods*. Oxford University Press.
- Pakarinen, E., Lerkkanen, M.-K., Poikkeus, A.-M., Kiuru, N., Siekkinen, M., Rasku-Puttonen, H., & Nurmi, J.-E. (2010). A validation of the Classroom Assessment Scoring System in Finnish kindergartens. *Early Education and Development*, 21(1), 95–124. <https://doi.org/10.1080/10409280902858764>
- Pakarinen, E., Lerkkanen, M.-K., & von Suchodoletz, A. (2020). Teacher emotional support in relation to social competence in preschool classrooms. *International Journal of Research & Method in Education*, 43(4), 444–460. <https://doi.org/10.1080/1743727X.2020.1791815>
- Peisner-Feinberg, E. S. (2007). Child care and its impact on young children’s development. *Child Care - Early Childhood Education and Care*, 6.
- Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., & Yazejian, N. (2001). The relation of preschool child-care quality to children’s cognitive and social developmental trajectories through second grade. *Child Development*, 72(5), 1534–1553. <https://doi.org/10.1111/1467-8624.00364>
- Perlman, M., Falenchuk, O., Fletcher, B., McMullen, E., Beyene, J., & Shah, P. S. (2016). A systematic review and meta-analysis of a measure of staff/child interaction quality (the Classroom Assessment Scoring System) in early childhood education and care settings and

- child outcomes. *PloS One*, *11*(12), e0167660-. <https://doi.org/10.1371/journal.pone.0167660>
- Phares, V., Compas, B. E., & Howell, D. C. (1989). Perspectives on child behavior problems: Comparisons of children's self-reports with parent and teacher reports. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, *1*(1), 68. <https://doi.org/10.1037/1040-3590.1.1.68>
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System Manual, PRE-K*. Teachstone Training, LLC.
- Planalp, E. M., & Braungart-Rieker, J. M. (2015). Trajectories of regulatory behaviors in early infancy: Determinants of infant self-distraction and self-comforting. *Infancy: The Official Journal of the International Society on Infant Studies*, *20*(2), 129–159. <https://doi.org/10.1111/inf.12068>
- Raftery, A. E. (1995). Bayesian Model Selection in Social Research. *Sociological Methodology*, *25*, 111–163. <https://doi.org/10.2307/271063>
- Rapoport, E., Reisert, H., Schoeman, E., & Adesman, A. (2020). Reporting of child maltreatment during the SARS-CoV-2 pandemic in New York City from March to May 2020. *Child Abuse & Neglect*. <https://doi.org/10.1016/j.chiabu.2020.104719>
- Ree, M., & Emilson, A. (2020). Participation in communities in ECEC expressed in child–educator interactions. *Early Child Development and Care*, *190*(14), 2229–2240. <https://doi.org/10.1080/03004430.2019.1566230>
- Reifman, A., & Garrett, K. (2010). Winsorize. *Encyclopedia of Research Design*, 1636–1638.
- Reis, A. H., Oliveira, S. E. S., Bandeira, D. R., Andrade, N. C., Abreu, N., & Sperb, T. M. (2016). Emotion regulation checklist (ERC): Preliminary studies of cross-cultural adaptation and validation for use in Brazil. *Temas Em Psicologia*, *24*(1), 97–116. <https://doi.org/10.9788/TP2016.1-07>
- Richter, J. (2015). Preliminary evidence for good psychometric properties of the Norwegian version of the Brief Problems Monitor (BPM). *Nordic Journal of Psychiatry*, *69*(3), 174–178. <https://doi.org/10.3109/08039488.2014.951070>

- Salkind, N. (2010). Demographics. In *Encyclopedia of research design* (Vol. 1, p. 347).
<https://doi.org/10.4135/9781412961288>
- Sandilos, L. E., & DiPerna, J. C. (2014). Measuring quality in kindergarten classrooms: Structural analysis of the Classroom Assessment Scoring System (CLASS K-3). *Early Education and Development, 25*(6), 894–914. <https://doi.org/10.1080/10409289.2014.883588>
- Shields, A., & Cicchetti, D. (1998). Emotion regulation among school-age children: The development and validation of a new criterion Q-sort scale. *Developmental Psychology, 33*(6), 906.
<https://doi.org/10.1037/0012-1649.33.6.906>
- Skoglund, R. I., & Åmot, I. (2020). When anger arises in the interaction with children in kindergartens – the staff’s reactions to children’s resistance. *Scandinavian Journal of Educational Research, 0*(0), 1–12. <https://doi.org/10.1080/00313831.2020.1739138>
- Slot, P. L., Boom, J., Verhagen, J., & Leseman, P. P. M. (2016). Measurement properties of the CLASS Toddler in ECEC in the Netherlands. *Journal of Applied Developmental Psychology*.
<https://doi.org/10.1016/j.appdev.2016.11.008>
- Solheim, E. (2013). *Effects of childcare on child development. Time in care, group size, and the teacher-child relationship* [Thesis for the degree of Philosophiae Doctor, Norges teknisk-og naturvitenskapelige universitet, NTNU].
https://ntnuopen.ntnu.no/ntnu-xmlui/bitstream/handle/11250/271006/628180_FULLTEXT02.pdf?sequence=2&isAllowed=y
- Sommersel, H. B., Vestergaard, S., & Larsen, M. S. (2013). *Kvalitet i barnehager i skandinavisk forskning 2006-2011: En systematisk forskningskartlegging*. Dansk Clearinghouse for Uddannelsesforskning.
<https://www.udir.no/globalassets/filer/tall-og-forskning/rapporter/2013/kunnskap-om-kvalitet-i-barnehager-2.pdf>
- Spinrad, T. L., & Eisenberg, N. (2019). Prosocial emotions. In V. LoBue, K. Pérez-Edgar, & K. A. Buss (Eds.), *Handbook of Emotional Development* (pp. 351–372). Springer International Publishing. https://doi.org/10.1007/978-3-030-17332-6_14

- Spinrad, T. L., Eisenberg, N., Gaertner, B., Popp, T., Smith, C. L., Kupfer, A., Greving, K., Liew, J., & Hofer, C. (2007). Relations of maternal socialization and toddlers' effortful control to children's adjustment and social competence. *Developmental Psychology*, *43*(5), 1170–1186. <https://doi.org/10.1037/0012-1649.43.5.1170>
- Spinrad, T. L., Eisenberg, N., Kupfer, A., Gaertner, B., & Michalik, N. (2004, May). *The Coping with Toddlers' Negative Emotions Scale*. Poster session presented at the Biennial international conference on infant studies, Chicago, IL.
- SSB. (2020a). *Barn 0-17 år, etter antall hjemmeboende søsken, 2001—2020*. PX-Web SSB. <http://www.ssb.no/statbankstatbank/table/06242/>
- SSB. (2020b). *Befolkningens utdanningsnivå, 1. Oktober 2019*. ssb.no. <https://www.ssb.no/utdanning/statistikker/utniv/aar/2020-06-19>
- SSB. (2020c). *Innvandrere og norskfødte med innvandrerforeldre*. ssb.no. <https://www.ssb.no/befolkning/statistikker/innvbef/aar/2020-03-09>
- SSB. (2020d, March 13). *Barnehager*. ssb.no. <https://www.ssb.no/utdanning/statistikker/barnehager/aar-endelige/2020-03-13>
- Stifter, C. A., & Augustine, M. (2019). Emotion regulation. In V. LoBue, K. Pérez-Edgar, & K. A. Buss (Eds.), *Handbook of Emotional Development* (pp. 405–430). Springer International Publishing. https://doi.org/10.1007/978-3-030-17332-6_20
- Stifter, C. A., & Braungart, J. M. (1995). The regulation of negative reactivity in infancy: Function and development. *Developmental Psychology*, *31*(3), 448–455. <https://doi.org/10.1037/0012-1649.31.3.448>
- Stuck, A., Kammermeyer, G., & Roux, S. (2016). The reliability and structure of the Classroom Assessment Scoring System in German pre-schools. *European Early Childhood Education Research Journal*, *24*(6), 873–894. <https://doi.org/10.1080/1350293X.2016.1239324>
- Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, *48*(6), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>

- Teachstone Home*. (n.d.). Teachstone. Retrieved 3 March 2021, from <https://teachstone.com/>
- Tronick, E., Als, H., Adamson, L., Wise, S., & Brazelton, T. B. (1978). The infant's response to entrapment between contradictory messages in face-to-face interaction. *Journal of the American Academy of Child Psychiatry*, *17*(1), 1–13.
[https://doi.org/10.1016/S0002-7138\(09\)62273-1](https://doi.org/10.1016/S0002-7138(09)62273-1)
- Udir. (2017). Hva er kvalitetsutvikling i barnehagen? In *Utdanningsdirektoratet*.
Utdanningsdirektoratet.
<https://www.udir.no/kvalitet-og-kompetanse/kvalitet-i-barnehagen/hva-er-kvalitet-i-barnehagen/>
- Udir. (2020). *Konsekvenser av smitteverntiltakene i barnehager våren 2020*.
<https://www.udir.no/tall-og-forskning/finn-forskning/tema/konsekvenser-av-smitteverntiltakene-i-barnehager-varen-2020/>
- UNICEF. (2008). *The child care transition* (Innocenti Report Card 8). UNICEF Innocenti Research Center, Florence. https://www.unicef-irc.org/publications/pdf/rc8_eng.pdf
- Walker-Andrews, A. S., & Lennon, E. (1991). Infants' discrimination of vocal expressions: Contributions of auditory and visual information. *Infant Behavior and Development*, *14*(2), 131–142. [https://doi.org/10.1016/0163-6383\(91\)90001-9](https://doi.org/10.1016/0163-6383(91)90001-9)
- Wang, M. V., Engvik, M., Evensen, L. A., Gustavson, K., Jin, F., Johansen, R., Lekhal, R., Schjølberg, S., & Aase, H. (2014). *Sammenhenger mellom barnehagekvalitet og barns fungering ved 5 år: Resultater fra Den norske mor og barn-undersøkelsen* (2014:1; p. 115).
Folkehelseinstituttet.
<https://www.fhi.no/globalassets/dokumenterfiler/rapporter/2014/Sammenhenger-mellom-barnehagekvalitet-og-barns-fungering-ved-5-aar.pdf>
- Wathne Oseland, S. (2019). *The Emotion Regulation Checklist—Factor form invariance and acquiescence in Norway* [NTNU].
<https://ntnuopen.ntnu.no/ntnu-xmlui/bitstream/handle/11250/2630057/Oseland,%20Sara%20Wathne.pdf?sequence=1>

- Westergård, E., Ertesvåg, S. K., & Rafaelsen, F. (2019). A preliminary validity of the Classroom Assessment Scoring System in Norwegian lower-secondary schools. *Scandinavian Journal of Educational Research*, 63(4), 566–584. <https://doi.org/10.1080/00313831.2017.1415964>
- Wilkins, D., Khan, M., Stabler, L., Newlands, F., & McDonnell, J. (2018). Evaluating the Quality of Social Work Supervision in UK Children's Services: Comparing Self-Report and Independent Observations. *Clinical Social Work Journal*, 46(4), 350–360. <https://doi.org/10.1007/s10615-018-0680-7>
- World Health Organization. (2004). *The importance of caregiver-child interactions for the survival and healthy development of young children: A review*. Dept. of Child and Adolescent Health and Development, World Health Organization.
- Yan, X., Andrews, T. J., & Young, A. W. (2015). Cultural similarities and differences in perceiving and recognizing facial expressions of basic emotions. *Journal of Experimental Psychology: Human Perception and Performance*, 42(3), 423. <https://doi.org/10.1037/xhp0000114>
- Zahn-Waxler, C., Friedman, R. J., Cole, P. M., Mizuta, I., & Hiruma, N. (1996). Japanese and United States preschool children's responses to conflict and distress. *Child Development*, 67(5), 2462–2477. JSTOR. <https://doi.org/10.2307/1131634>