

Doctoral theses at NTNU, 2021:144

Kenneth Stensen

Early detection of preschool children at risk for mental health problems

NTNU
Norges teknisk-naturvitenskapelige universitet
Avhandling for graden
philosophiae doctor
Fakultet for medisin og helsevitenskap
Institutt for psykisk helse



Norwegian University of
Science and Technology

Kenneth Stensen

Early detection of preschool children at risk for mental health problems

Thesis for the Degree of Philosophiae Doctor

Trondheim, May 2021

Norwegian University of Science and Technology
Faculty of Medicine and Health Sciences
Department of Mental Health

NTNU

Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Medicine and Health Sciences

Department of Mental Health

© Kenneth Stensen

ISBN 978-82-326-5351-5 (printed ver.)

ISBN 978-82-326-5460-4 (electronic ver.)

ISSN 1503-8181 (printed ver.)

ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2021:144

Printed by NTNU Grafisk senter

Tidlig identifisering av barnehagebarn i risiko for å utvikle psykiske vansker

Denne avhandlingen omhandler identifisering av barnehagebarn i risiko for å utvikle psykiske vansker, nærmere bestemt barnehagelæreres evne til å oppdage barn i risiko for denne typen vansker. Så mange som 20% av barn opplever symptomer på psykiske vansker, og med tanke på at så å si nesten alle norske barn går i barnehagen, har barnehagearenaen et stort potensial til å identifisere og hjelpe barn i risiko for å utvikle psykiske vansker. For å kunne gi barn riktig hjelp til rett tid må de barna som vil kunne dra nytte av hjelpen først ha blitt identifisert.

Avhandlingens tre studier er basert på data fra prosjektet *Barn i Midt-Norge*, hvor hovedfokuset var på barns psykiske helse og relasjonskvaliteten mellom barna og barnehagelærere. Dette prosjektet ble gjennomført fra 2012 til 2017 i kommunene Steinkjer, Volda, og det som tidligere var Klæbu kommune. Hovedformålet med avhandlingen var å undersøke om barnehagelærernes bekymringer for barns utvikling gjenspeilet deres oppfatning av barnas psykiske vansker. I artikkel 1 ble det undersøkt hvorvidt barnehagelærere bekymret seg for barna som de selv hadde vurdert til å ha et klinisk symptomtrykk av psykiske vansker. I artikkel 2 var hensikten å gjennomføre en nærmere undersøkelse av de barna som ble klassifisert som *falske positive*, altså de tilfellene hvor barnehagelærere uttrykte bekymring for barn uten at et klinisk symptomtrykk faktisk ble rapportert. Mer spesifikt ble det undersøkt hvordan ulike faktorer påvirket sannsynligheten for om barnehagelærere uttrykte bekymring eller ikke, samt om barna de uttrykte bekymring for med ikke-klinisk symptomtrykk skilte seg ut fra de barna de ikke uttrykte bekymring for med ikke-klinisk symptomtrykk (*sanne negative*). Da det eksisterer få screeningverktøy for barnehagebarn som også inkluderer de yngste barna, ble det gjennomført en valideringsstudie

av *Ages & Stages Questionnaire: Social-Emotional (ASQ:SE)* i artikkel 3. ASQ:SE er et mye brukt screeningverktøy internasjonalt, men det har også blitt tatt i bruk i Norge uten at det først har blitt validert eller at norske normer forelå. For å være sikre på at barn får riktig hjelp til rett tid er det viktig å undersøke validiteten og nøyaktigheten av instrumenter som brukes til identifisering, da feilklassifiseringer kan belaste barn og deres familier unødige, og i verste fall føre til at barn med hjelpebehov blir oversett.

Avhandlingen konkluderer med at barnehagelæreres bekymringer bør bli tatt på alvor, da deres bekymringer omfatter en stor del av barna med et klinisk symptomtrykk. Det er også et forbedringspotensial med tanke på klassifiseringsnøyaktighet, da det også er svært mange falske positive tilfeller. I tillegg er det en liten andel barnehagebarn som barnehagelærere vurderer til å ha et klinisk symptomtrykk, uten at barnehagelærere uttrykker bekymring for dette. Når det er sagt, de falske positive tilfellene viser betydelig mer psykiske vansker og dårligere relasjonskvalitet enn de sanne negative tilfellene, derfor foreslås det i denne avhandlingen at barnehagelæreres bekymring kan anvendes som en for-screening for å identifisere barn som bør følges opp med et standardisert screeningverktøy, slik som f.eks. ASQ:SE, i en sekvensiell screeningprosess, noe som kan bidra til at tiltak iverksettes for barna som trenger hjelp. Et annet viktig funn er at ikke-bekymring i stor grad sammenfaller med lave skårer for psykiske vansker. Med andre ord, når barnehagelærere ikke uttrykker bekymring for et barn, så vil det barnet stort sett befinne seg i det som ansees som normalsjiktet av symptomtrykk. Dermed kan ikke-bekymring være en effektiv måte å sile ut de barna som ikke trenger å følges opp med et grundigere verktøy. ASQ:SE viste gode måleegenskaper når man brukte norske normer og kan derfor anbefales om et oppfølgingsscreeningverktøy i de tilfellene barnehageansatte er bekymret eller usikre, forutsatt at barnet er eldre enn to år. For barn under to år viste ASQ:SE utilfredsstillende måleegenskaper, derfor bør andre instrumenter eller observasjonsmetoder brukes i disse

tilfellene. Funn fra denne avhandlingen kan bidra til bevissthet rundt styrker og svakheter ved barnehageansattes «magefølelse» i det å gjenkjenne barn med mer eller mindre psykiske vansker. Beskrivelser av barnehagelæreres bekymringer sammen med screeningdokumentasjon kan føre til bedre henvisninger til Pedagogisk-Psykologisk Tjeneste (PPT), og eventuelt videre til spesialisthelsetjenesten. Kunnskap om barnehagelæreres svakheter i sine vurderinger kan bidra til økt fokus på kompetanseheving, samt at PPT kan selv også utvikle sitt eget utredningsarbeid, f.eks. ved å anbefale eller selv gjennomføre gode observasjonsmetoder for de yngste barnehagebarna.

Navn kandidat: Kenneth Stensen

**Institutt: Regionalt kunnskapssenter for barn og unge – Psykisk helse og barnevern
(RKBU Midt-Norge), Institutt for psykisk helse, Fakultet for medisin og helsevitenskap,
NTNU**

Veiledere: May Britt Drugli, Frode Stenseng, Jan Wallander

Finansieringskilde: Internfinansiert

Contents

- 1 Introduction 1
 - 1.1 Topic of the thesis 1
 - 1.2 Theoretical framework 3
 - 1.3 Rationale for the identification of preschool children with mental health problems ..5
 - 1.4 Identifying mental health problems in preschool children 9
 - 1.5 Early intervention 16
 - 1.6 Screening accuracy 18
 - 1.7 Classification accuracy of two screening procedures 22
 - 1.8 Aims of the thesis 26
- 2 Methods 29
 - 2.1 Sample and procedures 29
 - 2.2 Measures 30
 - 2.3 Data analyses 33
- 3 Results overview 35
 - 3.1 Study 1: Childcare Providers’ Nominations of Preschool Children at Risk for Mental Health Problems: Does it Discriminate Well Compared to the Caregiver-Teacher Report Form (C-TRF)? 35
 - 3.2 Study 2: Teacher Nominations of Preschool Children at Risk for Mental Health Problems: How False is a False Positive Nomination and What Make Teachers Concerned? 38

3.3	Study 3: Screening for mental health problems in a Norwegian preschool population. A validation of the Ages and Stages Questionnaire: Social- Emotional (ASQ:SE)	39
4	Discussion.....	41
4.1	Considering identification, non-identification, and misclassification.....	43
4.2	“Gold standards” and accuracy estimates	50
4.3	Preschool teachers as informants	54
4.4	Strengths and limitations.....	58
4.5	Clinical implications	60
4.6	Conclusions and suggestions for future research	64
5	References	67
6	Appendix	85

Acknowledgment

First, I would like to give a big heartfelt thanks to the Regional Center for Child and Youth Mental Health and Child Welfare – Central Norway for the opportunity to work with the PhD-project in such a great environment with colleagues and fellow PhD-candidates, while at the same time performing my other assignments.

This thesis is the result of the relentless work of the project group in the *Children in Central Norway project*, which spanned approximately seven years. A big thanks to Anne Synnøve Brenne, Odd Sverre Westbye, May Britt Drugli, Frode Stenseng, and Håvard Tveit for making this PhD-project possible. I would also like to thank all parents who participated and provided consent, preschool teachers, and administrators in the municipalities Steinkjer, Volda, and former Klæbu for their effort. Without them this project would not have been possible.

I am deeply grateful to my three supervisors May Britt Drugli, Frode Stenseng, and Jan Wallander. Their support, guidance, and knowledge have been profound in the work with the thesis. I also wish to thank my co-authors Stian Lydersen and Marianne de Wolff. Their contributions have been invaluable in the realization of this PhD-project.

Finally, I would like to give a big thanks to my wife Oda, my family, and my in-laws for all the patience when I got lost in my work or my mind drifted away, as well as their interest for the subject and the continuous support and encouragement throughout the work process.

Trondheim, January 2021

Kenneth Stensen

List of papers

Study 1: Childcare Providers' Nominations of Preschool Children at Risk for Mental Health Problems: Does it Discriminate Well Compared to the Caregiver-Teacher Report Form (C-TRF)? *Published in Scandinavian Journal of Educational Research (2021).*

Study 2: Teacher Nominations of Preschool Children at Risk for Mental Health Problems: How False is a False Positive Nomination and What Make Teachers Concerned? *Undergoing review in Journal of Psychopathology and Behavioral Assessment.*

Study 3: Screening for mental health problems in a Norwegian preschool population. A validation of the Ages and Stages Questionnaire: Social- Emotional (ASQ:SE). *Published in Child and Adolescent Mental Health (2018).*

Abbreviations and acronyms used in the thesis

ADHD= Attention Deficit Hyperactivity Disorder

ASEBA= Achenbach System of Empirically Based Assessment

ASQ:SE= Ages & Stages Questionnaire : Social- Emotional

AUC= Area under curve

BITSEA= the Brief Infant-Toddler Social and Emotional Assessment

CAPA= The Child and Adolescent Psychiatric Assessment

CBCL= Child Behavior Checklist

CI= Confidence interval

C-TRF= Caregiver-Teacher Report Form

DAWBA= The Development and Well-being Assessment

DIF= Differential item functioning

DISC= The Diagnostic Interview Schedule for Children

DSM= Diagnostic and statistical manual of mental disorders

FN= False negative

FP= False positive

ICD= International Classification of Disease

OR= Odds ratio

ROC= Receiver operating characteristic

NPV= Negative predictive value

PPV= Positive predictive value

SDQ= Strength and Difficulties Questionnaire

S-TRS= Student-Teacher Relationship Scale

TN= True negative

TP= True positive

1 Introduction

1.1 Topic of the thesis

As development in the early years of life establishes foundation for development later in life (Center on the Developing Child Harvard University, 2010), an increasing number of researchers have suggested that the preschool period is an important time for identifying and preventing the development of mental health problems before stable patterns emerge and problems evolve into disorders (Dougherty et al., 2015; Doyle, et al., 2009; Essex et al., 2009; Feeney-Kettler et al., 2010; Heckman, 2006; Heo & Squires, 2011; Kauffman, 1999; McCabe & Altamura, 2011; Njoroge & Bernhart, 2011; Nores & Barnett, 2010; Poulou, 2015; Raver et al., 2009; de Wolff et al., 2013). To provide support for children who would most likely benefit from it requires that they first are identified; thus, it is important to develop psychometrically valid identification procedures to ensure that children who are in need of help receive it (Feeney-Kettler et al., 2010). However, this may be challenging during a period where development occurs rapidly (Keenan et al., 1998); some behaviors may be perceived appropriate and normative at one age or context, but inappropriate and abnormal at another. This thesis examines baseline data collected in 2012-2014 from the Children in Central Norway study. The overall aims of that study were to enhance the competence of preschool teachers in addressing preschool children's mental health and to improve the relational quality between preschool teachers and children.

Parents are usually the people who contact services when there are concerns regarding young children's development (Ellingson et al., 2004). Since preschools are a promising venue for the early identification and screening of mental health problems, more attention should be directed towards preschool teachers' perception of problem behaviors among children (Poulou, 2015). Because preschool teachers frequently interact with and spend so

much time with children, they are uniquely positioned to raise concerns for children who they perceive at risk for mental health problems and follow up with necessary steps to ensure that support and help are provided for children in need. However, there has been little focus on the accuracy of preschool teachers' concerns for children's development and how it reflects children's mental health problems, especially for the youngest children. Thus, this thesis sought to examine the accuracy of preschool teachers' concerns about children's development as well as to investigate the psychometric properties of a brief screening instrument to be used by preschool teachers for preschool children.

When discussing classification accuracy, one must also address the issues and consequences of misclassification. Misclassifications in the form of false positives may be a particular issue, as they might not be false after all (Glascoe, 2001; Jensen & Watanbe, 1999). Based on preschool teachers' concerns, we wanted to take a closer look at which factors contributed to false positives (e.g. preschool teachers' identifying concern about children who in fact are displaying a non-clinical symptom level of mental health problems). By investigating both an unstandardized screening approach (preschool teachers' concern) and a standardized approach (a standardized scalar screening instrument), we can illuminate preschool teachers' ability to discriminate between children with and without mental health problems and factors that leads them to express concerns for children's development, as well as the utilization of a standardized screening instrument and the applicability of this instrument in a Norwegian preschool population. Examining these elements may provide valuable information on how screening procedures could be organized in a preschool context.

To inform the discussion about screening preschool children, an introductory theoretical framework for understanding young children's development will be presented, followed by prevalence and stability of mental health problems in young children. Second, the current status of the identification of children with mental health problems will be

discussed before reviewing the classification accuracy and psychometric aspects of the teacher nomination method and the screening instrument of interest, the Ages & Stages Questionnaire: Social-Emotional (ASQ:SE). After presenting the methodology and results from the studies constituting this thesis, the findings will be interpreted in the discussion with respect to identification, non-identification, and misclassification, and issues with screening and accuracy will also be examined. Finally, preschool teachers' roles in identifying children with mental health problems will be discussed and suggestions about the clinical implications of the findings will be presented.

1.2 Theoretical framework

As development occurs rapidly from birth to school entry, developmental issues need to be addressed to understand the rationale behind screening. Because no single theory or model is sufficient to explain children's development of mental health problems, system theory models in the form of transactions and developmental cascades have been chosen as the framework for the present thesis. An understanding of these models may shed light on the complexity of screening young children and inform the discussion of classification accuracy in distinguishing normative from atypical behavior in early development.

A transactional understanding of development suggests that any process in the individual is influenced by interactions with the individuals' context. Thus, behavior at any one time is a result of the individual and his or her experiences in the environment (Sameroff & Mackenzie, 2003). In other words, development occurs due to dynamic interactions of systems within and outside of an individual over time (Cox et al., 2010). Children and their environment interchangeably influence each other (Sameroff, 2009). For instance, a child's behavior may invoke certain responses from the child's environment (e.g., parents), where

environmental responses will influence the child's subsequent behavior by responding to the initial behavior and so on. The effect of developmental interactions and transaction spreads across levels, domains, and systems. Masten and Cicchetti (2010) have termed this concept *developmental cascades*, which may alter the course of development. Two related concepts are those of *equifinality* and *multifinality*, whereas the former refers to initial individual differences that may lead to the same outcome, while the latter refers to that similar initial conditions may lead to different outcomes (Cicchetti & Rogosch, 1996; Feiring & Lewis, 1987). The flexibility of developmental trajectories makes equifinality and multifinality particularly relevant for developmental psychopathology due to its commonality (Coghill & Sonuga-Barke, 2012). Thus, development should be regarded as probabilistic rather than absolute. As children develop, the number of proximal factors directly influencing their lives expands beyond the immediate family (Sameroff, 2000), to include childcare centers, teachers, friends, and leisure activities. In turn, these factors are introduced into the transactional model and interact with other factors in children's lives in a circular manner.

Complimentary to a transactional understanding of development is that of attachment theory, which focuses on the dyad between children and caregivers. Young children are dependent on their caregivers to meet their basic needs. According to Bowlby (1979), the quality of early relationships is shaped by internal working models that the child develops through experiences with its caregivers. These internal working models will also influence how the child establish relationships to others. The influence of early attachment to the caregiver on development, for better or for worse, has been documented repeatedly (e.g., Cortazar & Herreros, 2010; Sroufe, 2005; Thompson, 2000). The role of attachment and context in developing and maintaining behavior problems has long been recognized (Zeanah et al., 2011). Increased attention to the role of context has in the expression of psychological symptoms could translate into more accurate assessments of clinical phenomena (Dirks et al.,

2012). Even though development is probabilistic in nature, different factors have varying probabilities and may be related to specific outcomes (Cox et al., 2010). Thus, attempts to intervene against maladjustment may be of great importance for both the individual and society. Given the effects psychopathology can have over time, well-timed and targeted interventions can interrupt negative developmental cascades and promote positive ones (Masten & Cicchetti, 2010).

1.3 Rationale for the identification of preschool children with mental health problems

As mentioned earlier, there is a growing consensus of the importance of the early identification of mental health problems in children so that support can be provided at the earliest point possible. The primary rationale behind intervening early in childhood is that the first 2-3 years constitute a period with high neuroplasticity and critical learning periods (Bilancia & Rescorla, 2010; Davidson & McEwen, 2012; Dougherty et al., 2015; Luby, 2010). This can explain why some early interventions targeting mental health problems are more successful in creating a positive outcome, compared to interventions later in life (Heckman, 2006; Masten & Cicchetti, 2010). However, approximately half of all children with mental health problems are not identified before school entry (Glascoe & Marks, 2011), indicating that the opportunity for early intervention and treatment may be lost for these children who might developmentally have benefited from support and help.

The observational accuracy of adults and access to psychometrically sound screening instruments may contribute to an improved ability to identify children with mental health problems and provide them with mental health services. However, screening and identification are a futile exercise unless they lead to children receiving help for their

problems. In some cases, minor interventions carried out by preschool staff may be sufficient to meet children's needs, while in other cases, a more thorough assessment and tailored professional intervention in mental health services are necessary.

Prevalence and stability of problems

The main distinction between mental health problems and psychiatric disorders is that the former represents a symptom or behavioral phenotype that may be measured categorically or dimensionally, while the latter is usually measured through behavioral criteria, such as onset, duration, and intensity of symptoms. In addition and in contrast to mental health problems, to qualify for a psychiatric disorder the individual needs to display some form of functional impairment or distress across situations and contexts. Mental health problems are socially defined and thus highly interactive with the context individuals lives and the social expectations placed upon them (Costello & Bouras, 2006). Globally, approximately 20% of children are burdened with mental health problems (Belfer, 2008), and pooled prevalence estimates show that 13% to 20% of children meet diagnostic criteria for a psychiatric disorder (Charach et al., 2020; Polanczyk et al., 2015; Vasileva et al., 2020). Moreover, every third 1 to 7 years old child who meets criteria for a psychiatric disorder also fulfills the criteria for at least one additional psychiatric disorder (Vasileva et al., 2020).

Compared to prevalence estimates reported globally, the estimates are somewhat lower in Norway with 15% to 20% of children displaying some mental health problems (Lekhal, 2020; Skogen et al., 2014) and 7% of preschoolers meets diagnostic criteria that would qualify for a psychiatric disorder (Wichstrøm et al, 2012). Parents, preschool teachers, and primary school teachers from Nordic countries tend to report lower symptom scores for emotional and behavioral problems on dimensional measures, such as the Child-Behavior

Checklist (CBCL), the Caregiver-Teacher Report Form (C-TRF), and the Strength and Difficulties Questionnaire (SDQ), compared to other countries (Heiervang, Goodman, A., & Goodman, R., 2008; Rescorla et al., 2012; Rescorla et al., 2014). This trend has also been observed in Norway, where the mean of preschool teacher-rated problem behaviors in preschoolers places Norway as a low scoring society regarding emotional and behavioral problems (Drugli & Stensen, 2019), which corresponds well with teacher-rated problem behaviors for children in primary school (Larson & Drugli, 2011). Despite these lower ratings of problem behaviors, only one-tenth of children with emotional and behavioral problems have received professional help for their problems (Wichstrøm et al., 2014) and only 4% of Norwegian 4-5 years old children receives special educational support, with the majority of them being boys (79% vs. 21%) (Lekhal, 2020). This may be because Norwegian parents and teachers tend to under-recognize or under-report emotional problems, thus, their ratings do not reflect the actual prevalence of problems (Heiervang et al., 2008). Additionally, only behavioral problems (e.g., aggression) increase the probability of service use in Norwegian preschoolers, while emotional problems (e.g., anxiety) do not (Wichstrøm et al., 2014). Although preschool children show similar prevalence estimates for mental health problems and disorders as older children, these problems seem to be under-identified, under-referred, and under-treated (Egger & Angold, 2006; Horwitz et al., 2003; Horwitz et al., 2007).

Some emotional and behavioral problems will come and go during early childhood as children develop, but some problems may persist or even escalate over time (Powell et al., 2006). Developmental precursors for emotional and behavioral problems have been identified in the first two years of life. For example, Keenan and colleagues (1998) found that a difficult temperament when children were 18 months old was significantly related to boys' and girls' internalizing problems at 3 and 5 years old, while noncompliance in girls and aggression in boys at the early stage were related to later externalizing problems. It has been reported that

37% of 18-months-old children with extreme emotional or behavioral problems continue to display extreme difficulties one year later (Mathiesen & Sanson, 2000). Homotypic problem persistence has also been found in infants and toddlers, where 38%-50% of the children demonstrated the same type of problems one year later (Briggs-Gowan et al., 2006). Additionally, 30% of 3-year-old children with a high parent-reported score of internalizing and externalizing problems also received high parent-reported scores when they were 12 years old (Pihlakoski et al., 2006). Preschoolers who meet diagnostic criteria at age 3 are five times more likely to meet diagnostic criteria at age 6 than those who did not (Bufferd et al., 2012). Angold and colleagues (1999) found that children and adolescents who had symptoms of psychopathology but did not meet the DSM criteria for a disorder, still experienced considerable impairment. Thus, functional impairment can occur at a lower level of disturbance than implied by a diagnosis. Approximately 50% of preschoolers with a diagnosis still have a diagnosed psychiatric disorder in middle childhood or early adolescence (Finsaas et al., 2018a).

Whereas some children outgrow their disorders, an equally large number do not (Lavigne et al., 1998). Especially children with co-occurring internalizing and externalizing problems are more likely to display persistent problems (Basten et al., 2015; Beyer et al., 2012). Fanti and Henrich (2010) followed children over a 10 year period from 2 to 12 years old and found that children exhibiting continuous externalizing problems or co-occurring internalizing and externalizing problems were more likely to engage in risky behaviors, associate with deviant peers, be rejected by peers, and be asocial during early adolescence. Children who exhibit continuous internalizing problems were only at higher risk for being asocial with peers during early adolescence. Long-lasting deleterious effects of childhood psychopathology have also been found by Finsaas and colleagues (2018b). In their study, having a disorder in early childhood predicted greater impairment and poorer functioning in

adolescence, even when controlling for concurrent mental health problems. Additionally, a subthreshold subsample without diagnosis in early childhood also exhibited significant longitudinal association between depression and ADHD symptoms in childhood and depression and functional outcomes in adolescence. Many subthreshold conditions have predictive validity and may be precursors for disorders later in life. Thus, subthreshold conditions may be effective targets for preventive interventions (Shankman et al., 2009).

The strongest predictor for future mental health problems is usually the status of current mental health problems (Bilancia & Rescorla, 2010). However, predicting who will not develop mental health problems tends to be more accurate than predicting who will develop mental health problems, a result of low positive predictive value (PPV) of measurements, which will be discussed in a later section in the thesis. However, prior research suggests that a substantial proportion of children experience non-transient problems that lead potentially to serious consequences for their development and well-being. As mental health problems are relatively common for young children and they may be precursors for later deleterious disorders, the prevalence estimates and low referral rates makes identification of children with mental health problems at an early stage and providing help for those in need a major public health concern.

1.4 Identifying mental health problems in preschool children

Traditionally, pediatricians have identified developmental delays or mental health problems in children. However, a review of 11 studies from the USA showed that pediatricians working without standardized screening instruments had a low accuracy in identifying developmental and/or behavioral problems in children (sensitivity range 14-54%) but had an acceptable to high accuracy identifying children without problems (specificity range 69-

100%) (Sheldrick et al., 2011). Low rates of identifying children with problems were also reported in a Danish study investigating the accuracy of health nurses' concerns about children's development, the mother-child relationship, or family functioning against ICD10 diagnoses. In addition, health nurses exhibit a low accuracy in identifying children without problems and a high false positive rate (76%) (Skovgaard et al., 2008). When parents raise concerns about their child and directs pediatricians' attention towards particular behaviors, the identification rate increases considerably, thereby underlining the importance of including parents in the identification process (Glascoe, 1997; Glascoe, 2003; Glascoe & Marks, 2011). Parental concerns are also a robust predictor of seeking help for children with behavioral problems (Ellingson et al., 2004). However, among parents with a child who has been identified with mental health problems by primary care professionals, only 67-72% of them receives advices, and only 26-42% receive a specialist referral (Charach et al., 2020). Thus, pediatricians and health nurses working without standardized screening instruments seem to have a hard time accurately identifying children at risk for mental health problems and making accurate clinical decisions, unless parents voice their concerns. If parents do not show concern, many cases may go unidentified if the responsibility of identification and help-seeking behaviors are placed solely upon parents or pediatricians (Lavigne et al., 2016a).

In addition to parents, preschool teachers are the only viable source of information regarding young children's development (Sveen et al., 2013), and because emotional and behavioral problems may be context specific (Achenbach & Rescorla, 2000), preschool teachers' may play an important role in identifying children with mental health problems and connecting them with relevant mental health services (Berkhout et al., 2012; Eklund et al., 2009; Poulou, 2015). In Norway, approximately 92% of children aged 1 to 5 years old attend childcare centers (Statistics Norway, 2020). During the week, children spend a considerable number of hours in the childcare center together with their preschool teachers and other staff

members, such as assistants. In contrast to assistants, who usually have limited formal education in child development, preschool teachers hold a bachelor's degree and have at least some education in child development, which together with experience with multiple children could make a potentially valuable reference base for discriminating between age appropriate normal from abnormal behaviors.

The Norwegian Kindergarten Act (2017) states that preschool teachers have a responsibility to ensure that the necessary next steps are undertaken if they have behavioral concerns about a child (e.g., apply observational procedures or relevant tests if needed). If preschool teachers are still uncertain, a referral to the Educational and Psychological Counselling Service for a more thorough follow-up evaluation should be given (with parental consent), which in turn could lead to further referrals to specialized mental health services for a clinical assessment. Thus, preschool teachers may have an important role as vanguards by identifying children with needs and providing access to relevant mental health services.

However, surprisingly little research has been conducted on preschool teachers' ability to classify children at risk for mental health problems, especially the youngest preschool children (one to three years old). Consequently, preschool teachers' ability to discriminate between children at risk for mental health problems and those not at risk will be examined in the present thesis.

Screening

Screening may only be regarded as an indicator of the presence or absence of some aspects of a target condition to determine whether a referral for a more thorough assessment should be made. The screening result will fall within one of the four following categories: true positive (positive screening and condition present), true negative (negative screening and condition

absent), false positive (positive screening but condition absent), and false negative (negative screening but condition present). Screening procedures may also capture those who fall beneath a clinical/diagnostic threshold, but still may have elevated levels of the relevant symptoms. Compared to diagnostic assessments, screening procedures are usually quicker, simpler, and easier to administer. Using an unstandardized approach, such as just asking “Do you have any concerns about this child?”, “Do I have any concerns about this child?”, or “Are there any risk factors present in the child’s life?”, is probably the shortest and simplest way to perform screening. However, as mentioned earlier, pediatricians and health nurses relying solely on their subjective assessment have low accuracy in identifying children with mental health problems when working without standardized instruments (Sheldrick, Merchant, & Perrin, 2011; Skovgaard et al., 2008). This may be due to the frequency and brevity of the basis for their screening, such as child health check-ups during the preschool period. Unless parents elicit concerns, pediatric practitioners may pay attention to aspects of development other than mental health.

Standardized instruments may be a more effective approach to screening. Preschoolers can be screened for emotional and behavioral disorders as efficiently as older children and adults for whom this is a standard approach (Sveen et al., 2013). Screening instruments may be brief and simple and provide a global indication of risk status, making them suitable for universal screening or with larger groups of children. In contrast, longer, time consuming, and complex tools may be more appropriate for targeted or follow-up screening, as they usually contain more information, which could indicate a child’s risk status both globally and specifically (i.e., regarding specific sets of problems).

The accuracy of a screening instrument, whether it be an unstandardized or a standardized instrument, depends on the observational accuracy of the informant. Thus, it is necessary for the informant to have some prior knowledge about the child in question before

screening. For example, the Ages and Stages Questionnaire: Social-Emotional (ASQ:SE), which will be presented in-depth later in this thesis, was designed to be completed by parents. Preschool teachers may also complete this questionnaire, but the ASQ:SE manual recommends that preschool teachers have seen the child 15-20 hours/week prior to completing it (Anuniação et al, 2019; Squires et al., 2002). Thus, this particular instrument is more suitable for parents and preschool teachers rather than pediatricians due to the required time spent with the same children across various situations. Investigating the psychometric properties of screening instruments, such as its reliability, validity, and classification accuracy, is of great importance when evaluating the extent to which an instrument can be used to identify children at risk.

Ethical considerations in screening

As mentioned earlier, The Norwegian Kindergarten Act (2017) states that preschool teachers have a responsibility to ensure that the necessary next steps are taken if they have concerns about a child. If screening is deemed to be necessary and indicates that that the child is at risk (i. e., a positive test), preschool teachers have an ethical obligation to follow-up (American Academy of Child and Adolescent Psychiatrists, 2001). Parental consent is required before screening with standardized instruments. Thus, the positive and negative aspects of screening should be conveyed parents, so they can make an informed decision regarding whether to allow their child be screened. Factors such as the cost and availability of interventions, the intrusiveness of the screening for the child and the family, and the possibility of stigmatizing should be discussed prior to screening (Sawyer et al., 2013). After a child receives a positive score on a screening test, parents may experience anxiety, guilt, and shame. However, good communication skills from professionals can reduce parental stress and anxiety (Hewlett &

Waisbren, 2006). If support systems or mental health services do not have sufficient resources available to accommodate referrals based on screening procedures, it would be ethically questionable to perform screening in the first place.

Some children will outgrow their problems while others will not bounce back to a developmentally normal trajectory (Powell et al., 2006). The relative transient nature of problem behaviors in preschool children poses other ethical questions regarding screening. Some children might display elevated symptom levels at some point in time, but an absence of symptoms at a later time point without any intervention being administered between. A positive screening test at the time when symptoms were present could have warranted further testing and intervention, despite in this case, the child's behavior would normalize as a result of time and maturation. In a worst-case scenario, administering an intervention, including possibly medication, to children who it turns out do not need it may inflict negative developmental consequences on what already is a healthy development camouflaged with transient symptoms. Although ample research exists regarding the stability of mental health problems and disorders, few children with a positive screening test for mental health problems in preschool meets diagnostic criteria for a mental disorder in early adulthood. Moreover, those who meet diagnostic criteria for a mental disorder in early adulthood are rarely identified with symptoms in early childhood (Neyman et al., 2007). These findings indicate that the predictive accuracy of early screening may be limited, further raising ethical questions about its use.

Another issue is the misclassification rates. A screening instrument producing a high rate of false negatives fails to identify children who potentially would benefit from further testing and intervention. A high rate of false positives, on the other hand, could result in the stigmatization of the children mistakenly identified with the condition. In addition, it could create unnecessary worries for parents and placing an unnecessary workload on the mental

health services, which ideally should allocate their resources to help children in actual need of support. The issue of misclassification is attenuated when screening in a normal population, where the base rate of mental health problems is lower compared to an at-risk population. The efficacy of screening instruments in correctly identifying children with potential for the target condition decreases as the base rate of that condition decreases (Lavigne et al., 2016a). In other words, the rarer the targeted condition is, the harder it is to identify. Thus, even psychometrically sound screening instruments can have considerable misclassification rates if the base rate of problems is low, usually in the form of overidentification of potential cases but which turn out to be false (false positives).

The choice between universal or targeted screening may also affect the misclassification rate. Universal screening is conducted at a population level, for example, semi-annually screening all children in childcare centers. Such an approach increases the likelihood of children at risk for mental health problems being identified and referred for further assessment, but it also increases the likelihood of false positive cases. Targeted screening on the other hand, which usually involves subsamples of children identified by means other than screening, based on certain characteristics that are in some way related to the target condition (e.g., concerns from parents or preschool teachers; very low birthweight), could result in fewer cases being misclassified just by nature of being targeted. However, targeted screening may be more prone to false negative cases, as children in need of help might be missed in the selection of the target group.

Which type of screening to perform and the acceptable level of misclassification rates need to be carefully discussed prior to screening, as these choices involves a trade-off between identifying children in need of help versus avoiding burdening those who do not need help. Even though there are high rates of false positive identifications associated with screening in the broad population or with a universal approach, the accuracy of a false

positive screening test result may be debated. As children classified as false positives through screening procedures receives significantly poorer outcome scores on clinical and other risk factor measures compared to children classified as true negatives (Glascoe, 2001; Jensen & Watanbe, 1999), discarding a false positive case blatantly may be a premature decision.

1.5 Early intervention

The prevalence estimates of mental health problems and disorders, the stability of problems, the referral rates, the potential of contextual factors and the biological processes occurring in the preschool period (e.g., neuroplasticity) (Davidson & McEwen, 2012; Halperin et al., 2012; Luby, 2010) strongly indicate that the early identification of and intervention for mental health problems are necessary. In accordance with developmental cascades (Masten & Cicchetti, 2010), reducing problems in one domain may cause a reduction of problems in another domain, and building competency in one domain may yield increased competency in another domain. Consequently, timely interventions may interrupt negative cascades or promote positive cascades (Masten & Cicchetti, 2010), thus increasing the probability of healthy development. This may explain why interventions administered earlier in life, during a period with rapid development provide a better investment than interventions administered later in life (Duncan & Magnuson, 2013; Heckman, 2006; Nores & Barnett, 2010). However, identifying all who would benefit from interventions would be a meaningless exercise if effective interventions were not readily available to be administered.

Universal preventive intervention programs usually aim to decrease the incidence rate of a targeted condition. For example, the Incredible Years Teacher Classroom Management program has been found to be effective in reducing internalizing and externalizing problems in preschoolers by improving the teacher-child relationship. This effect was found for

children scoring with no or few symptoms as well as for children displaying an elevated symptom load of problem behaviors (Tveit et al., 2019). Selective preventive intervention programs are usually carried out by primary health services with the aim of disrupting developmental trajectories where negative outcomes are probable, by reducing or eliminating problems before patterns that are more serious emerge. Training parents of preschool children at risk for developing internalizing or externalizing disorders has been shown to have promising long-term effects (Brotman et al., 2008; Rapee, 2013). Last, targeted preventive interventions are performed by special health services at signs of disorders, but before disorders have fully been manifested. For internalizing disorders such as depression and anxiety in clinical samples, parent-child interaction therapy focusing on, for example, emotional development, has shown promising results in reducing the severity of depression as well as reducing externalizing problems and functional impairment (Lenze et al., 2011; Luby et al., 2012; Luby, 2013). For externalizing disorders, such as ADHD and other disruptive problems, parental training should be the first choice of intervention, and medication should only be used in cases where parent training is ineffective (Charach et al., 2017; Daley et al., 2009). A meta-analysis of 55 studies showed that parental training was more effective than medication for children at risk for ADHD (Charach et al., 2013). The Brief Behavioral Intervention, a parental management program, has exhibited promising results in reducing disruptive behaviors in referred preschool children. At posttreatment, both parents and preschool teachers reported significantly fewer disruptive problems, and there were long-term effects one year after treatment had ended (Axelrad et al., 2013).

Although the abovementioned studies do not represent a complete review of the available studies of intervention as that would be beyond the scope of the current thesis, they indicate that effective interventions are available for children who are identified to be in need of help via early screening for mental health problems. A proposed framework for working

with screening and interventions in early education is the Pyramid Model as a Response to Intervention (Fox et al., 2009). This model emphasizes the use of universal screening to identify children with developmental delays who might need more systematic support (universal, selective, or targeted). Other core features of this model is the continuous monitoring of children's problem behaviors and the collaborative problem-solving process (mainly between parents and preschool teachers, but also specialists if more selective or targeted interventions are needed) to determine if children and teachers are in need of more support or additional interventions to ensure children's healthy development. Even though several universal screening instruments intended for preschool children exists, their accuracy estimates in identifying children in need of support varies (Feeney-Kettler et al., 2010; Lavigne et al., 2016b).

1.6 Screening accuracy

Screening instruments need to have demonstrated validity and an acceptable ability to classify cases and non-cases for this task to be useful. When investigating a screening instruments criterion related validity and accuracy, one or more well-established instruments that should measure the same underlying construct as the screening instrument in question are applied as comparators (sometimes referred to as a "gold standards") to measure the agreement between them (Fayers & Machin, 2007). What constitutes a "gold standard" is debatable. However, in validation and accuracy studies of instruments, the choice of a "gold standard" or a comparator is commonly based on the instrument's psychometric properties and its status among researchers and clinicians in the field. Consequently, criterion related validity and accuracy are established if a high degree of agreement (e.g., a high correlation) is demonstrated between the screening instrument and the comparator. The classification

accuracy of a screening instrument is usually represented by its sensitivity and specificity estimates, which are measured by comparing the classification agreement between the screening instrument and a “gold standard” criterion that is thought to reflect the true condition of the subject. In the present thesis, the Caregiver-Teacher Report Form (C-TRF, (Achenbach & Rescorla, 2000)) has been selected to represent the “gold standard” measure of mental health problems among preschoolers based on its extensive use and its status in the mental health field. The C-TRF is the teacher-reported questionnaire of the Achenbach System of Empirically Based Assessment (ASEBA), which has been proven to have excellent psychometric properties across cultures and has been translated to more than 110 languages (ASEBA.org, 2020; de Groot et al., 1994; Ivanova et al., 2007; Ivanova et al., 2010; Ivanova et al., 2011; Koot, Verhulst, & Boomsma, 1997; Rescorla et al., 2012; Rescorla et al., 2014; Verhulst & Koot, 1992). The C-TRF and its parent-reported counterpart, the Child Behavior Checklist (CBCL), are commonly used comparators when investigating the psychometric properties of other screening or assessment instruments (Lavigne et al., 2016b).

Regarding the accuracy of a screening instrument, sensitivity refers to the instrument’s ability to identify true positive cases (both screening instrument and comparator agree that the condition is present), while specificity is the instrument’s ability to identify true negative cases (both screening instrument and comparator agree that the condition is absent). Misclassification is represented by the screening instrument’s false positives rate (positive on the instrument and negative on the criterion) and false negative rate (negative on the instrument and positive on the criterion). Additional estimates of classification accuracy include the positive and negative predictive value (PPV and NPV) of an instrument. In contrast to sensitivity and specificity, the PPV and NPV account for the population prevalence, and may be considered more clinical useful (Altman, 1991). Thus, the sensitivity and specificity of an instrument may be the same, but the PPV and NPV will vary as a

function of the prevalence. In other words, the same test may perform differently in different populations, as the prevalence of targeted conditions may vary.

Table 1. Two by two table of a screening test against a “gold standard” criterion

	“Gold standard” criterion (e.g., the C-TRF)		
Screening instrument (e.g., the ASQ:SE)	Positive	Negative	Total
Positive	True positive (TP)	False positive (FP)	TP+FP
Negative	False negative (FN)	True negative (TN)	FN+TN
Total	TP+FN	FP+TN	

Sensitivity (or true positive rate) = $TP/(TP+FN)$

Specificity (or true negative rate) = $TN/(FP+TN)$

False positive rate = $FP/(FP+TN)$ or $FP/(TP+FP)$

False negative rate = $FN/(TP+FN)$ or $FN/(FN+TN)$

Positive predictive value (PPV) = $TP/(TP+FP)$

Negative predictive value (NPV) = $TN/(TN+FN)$

The abovementioned PPV and NPV formulas are based on sample prevalence. However, other PPV and NPV estimates can be calculated by inserting different prevalence estimates into the following formula, for example the population prevalence:

$$PPV = \frac{\text{Sensitivity} * \text{Prevalence}}{\text{Sensitivity} * \text{Prevalence} + (1 - \text{Specificity})(1 - \text{Prevalence})}$$

$$NPV = \frac{\text{Specificity} * (1 - \text{Prevalence})}{(1 - \text{Sensitivity}) * \text{Prevalence} + \text{Specificity} * (1 - \text{Prevalence})}$$

Usually, the accuracy estimates can be calculated in a two-by-two contingency table if the results of the screening test and the criterion are known. An alternative approach to binary calculations is performing receiver operating characteristic (ROC) analysis. ROC analyses may be performed to investigate the classification accuracy of an ordinal or continuous indexed screening instrument against a binary, ordinal, or continuous criterion. This would yield an Area Under the Curve (AUC) estimate, which reflects the screening instrument's ability to discriminate between those with the targeted condition and those without. In other words, the AUC can be interpreted as the probability that a person randomly drawn from a sample of individuals with the targeted condition receives a higher score on the screening instrument compared to a person randomly drawn from a sample of individuals without the targeted condition (Lydersen, 2012). An AUC of .50 reflects a lack of discrimination, .70 to .79 indicates an acceptable strength, .80 to .89 indicates an excellent strength, and .90 to 1.00 indicates an outstanding strength of discrimination (Hosmer & Lemeshow, 2000). In other words, an AUC of 1.00 reflects a perfect discrimination based on the screening instrument of individuals with or without the targeted condition as determined by the criterion, while an AUC of .50 indicates that the instrument performs not better than random assignment upon screening and is in essence worthless (one might as well flip a coin). In addition, ROC analyses provide a cutoff matrix for every single possible score attained on the screening instrument, as well as the sensitivity and specificity associated with that particular score.

The ROC analysis is particularly useful when adapting screening instruments to new cultures or populations, or when deliberating about whether the cutoff value needs adjustment to reach a particular goal (e.g., decreasing the cutoff value to identify more with the targeted condition). Thus, adjusting the cutoff value involves a trade-off. Decreasing the cutoff value increases the sensitivity, but it may also increase the rate of false positives. Increasing the cutoff value, on the other hand, can increase the specificity, but at the expense of more false negative cases. Thus, ROC analysis would be an appropriate approach to investigate the consequences of adjusting the cutoff value as this analysis uses the classification and misclassification estimates to indicate whether an instrument performs satisfactorily to achieve the targeted goal (e.g., identifying or ruling out more children).

1.7 Classification accuracy of two screening procedures

In the present thesis, both a two-by-two contingency table and ROC analyses are used to investigate the accuracy of an unstandardized (nomination method) and a standardized (the ASQ:SE) approach to screening for mental health problems in a preschool population.

The nomination method

In its simplest form, the nomination method refers to an informant nominating a child that the nominator believes has met a given criterion or criteria (e.g., the presence of risk factors, certain symptoms, or an overall judgement of developmental concerns). Thus, the nomination method is probably the most time- and cost-effective screening instrument available, as it involves the informant “only” making a global subjective judgment. Even though such an approach provides little information compared to standardized and lengthier screening instruments, it can direct attention towards children or subsets of children who may be in

need of a more thorough screening or follow-up assessment. In other words, the nomination method can also work as a pre-screening method for a more targeted screening approach.

As mentioned earlier, pediatricians working without standardized screening instruments (e.g., relying solely on clinical judgment) have low accuracy in identifying children with developmental and/or behavioral problems (sensitivity range 14-54%), but an acceptable to high accuracy in identifying those without problems (specificity range 69-100%) (Sheldrick et al., 2011). Parents, on the other hand, have demonstrated higher accuracy when voicing concerns about their child's developmental issues. Among 100 children aged 0 to 6 years attending a general hospital outpatient clinic, 20 were identified with developmental problems and 80 were identified to have no developmental problems. Of the 20 children identified with developmental problems, parents elicited concerns about 80% of them (sensitivity 80%), while for the 80 children without problems, parents did not elicit any concerns for 94% (specificity 94%) (Glascoe & Marks, 2011). Similar accuracy estimates regarding parental concerns have previously been reported by Glascoe (1997), but there was a considerably lower specificity among children aged 21 to 84 months (sensitivity 79%, specificity 72%, PPV 31%). Lower accuracy estimates have been observed when parents are asked to nominate children who they think have a higher than average chance to develop a behavioral, emotional, or mental health problem in the future. When asked this question, parents nominated 46% of the children who went on to have a clinical level of internalizing problems one year later and 53% of the children who went on to have a clinical level of externalizing problems. The specificity was 80% for both internalizing and externalizing problems (Dwyer et al., 2006). In other words, parents correctly identified approximately half of those children who would subsequently display a clinical level of symptoms one year later. However, they were more accurate in predicting which children would not develop clinical levels of symptoms compared to children who would display clinical levels of symptoms.

Very little is known about the accuracy of the nomination method among preschool teachers. However, some studies have been carried out with primary school teachers. Primary school teachers tend to nominate children who exhibit externalizing problems more frequently than those with internalizing problems, and teachers report greater concern about externalizing problems than for internalizing problems (Loades & Mastroyannopoulou, 2010; Soles et al., 2008). When asked to nominate children who they thought had a higher than average chance of developing a behavioral, emotional, or mental health problem in the future, teachers correctly identified 34% of those who would subsequently display a clinical level of internalizing problems one year later and 69% of children who would display a clinical level of externalizing problems. The specificity estimates for internalizing problems and externalizing problems were 75% and 77% respectively (Dwyer et al., 2006). Difficulties in correctly nominating those with internalizing problems have been demonstrated in other studies as well. Approximately 50% of children nominated by teachers as being at risk of developing an anxiety disorder are found to have an anxiety disorder, while about 83% children nominated by teachers as being at risk of depression do not meet the criteria for a depressive disorder (Dadds et al., 1997; Moor et al., 2000). When teachers are asked to nominate the children in their class who they perceive as most anxious, the identified children display significantly more anxiety problems compared to those not nominated (Layne et al., 2006). Ollendick and colleagues (1990) asked fourth grade teachers to nominate well-adjusted and at-risk (aggressive or withdrawn) children. Over a five-year period, children nominated as well-adjusted continued to out-perform the children nominated to be at risk, in terms of academic grades, sociometric status, and social behavior. Additionally, children nominated by teachers as being in need of help for mental health services struggle more with self-esteem, academic efficacy, and internalizing symptoms than non-nominees (Roeser & Midgley, 1997).

The ASQ:SE

The ASQ:SE (Squires et al., 2002) is a brief parent-reported screening instrument that was developed in the United States and designed to assist in identifying developmental delays in children aged six to 60 months, where a high score on this instrument may indicate developmental problems. The ASQ:SE has various forms depending on the child's age, and each form is thought to reflect the developmental level associated with that age. The ASQ:SE has demonstrated good reliability (α ranging from .67 to .91 for the various forms and 94% test-retest agreement) in the United States, and has exhibited satisfactory validity indices (Gokiert et al., 2014). Additionally, promising screening properties have been exhibited in other regions of the world (Anuniação et al., 2019; ; Chen et al., 2015; Heo & Squires, 2012; Kucuker et al., 2011; Squires et al., 2002; Vaezghasemi et al., 2020), indicating the potential for this instrument to be adapted cross-culturally. Most of the studies on the ASQ:SE are based on parents' reports. However, Anuniação and colleagues (2019) used preschool teachers as informants and found good factorial validity when using the ASQ:SE in Brazil.

Research on the classification accuracy of the ASQ:SE exists. Feeney-Kettler and colleagues (2010) reported sensitivity estimates ranging from 78% to 85% for identifying mental health problems, and specificity estimates ranging from 93% to 96% depending on the form of the instrument. In a slightly newer review than the abovementioned one by Feeney-Kettler and colleagues (2010), based on five studies, the ASQ:SE exhibited a mean sensitivity of 73% and a mean specificity of 88% (Lavigne et al., 2016b). However, less promising estimates have been reported when used with younger children in the Netherlands. Using the Child Behavior Checklist as a comparator (the parent-reported questionnaire of the ASEBA), de Wolff and colleagues (2013) reported sensitivity estimates of 28% for 6-months old

children (specificity 93%), 38% sensitivity for 14-months old children (specificity 91%), and 66% sensitivity for 24-months old children (specificity 91%). Similar to results for two-year-old children, sensitivity estimates of 65% (specificity 91%) and 63% (specificity 91%) have been reported for three-year-old and four-year-old children, respectively (Theunissen et al., 2015). This said, one possible explanation for the low sensitivity for the youngest children in the abovementioned studies is the choice of comparator, as the Child Behavior Checklist (and the C-TRF) are not intended for children younger than 1.5 years old. All studies mentioned here are based on parent reports. The present thesis will extend knowledge about the screening accuracy of the ASQ:SE by investigating its use in a Norwegian preschool sample based on preschool teacher reports. As the prevalence of problems and how informants perceive items may vary around the world, it is important to conduct validation studies to ensure the appropriateness of an instrument for the intended population. Accessing screening instruments with good psychometric properties is important for ensuring that children in need of help receive support in a timely manner.

1.8 Aims of the thesis

The overall aim of the current thesis was to investigate the classification accuracy of mental health problems in a preschool population based on preschool teacher reports using the nomination method and a standardized screening instrument (the ASQ:SE). In addition, a closer examination of the false positives from the preschool teacher nominations was carried out to understand better this screening process. More specifically, the papers investigated the following:

Aim of Paper 1:

Investigating the accuracy of preschool teacher nominations of preschool children at risk for mental health problems against a well-established comparator, the C-TRF:

- Do preschool teacher nominations of preschool children at risk for mental health problems accurately discriminate between children at risk and those not at risk based on assessment with a standardized instrument?
- Do preschool teachers exhibit higher accuracy for externalizing problems than for internalizing problems?
- Do preschool teachers demonstrate higher accuracy for boys and older children than for girls and younger children?

Aim of Paper 2:

A closer examination of the false positives from paper 1 by comparing children classified as false positives to those classified as true negatives:

- Do children classified as false positives display more behavior problems and experience a poorer teacher-child relationship compared to those classified as true negatives?
- Do children displaying more behavior problems have increased odds of being classified as false positives?
- Does a more negative teacher-child relationship increase the odds of a false positive classification?

Aim of Paper 3:

Investigating the criterion-reference validity and classification accuracy of the ASQ:SE against the C-TRF based on preschool teachers' reports:

- Does the screening accuracy differ for different age forms of the ASQ:SE compared to the C-TRF?
- What are the optimal cutoff values for each form in a Norwegian preschool context?

2 Methods

Data material in the studies

The data are based on the Children in Central Norway study, which was an intervention study focusing on enhancing preschool staff competency with respect to children's mental health, as well as improving the relationship quality between preschool teachers and children. The data applied in the present thesis are based on the baseline data, which were collected in 2012-2014 before any intervention had commenced.

2.1 Sample and procedures

Sample and procedure in studies 1 and 2

Parents with children in childcare centers that served children from age one to six years old in three municipalities in Central Norway (Steinkjer, former Klæbu, and Volda) received recruitment letters with information regarding the project as well as an informed consent form. Information was also provided in parent meetings before the project started. The recruitment letter provided the option for parents to consent either by logging in with a personal invitation code or by returning the consent form to the childcare center. Parental consent gave the preschool teacher in the childcare center who was most familiar with the child permission to complete a survey regarding that parent's child. Preschool teachers provided consent electronically via the survey with their own invitation codes. Participation was voluntary and parental consent could be withdrawn at any time without reprisal until the participation registry was deleted. Of the invited parents, 1631 (77%) consented to enroll their child in the study and the teachers reported on 1431 children (68%). The gender distribution was 51% boys and 49% girls with a mean age of 45 months, and 169 preschool teachers (7% males) from 57 childcare centers participated. Usually, the preschool teachers

reported on 6-12 children each, depending on the size of the childcare centers and the group sizes within. The preschool teachers had at least a bachelor's degree (three years of higher education) in early childhood education.

Sample and procedure in studies 3

In study 3, the N of children was reported to be 1486. After excluding those children who did not meet the lower age criterion of the comparator and those who were administered an age-inappropriate ASQ:SE form, the sample contained 1428 children, 1395 of whom had both complete ASQ:SE and C-TRF data. This n is slightly inaccurate, as 28 children were discovered after the paper had been published to have been registered twice at baseline (the newest of the double data entries were removed before the analyses for studies 1 and 2 were performed, which have the correct N). Thus, the N in study 3 is slightly higher than it should be. However, as the double entries mainly consisted of children in the older age range, the chances of possible inflation or deflation of the outcome are probably miniscule.

2.2 Measures

(more detailed descriptions of the measures can be found in the papers)

Preschool teacher nomination

The preschool teachers were asked to make a global judgment concerning each child's risk status by answering "yes" or "no" to the question "Do you have any concerns for this child?". If "yes" was answered, preschool teachers could provide more details about their nomination by checking one or more reasons for nomination (aggression, attention, emotional, social,

motoric, language, home). However, only those nominated with specifications of aggression, attention, emotional, or social were considered to be “at risk” to correspond with types of problems addressed in the comparator (the C-TRF).

Ages & Stages Questionnaire: Social-Emotional (ASQ:SE)

The ASQ:SE is a brief parent reported instrument intended to identify developmental delays in children aged six to 60 months in the social and emotional domains (Squires et al., 2002). Different forms are used depending on the child’s age and the number of scored items ranges from 19 (six months form) to 33 (48 and 60 months forms). The following age intervals are covered by the different forms: the ASQ:SE 18 (15 to 20 months old), ASQ:SE 24 (21 to 26 months old), ASQ:SE 30 (27 to 32 months old), ASQ:SE 36 (33 to 41 months old), ASQ:SE 48 (42 to 53 months old), and ASQ:SE 60 (54 to 65 months old). There are three response options (rarely or never, sometimes, most of the time) for each item, which are scored zero, five, and ten, respectively. An additional five points can be added for each item if the informant check this specific behavior as worrisome. A total difficulty score is calculated by adding the points from all the items and the items related to expressed concerns. The cutoff scores indicating further assessment provided by the manual vary for each age interval form based on validation studies conducted in the US (Squires et al., 2002).

Caregiver-Teacher report form (C-TRF)

Preschool teachers completed the C-TRF (Achenbach & Rescorla, 2000), which contains 100 items describing problem behaviors for children aged from 1.5 to 5 years old. Each item has three response options: “not true (as far as you know)”, “somewhat or sometimes true” and “very often or often true” corresponding to a scores between zero and two. The C-TRF

contains the following subscales: *emotional reactive* (7 items), *anxious/depressed* (8 items), *withdrawn* (10 items), *somatic complaints* (7 items), *attention problems* (9 items), and *aggressive behavior* (25 items). A total problem score (ranging from zero to 200) can be calculated by adding the scores across all items. In addition, two broadband scales can be calculated by adding certain subscales for *internalizing problems* (*emotional reactive*, *anxious/depressed*, *withdrawn*, and *somatic complaints*) and *externalizing problems* (*attention* and *aggressive behavior*).

Student-Teacher Relationship Scale (S-TRS)

The S-TRS (Pianta, 2001) is a teacher-report form developed to measure teachers' perception of their relationship with a child or student. This scale is extensively used in preschool and school research and contains *closeness*, *conflict*, and *dependency* subscales. The response option for each item range from 1 ("definitely does not apply") to 5 ("definitely applies"). In the current thesis, only the *closeness* (11 items) and *conflict* (12 items) subscales were used, as the meaning and interpretation of the *dependency* scale may be subject to cultural differences, and it has exhibited some factorial issues (Solheim et al., 2012). A total score for each subscale was obtained by summing the individual items, where higher scores on the *closeness* subscale indicates a higher degree of warmth in the relationship and higher scores on the *conflict* subscale indicates a higher degree of problematic interactions in the relationship.

2.3 Data analyses

Data analyses in studies 1 and 2

(described more thoroughly in the papers)

Following the instructions in the manual, children with a score at or above the 90th percentile on the C-TRF's Total Problem, Internalizing, or Externalizing scales were defined to be at elevated risk for mental health problems in these respective domains. In addition, children in the top 2% on at least one subscale (except somatic complaints), but who were not rated in the clinical range (90th percentile) on any of the three abovementioned broadband C-TRF scales, were also considered to be at elevated risk. For the Total Problem scale, the top 2% on any subscale (excluding somatic complaints) was included in the clinical range, while for the Internalizing and Externalizing scale only the top 2% on the corresponding subscales was included. This procedure was performed separately for boys and girls, because norms are provided separately for each, yielding gender specific cutoff values. As the C-TRF has demonstrated very small age effects, but significant gender effects (although small in magnitude) (Achenbach & Rescorla, 2000; Drugli & Stensen, 2019; Kristensen et al., 2010; Rescorla, 2005), only gender specific cutoff values were applied.

In study 1, an independent sample t-test was performed to investigate group differences for preschool teachers' nominations based on children's gender and age (1-2 years old vs. 3-6 years old). Following this, we calculated the sensitivity, specificity, false positive rates and false negative rates. In addition, the positive predictive value (PPV) and negative predictive value (NPV) at the sample prevalence of the target condition were calculated. As the organizational structure of Norwegian childcare centers are divided into younger and older preschool children, calculations were performed separately for each age group (ages 1-2 and 3-6), for each gender, and for the overall sample.

In study 2, an independent sample t-test was carried out to test for group differences on children's age, gender, internalizing and externalizing problems, as well as relational quality in the form of closeness and conflict, between the true negatives and the false positives. This was followed by a two-level (children nested within preschool teachers) binominal logistic regression analyses to investigate covariates of group membership for the false positives compared to the true negatives.

Data analyses in study 3

(described more thoroughly in the papers)

Spearman's correlation was used to evaluate the criterion-reference validity of the ASQ:SE by using the C-TRF as the comparator. The sensitivity, specificity, PPV and NPV were calculated by receiver operating characteristic (ROC) analysis for the ASQ:SE using the criterion of a score at or above the 90th percentile for the C-TRF total problem score. The PPV and NPV were calculated at the sample prevalence of 10%. To identify an "optimal" cutoff value for each form of the ASQ:SE, a criterion of specificity of at least 90% was established.

3 Results overview

3.1 Study 1: Childcare Providers' Nominations of Preschool Children at Risk for Mental Health Problems: Does it Discriminate Well Compared to the Caregiver-Teacher Report Form (C-TRF)?

In study 1, we investigated the accuracy of preschool teachers' (referred to as childcare providers in the paper) nominations in discriminating children in the clinical range of the C-TRF from those in the normal range. In other words, were their nominations of whether children were developmentally at risk or not reflected in their scores on the C-TRF. Based on 169 preschool teachers' reports of 1430 children (49% girls) aged 1-6 years old, the accuracy of nominations was explored for the overall sample and different age groups (1-2 years versus 3-6 years) and genders. This was done for nominations against symptoms of mental health problems in general (Total Problems on the C-TRF) and more specific sets of symptoms (internalizing or externalizing problems). Our results indicate that preschool teachers' nominations were relatively accurate according to the C-TRF scores. However, their nominations also led to a considerable number of false positive cases (i.e., concerns about children who in fact scored in the normal range on the C-TRF).

Considering the findings in more detail, the nomination accuracy varied as a function of children's age, gender, and type of symptom scale of the C-TRF used. Significantly more boys compared to girls were nominated by the preschool teachers ($p=.018$), as well as significantly more children from the oldest age group (3-6 years old) compared to the youngest age group (1-2 years old) ($p<.001$). Preschool teachers demonstrated a higher accuracy for boys than for girls and for older children than for younger children. More specifically, Table 2 shows that preschool teachers correctly nominated 57% of the 1- to 2-year-old children and 81% of the 3-6-year-old children in the clinical range of the C-TRF

Total Problem scale. Nominations of the youngest girls exhibited a sensitivity of 44%, while for the youngest boys the sensitivity was 71%. For the 3 to 6-years-old group, the sensitivity exceeded 78% for both genders. However, the highest false positive rate was also found for the children aged 3-6 years old, with approximately every other nomination or more being a false positive. The specificity estimates ranged from 86%-95% and the false negative rates were below 7%. The positive predictive values (PPV) ranged from 43% to 59%, while the negative predictive value (NPV) ranged from 93% to 98%.

Table 2. Accuracy of childcare providers' nominations below or in the clinical range on the Caregiver-Teacher Report Form's Total Problems scale, reported as percentages, proportions, and 95% confidence intervals for the percentages, for each gender, and total within each age group (N=1430)

Age group	Gender	Sensitivity	Specificity	Rate of FP	Rate of FN	PPV	NPV
1-2 years	Boys	71% (17/24)	94% (190/202)	41% (12/29)	4% (7/197)	59% (17/29)	96% (190/197)
		[51%, 85%]	[90%, 97%]	[26%, 59%]	[2%, 7%]	[41%, 74%]	[93%, 98%]
	Girls	44% (11/25)	95% (186/195)	45% (9/20)	7% (14/200)	55% (11/20)	93% (186/200)
		[27%, 63%]	[91%, 98%]	[26%, 66%]	[4%, 11%]	[34%, 74%]	[86%, 96%]
Total	57% (28/49)	95% (376/397)	43% (21/49)	5% (21/397)	57% (28/49)	95% (376/397)	
	[43%, 70%]	[92%, 97%]	[30%, 57%]	[3%, 8%]	[43%, 70%]	[92%, 97%]	
3-6 years	Boys	85% (49/58)	86% (379/443)	57% (64/113)	2% (9/388)	43% (49/113)	98% (379/388)
		[73%, 92%]	[82%, 89%]	[47%, 65%]	[1%, 4%]	[35%, 53%]	[96%, 99%]
	Girls	78% (42/54)	90% (387/429)	50% (42/84)	3% (12/399)	50% (42/84)	97% (387/399)
		[65%, 87%]	[87%, 93%]	[40%, 60%]	[2%, 5%]	[40%, 60%]	[95%, 98%]
Total	81% (91/112)	88% (766/872)	54% (106/197)	3% (21/787)	46% (91/197)	97% (766/787)	
	[73%, 87%]	[86%, 90%]	[47%, 61%]	[2%, 4%]	[39%, 53%]	[96%, 98%]	

Note. FP= false positives, FN= false negatives, PPV and NPV= positive and negative predictive value at sample prevalence

When separating the nomination by types of symptoms of mental health problems (internalizing or externalizing problems) as seen in Table 3, the lowest sensitivity estimate for internalizing problems was found for girls aged 1-2 years old (53%), and the highest was found for boys aged 3-6 years old (83%). Concerns about internalizing problems for boys

aged 3-6 years old also had the highest rate of false positives (70%). The highest specificity estimate was found for girls aged 1-2 years old (95%), while the lowest was observed for boys aged 3-6 years old (85%). The false negative rates ranged from 2% to 4%, the PPV ranged from 30% to 48%, and the NPV ranged from 96% to 98%.

Comparing concerns against externalizing problems with the scores on the Externalizing Problem scale of the C-TRF yielded a similar sensitivity estimate, with the highest sensitivity observed for boys aged 3-6 years old (83%) and the lowest sensitivity observed for girls aged 1-2 years old (24%). The lowest false positive rate was found for boys aged 1-2 years old (52%) and the highest for girls aged 1-2 years old (75%). Specificity estimates ranged from 84% (boys aged 3-6 years old) to 93% (boys aged 1-2 years old) and the highest false negative rate was found for girls aged 1-2 years old (8%). The PPV ranged from 25% to 48%, while the NPV was above 90% across all ages and genders.

Table 3. Accuracy of childcare providers' nominations below or in the clinical range on the Caregiver-Teacher Report Form's Internalizing and Externalizing scales, reported as percentages, proportions, and 95% confidence intervals for the percentages, for each gender, and total within each age group (N=1430)

Age group	Gender	Scale	Sensitivity	Specificity	Rate of FP	Rate of FN	PPV	NPV
1-2 years	Boys	INT	64% (14/22) [43%, 80%]	93% (189/204) [88%, 95%]	52% (15/29) [34%, 69%]	4% (8/197) [2%, 8%]	48% (14/29) [31%, 66%]	96% (189/197) [92%, 98%]
		EXT	64% (14/22) [43%, 80%]	93% (189/204) [88%, 95%]	52% (15/29) [34%, 69%]	4% (8/197) [2%, 8%]	48% (14/29) [31%, 66%]	96% (189/197) [92%, 98%]
	Girls	INT	53% (9/17) [31%, 74%]	95% (192/203) [91%, 97%]	55% (11/20) [34%, 74%]	4% (8/200) [2%, 8%]	45% (9/20) [26%, 66%]	96% (192/200) [92%, 98%]
		EXT	24% (5/21) [11%, 45%]	92% (184/199) [88%, 95%]	75% (15/20) [53%, 89%]	8% (16/200) [5% to 13%]	25% (5/20) [11%, 47%]	92% (184/200) [87%, 95%]
3-6 years	Boys	INT	83% (34/41) [69%, 91%]	83% (381/460) [79%, 86%]	70% (79/113) [61%, 78%]	2% (7/388) [1%, 4%]	30% (34/113) [22%, 39%]	98% (381/388) [96%, 99%]
		EXT	83% (40/48) [70%, 91%]	84% (380/453) [80%, 87%]	65% (73/113) [55%, 73%]	2% (8/388) [1%, 4%]	35% (40/113) [27%, 45%]	98% (380/388) [96%, 99%]
	Girls	INT	74% (35/47) [60%, 85%]	89% (387/436) [85%, 90%]	58% (49/84) [48%, 68%]	3% (12/399) [2%, 5%]	42% (35/84) [32%, 52%]	97% (387/399) [95%, 98%]
		EXT	73% (29/40) [57%, 84%]	88% (388/443) [84%, 90%]	66% (55/84) [55%, 75%]	3% (11/399) [2%, 5%]	35% (29/84) [25%, 45%]	97% (388/399) [95%, 98%]

Note. INT= internalizing, EXT= externalizing, FP= false positives, FN= false negatives, PPV and NPV= positive and negative predictive value at sample prevalence

3.2 Study 2: Teacher Nominations of Preschool Children at Risk for Mental Health Problems: How False is a False Positive Nomination and What Make Teachers Concerned?

Using the same procedure as in Study 1, the aim of Study 2 was to explore the falseness of a false positive preschool teacher nomination by comparing those classified as false positives against those classified as true negatives. Findings from this study showed that children classified as false positives (preschool teachers concerned for children who were in the normal range of the C-TRF) had significantly worse scores on the C-TRF and a poorer teacher-child relationship than those classified as true negatives (absence of concerns from preschool teachers and in the normal range of the C-TRF).

More specifically, the findings from this study showed that there were significant between-group differences between the preschool teacher nominated false positive group and the true negative group, with the former including significantly more boys and older children compared to the latter. Additionally, the false positive group was evaluated more negatively by the preschool teachers in the form of receiving significantly higher scores on internalizing and externalizing problems measured with a standardized instrument and having more conflict in the teacher-child relationship than the true negative group. Likewise, for closeness in the teacher-child relationship the false positive group received significantly lower scores, indicating less warmth in the teacher-child relationship in this group than in the true negative group. Fully adjusted logistic analyses demonstrated that increases in internalizing and externalizing scores increased the odds for a false positive classification (making preschool teachers more likely to express concern about a child, but which was not confirmed by the “gold standard”) by 55% and 21% respectively. In addition, male gender decreases the odds for false positive classification by 1%, while increases in the closeness scale (reflecting a

more positive relationship) reduced the odds for false positive classification by 8%. Neither child's gender nor conflictual teacher-child relationship were significant covariates in the fully adjusted analyses.

3.3 Study 3: Screening for mental health problems in a Norwegian preschool population. A validation of the Ages and Stages Questionnaire: Social-Emotional (ASQ:SE).

In Study 3, we aimed to investigate the validity and accuracy of a standardized screening instrument, the ASQ:SE. Based on reports from preschool teachers, we examined the criterion-reference validity of the ASQ:SE by comparing the ASQ:SE scores with scores on the C-TRF. Additionally, ROC analyses were performed to investigate how well the ASQ:SE could discriminate between children in the normal and clinical range of the C-TRF. The ROC analyses also yielded optimal cutoff values with associated sensitivity and specificity estimates based on a predetermined specificity of at least 90%.

Spearman's correlations between the different forms of the ASQ:SE and the C-TRF ranged from .49 (95% CI .33 to .63 for the ASQ:SE 24) to .72 (95% CI .66 to .77 for the ASQ:SE 60). ROC analyses yielded AUC values ranging from .87 (ASQ:SE 18) to .96 (ASQ:SE 30 and ASQ:SE 60). The selected cutoff values gave sensitivity estimates ranging from 50% (for the ASQ:SE 18) to 85% (for the ASQ:SE 48 and 60). The specificity was predetermined to be at least 90%, and consequently all specificity estimates were at 90% or above. The PPV at a 10% prevalence ranged from 36% to 59%, while the NPV ranged from 94% to 98%.

4 Discussion

This thesis investigated the accuracy of preschool teachers' nominations of children at risk for developing mental health problems compared with an extensively used and well-documented "gold standard", the Caregiver-Teacher Report Form (C-TRF). These studies focused on estimating preschool teachers' ability to discriminate between children in the clinical and normal ranges on the C-TRF by asking the preschool teachers whether they were concerned about a child's development or not. As a follow-up to this study, a closer examination of the characteristics of those children classified as false positives was performed, as previous research has suggested that children classified as false positives based on various screening instruments may still be in need of some attention. Unstandardized approaches to screening for socio-emotional or mental health problems in the form of subjective or clinical judgments have previously been deemed to have unsatisfactory accuracy. Finally, this thesis sought to validate and calibrate a standardized screening instrument intended to be used among young children and to examine how the instrument performs in a Norwegian context with preschool teachers as informants. The main findings from this thesis are as follows:

Examining preschool teachers' ability to discriminate between children in the clinical and the normal range on the C-TRF through a simple nomination process revealed that they were better at identifying children with a normal level of symptoms than identifying those with a clinical level of symptoms. In other words, preschool teachers were more accurate in evaluating normal behavior than abnormal behavior. Moreover, there were some variations regarding sensitivity. Preschool teachers' discriminative accuracy was better for older preschoolers (3 to 7 years old) than younger preschoolers (1 to 2 years old) and better for boys than for girls. In some instances, the accuracy exceeded the minimum requirement for screening instrument accuracy proposed by Glascoe (2005), that is sensitivity and specificity

estimates of at least 70%. Preschool teachers' nominations also left very few false negative cases (children reported with a clinical symptom load for which preschool teachers expressed no concern). However, preschool teachers' nominations also led to about every other nomination or more being a false positive (concerns for children reported with a non-clinical symptom load), which is a considerable proportion.

When comparing the false positive cases (children to be of concern but with a normal symptom level) with the true negatives (non-nominated children with a normal symptom level), those classified as false positives had worse outcomes in the form of more mental health problems and poorer teacher-child relationship. Certain children's characteristics, such as internalizing and externalizing problems and age, increased the odds of preschool teachers expressing concern, while preschool teachers' perceptions of conflictual teacher-child relationships and the child's gender were nonsignificant covariates when examining the odds for nomination. Teacher-perceived closeness in the relationship, however, reduced the odds for preschool teachers expressing concern.

Findings from the validation study of the Ages and Stages Questionnaire: Social-Emotional (ASQ:SE), indicated promising validity indices when comparing the instrument with the C-TRF, as well as satisfactory overall accuracy estimates. However, this does not apply for the youngest children (two years old and younger), for whom the screening instrument failed to meet the recommended benchmark and should therefore be applied with caution, if at all. Comparing the optimal cutoff values identified through the receiver operating characteristics (ROC) analyses to those reported in the ASQ:SE manual, the results suggest a general decrease of the cutoff values to optimize the identification rate (given the predetermined specificity of 90%), with the exception of the ASQ:SE form intended for three-year-old children, for which the results suggested a slight increase in the cutoff value.

In the following section, the findings from the current thesis will be discussed. First, screening aspects regarding the identification, nonidentification, and misclassification will be considered. Then, preschool teachers as informants and their role in the screening process will be discussed. Finally, proposals for implications for the practice field and clinicians, as well as suggestions for future directions will be made.

4.1 Considering identification, non-identification, and misclassification

The findings from Study 1 and Study 3 indicate that a considerable number of children with a clinical level of symptoms of mental health problems can be identified by preschool teachers, both through simple nomination and with the use of a standardized screening instrument (the ASQ:SE). Another promising finding is the low rates of false negative cases, indicating that few children in need of follow-up assessment are missed using either procedure. However, as the individual and societal consequences of untreated mental health problems may be high, efforts to lower the rates of false negatives must be pursued.

Over-identification (false positives) is far more common, especially when relying solely on preschool teacher nominations, which may lead to unnecessary strain on the support system and lead to stigma and anxiety among children and their parents. Another source of false positive classification is performing universal screening in the normal population where the prevalence of problems is low. Because the accuracy estimates of sensitivity and specificity (and thereby the rates of false negatives and false positives) are dependent on each other and involve a trade-off, an understanding of the psychometric properties of the instrument is essential to guide the selection of appropriate instruments and determination of cutoff values, as appropriate, to achieve the proposed goals.

If the aim is to increase the identification rate of those with mental health problems (prioritizing sensitivity), a lower cutoff value than those selected in Study 3 can be applied. This would increase the identification rate, but at the cost of lower specificity and thus a higher rate of false positives. In Study 3, the cutoff values that yielded the highest sensitivity given the predetermined specificity of at least 90% were selected. In other words, in this case specificity was prioritized, rendering a false positive rate at or below 10%. The ASQ:SE forms for children older than 2.5 years old yielded high sensitivity given the specificity criterion. However, the forms for children younger than 2.5 years failed to reach the benchmark of at least 70% sensitivity recommended by Glascoe (2005). A lower cutoff value for the youngest children could have been selected to increase the identification rate with the expense of a higher rate of false positives, if deemed acceptable and ethical. However, the inability of brief screening instruments to satisfactorily identify young children with mental health problems has also been previously demonstrated (Theunissen et al., 2015; de Wolff et al., 2013). In addition, findings from a Chinese sample show a lower convergence between the ASQ:SE and criteria measures for younger preschool children compared to older preschool children. Additionally, non-significant correlations between the ASQ:SE and internalizing problems were observed, suggesting that the ASQ:SE might not be sensitive enough to identify children with internalizing problems (Xie et al., 2019). Similar findings were reported for the second edition of the ASQ:SE, which exhibits some factorial reliability issues for the forms younger than 2.5 years old. Thus, the test results of the younger age forms should be interpreted cautiously, especially for the emotionally loaded items (Chen et al., 2020).

Similar to the accuracy estimates from the ASQ:SE in Study 3, the preschool teachers in Study 1 also had a harder time identifying the youngest preschool children at risk for mental health problems through their nomination compared to older children. Some of the

possible explanations might be the transient nature of problems for infants and toddlers, the more subtle symptom expression, young children's limited verbal communication repertoire, or that preschool teachers' lack the necessary knowledge and skills to identify abnormal behavior for such young children. Although previous studies indicate few age effects for problem behaviors in preschoolers (Achenbach & Rescorla, 2000; Drugli & Stensen, 2019; Rescorla, 2005), it seems that age influences preschool teachers' perception and accuracy when nominating children at risk, as preschool teachers are more reluctant and inaccurate for younger preschool children compared to older preschool children. One might think that if normal behavior is the standard in childcare centers, preschool teachers should be well equipped to identify deviancy from normality. However, similar to pediatricians and health nurses working without standardized screening instruments (Sheldrick et al., 2011; Skovgaard et al., 2008), preschool teachers also exhibit a considerable portion of misclassifications through their nominations in the form of false positive cases, which may suggest either inaccuracy in preschool teachers perception of problems or that preschool teachers' perceive sub-clinical level of mental health problems to be of concern.

Checklist instruments

Checklist instruments, such as the ASQ:SE and the C-TRF, are capable of providing information regarding the frequency and presence or absence of behaviors. However, they fail to capture the onset, severity, and duration of symptoms. Even many of the DSM diagnostic criteria specify that symptoms need to be developmentally inappropriate, but do not provide any guidelines for the distinction between age-appropriate and inappropriate behavior (Egger & Angold, 2006). Thus, what is perceived as age-appropriate behavior is largely a subjective decision based on knowledge and experience. For example, tantrums are

quite common for children aged two years old. Unless the frequency or intensity of the tantrums is unusually high, it might well represent a normative behavioral pattern. This said, the absence of tantrums may be a cause of concern, as tantrums are a precursor of the emerging autonomy (Carter et al., 2004), which is a normative developmental pattern. While some behaviors may be considered deviant or unusual regardless of age and thus raise concerns more easily, most behaviors operate dimensionally and depend on context, making the distinction between normal and abnormal behaviors more complicated (Carter et al., 2004).

Standardized checklist instruments with empirically derived cutoff values based on the population from which it was drawn might give a good and culturally appropriate indication of clinical or risk status. However, cutoff values may seem arbitrary even though they exhibit good discriminant ability and predict referral rates. Since distress and/or impairment are required for some DSM disorders (Egger & Angold, 2006), some clinically relevant cases might be missed, as symptom load and distress or impairment do not need to be perfectly correlated. Actually, symptoms of anxiety, depression, and ADHD only have low to modest correlations with functional impairment (Gordon et al., 2006; McKnight & Kashdan, 2009; McKnight et al., 2016). Hence, some children may display a significant symptom load with minimal impairment, while others may display few symptoms with significant impairment. For example, a child at the 89th percentile of the C-TRF may display more distress or impairment than a child at the commonly used 90th percentile, making the former more likely to be a clinically relevant case than the latter. A categorical approach to identification may increase the rates of false negative cases and fail to capture and intervene for emerging psychopathology. False positive cases may still hold some clinical relevance (Glascoe, 2001; Jensen & Watanbe, 1999), as children with a subclinical level of symptoms continue to display some impairment later in life (Finsaas et al., 2018b; Shankman et al.,

2009). Thus, it is important to identify relevant clinical cases below the clinical cutoff on checklist instruments and among children not meeting diagnostic criteria, as these children may still display elevated symptom loads or functional impairment.

The false positives

In accordance with previous studies reporting poorer outcomes and more psychosocial risk for children classified as false positives compared to true negative cases (Glascoe, 2001; Jensen & Watanbe, 1999), Study 2 indicates that children classified as false positives by preschool teacher nominations also display more problem behavior and poorer teacher-child relationships as perceived by preschool teachers. It seems plausible that a portion of the children classified as false positives in Study 1 are found in what is known as the “borderline range” or “the gray area”, which usually constitutes the range between what is considered normal and clinical on checklist instruments. The comparator, or “gold standard”, used in this thesis (the C-TRF) has a borderline range between the 83rd percentile and the 90th percentile where the clinical range starts. The C-TRF manual advises that if a child is scored in the clinical range by one preschool teacher or in the borderline range by multiple preschool teachers, further evaluations should be undertaken (Achenbach & Rescorla, 2000). If we were to apply the borderline percentile as cutoff for caseness, a portion of those classified as false positives in Studies 1 and 2 would be considered true positives. Consequently, this would have increased the identification rate at the expense of lower specificity, as some false positives indeed are false positives and are developmentally within normal parameters. This will be highlighted further when discussing the choice of comparator or “gold standard” later in this Discussion. The ASQ:SE does not have a borderline range in its first edition, which we used for Study 3. However, a borderline or “monitor range” was implemented in the second

edition (Squires et al., 2015), such that a child scored in the borderline range should be monitored to ensure healthy development.

Using the ASQ:SE, it is possible to make a checkmark for “concerned” by each particular behavior. Such a checkmark confers an additional five points and should warrant a scrutiny of said concern, regardless of whether the total score surpasses the cutoff value (Squires et al., 2002). As shown in Study 2, the probability of preschool teachers becoming concerned increases as the child’s emotional and/or behavior problems increase, and children for whom preschool teachers express concern display poorer outcomes than children for whom preschool teachers have no concerns. Thus, implementing either concern checkmarks and/or borderline ranges in screening instruments seems beneficial to increase the identification rate of children at risk for mental health problems. In this thesis, we have applied a more stringent cutoff value for specific problem behavior, thereby, ensuring that more extreme cases with specific behavior problems were included in the clinical range. However, concerns may arise from a single behavior or few behaviors that preschool teachers perceive as deviant or extreme. Concerns may also arise from behaviors that preschool teachers perceive negatively affecting a child’s function level. This could make a child with low to moderate scoring on many checklist items having a higher probability of reaching an empirically derived cutoff than a child with very few extreme scores. Taking two items from the C-TRF as an example, preschool teachers may mark “very true or often true” on “*hurting animals*” and “*quickly shifts activities*”, and the former may be the source of preschool teachers’ concern, while the latter is not. Even though these two items are scored the same, one might be perceived as more clinically significant than the other. Hence, the concern may stem from a single behavior rather than the cumulative symptom load. In a study by Nøvik (2000) with children aged 4 to 16 applying the Child Behavior Checklist (CBCL), the parent-

reported counterpart to the C-TRF, the two items *clumsy* and *unhappy, sad or depressed* were found to be strong predictors of psychiatric disorder, as well as exhibiting high PPV.

Findings from Nøvik's (2000) study and the present thesis are examples that shed light on the challenges in identifying caseness in children by dimensional measures with empirically derived cutoff values. As many symptoms of psychological disorders reflect typical behaviors that change phenotypically throughout early childhood, approaches that capture the full range of behaviors relevant to psychopathology are needed. Moreover, establishing developmental norms may also capture children that do not meet diagnostic criteria, but still may be at risk for developing psychopathology (Dougherty et al, 2015). Sheldrick and colleagues (2015) have suggested moving away from the group-based accuracy estimates of sensitivity, specificity, PPV, and NPV and instead examining the individual scores in a probabilistic matter. For example, children with a very high screen positive score on a validated screening instrument are more likely to have some psychopathology than children with a very low screen negative score. In addition, children approaching an empirically derived cutoff value would also be more likely to display some psychopathology compared to low-scoring children. Yet each increase in score is associated with an increased probability of psychopathology. This probabilistic approach could indeed increase the identification rate of children who fall beneath a clinical cutoff value, but still may be at risk. However, this approach assumes that each item and score constitute the same additive probability, which Nøvik's study (2000) suggests is not the case. Thus, unless item score probabilities are weighted for clinical relevance, this approach is unlikely to eradicate problems with misclassification. Implementing both a categorical and a dimensional approach with empirically derived cutoff values to the screening procedure might be beneficial to optimize the identification rate and keep the misclassification rate as low as possible, which will be discussed further in the Clinical implications section.

4.2 “Gold standards” and accuracy estimates

How golden are “gold standards”?

Commonly, when investigating the validity or the classification accuracy of an instrument, a comparator or “gold standard” criterion is applied. The C-TRF was used in Studies 1, 2, and 3 as the “gold standard” to establish which children were in the clinical range (at or above the 90th percentile on the broad problem scales and at or above the 98th percentile on the specific problem scales) and which children were in the normal range. Based on this dichotomization, children were classified as true positives, false positives, true negatives, or false negatives, based on preschool teacher nominations (Study 1) or based on their scores on the ASQ:SE (Study 3). The C-TRF was chosen as the “gold standard” in the current thesis due to its extensive use and status in the clinical field and research as well as its well-documented psychometric properties. This said, even though “gold standards” ideally represent the true nature of conditions, this is rarely the case, especially within the mental health and psychiatric field, where cases may be more gray than black or white. As there is no definite measure for assessing clinical conditions related to childhood psychopathology (Carneiro et al., 2020), one must consider the accuracy of the “gold standard” itself to truly understand the accuracy of an instrument beyond the selection of optimal cutoff values.

A gold standard refers to cases where the true status is known, while an imperfect “gold standard” reflects cases where the true status is unknown, although it might be indicated by the best test available (Hui & Zhou, 1998). A “gold standard” is in most cases an imperfect test, as it falls short of achieving 100% accuracy. Considering its limitations and biases is important to avoid misclassifications (Gold et al., 2010). For example, the CBCL is a commonly used “gold standard” for other instruments used in the assessment of children’s mental health. However, as it cannot be tested against itself, the instrument typically is tested against structured diagnostic interviews. A review of the classification accuracy of the CBCL

reports a mean sensitivity of 71% and a mean specificity of 86% against structured clinical interview when applying the clinical 90th percentile cutoff (Lavigne et al., 2016b), which is barely above the 70% sensitivity recommendation proposed by Glascoe (2005). If the CBCL accuracy estimates represent the practical upper limit of what to expect from screening instrument tested against the CBCL, the screening instrument in question does indeed need to excel to be deemed satisfactory. This said, even the “gold standard” for other “gold standards”, structured diagnostic interviews, does not show a perfect 100% sensitivity and specificity. In a study investigating the agreement between three diagnostic interviews, namely the Child and Adolescent Psychiatric Assessment (CAPA), the Diagnostic Interview Schedule for Children (DISC), and the Development and Well-being Assessment (DAWBA), it was found that these interviews varied in agreement, and thus varied in caseness, with the most accurate example exhibiting approximately 85% sensitivity and specificity (Angold et al., 2012). Consequently, whether or not a diagnosis is given may to some degree depend on which diagnostic interviews are administered.

When using the same informant for a screening instrument and the “gold standard”, an inflated correlation may be observed. The higher the correlation is, the more likely the screener and an imperfect “gold standard” are to make the same misclassifications. Moreover, any response biases will influence the results of both test and inflate the correlation, regardless of the child’s psychopathology (Sheldrick et al., 2015). One such response bias may be the teacher-perceived relationship with target child, which we will examine in the section *Preschool teachers as informants*. In Studies 1, 2, and 3, preschool teachers were the sole informants for both the test and the “gold standard”. As seen in Study 3, the correlation between the total scores of the ASQ:SE and the C-TRF ranges from .49 to .72, being generally lower for younger preschool children compared to older preschool children. Usually, higher correlations are interpreted as indices for criterion-reference validity between

two instruments (Fayers & Machin, 2007). Ironically, it can also be the source of a higher misclassification rate when using an imperfect “gold standard”. Consequently, even though findings from Study 3 report higher correlations and better accuracies for the older ASQ:SE forms, the older forms are also more susceptible to errors in the “gold standard”.

Additionally, the accuracy estimates reported in Studies 1 and 3 probably represent an upper limit for what to expect from nominations and a brief screening instrument. They might also be overly optimistic, as the estimates in many cases surpassed what is found as accuracy estimates for the parent-reported version of the same “gold standard” (Lavigne et al., 2016b). One might argue that it is unethical to conduct screening when the accuracy evidence is based on an imperfect “gold standard”, as the results may not reflect the true risk status for children. However, it may be equally unethical to refrain from screening, as this may lead to failure to provide help to children who need it.

Predictive values and the prevalence

Sensitivity and specificity are important estimates when evaluating the accuracy of an instrument, but they are of less use for clinicians and professionals in estimating the probability of the presence or absence of a targeted condition. PPVs and NPVs are the probability estimations of the presence or absence of the targeted condition for an individual once the test results are known (Akobeng, 2007). In other words, the PPV reflects the probability that a child with a positive screen has mental health problems, while the NPV reflects the probability of a child with a negative screen being within normal parameters. In contrast to sensitivity and specificity, these predictive values are a function of prevalence. Hence, these values will change as the prevalence of the targeted condition varies.

Based on the findings from Studies 1 and 3, preschool teachers using the ASQ:SE or their subjective judgment will correctly rule out children without a clinical mental health level of symptoms more than 9 out of 10 times, if the prevalence is approximately 10%. The PPV on the other hand, has slight variations of approximately 50%, indicating that approximately half of the children who are screened as positive using the ASQ:SE or preschool teacher nominations would display a clinical level of symptoms, while the other half would not. This said, the high rates of false positives obtained from preschool teacher nominations may have inflated the sensitivity, and thus the PPV, as the PPV and the NPV are a function of sensitivity, specificity, and prevalence. For example, if the preschool teachers had nominated all of the children, every child with a clinical level of symptoms would be identified, but at the expense of an extremely high false positive rate of approximately 90%. As the predictive values are extremely sensitive to variations in prevalence, it would be wrong to generalize the predictive values obtained in one population to another population, where the prevalence could be different (Akobeng, 2007). To exemplify the effect of prevalence, we can examine the findings from Study 3. In contrast to predictive values, the area under curve (AUC) reflects the instruments' intrinsic ability to discriminate between cases and noncases, regardless of population prevalence. The ASQ:SE demonstrates an AUC ranging from .89 to .96, which would represent an excellent to outstanding ability to discriminate between cases and noncases based on Hosmer and Lemeshow's (2000) strength of discrimination. The PPV on the other hand is considerably lower, ranging from 36% to 59%. If the next step was to refer those children with a positive screen to the special health service, they would have wasted about half of their clinical time assessing children with no clinical symptom load, time that ideally could be used on assessing and treating children who were actually in need of help. Screening in populations with a low prevalence of problems usually leads to a high rate of false positives, and as the prevalence decreases, the PPV and

the instruments' ability to identify true positives decrease (Lavigne et al., 2016a). The difficulty with accuracy is not related to statistics but the need to decide whether or not a test is clinically valuable (Altman, 1991). In the *Clinical implications* section, we will examine one approach to address the low PPVs and high rates of false positives commonly found when screening in the normal population, namely *sequential screening*.

4.3 Preschool teachers as informants

The amount of time that preschool teachers spend with numerous children should put them in an excellent position to discriminate deviancy from normality. However, the low to modest agreement between parents and preschool teachers are usually explained by the different contexts in which the informants observe children, different perception of what constitutes problem behaviors, and different expectancies to children's competency (Achenbach & Rescorla, 2000; Carneiro et al., 2020; Gross et al., 2004; Strickland et al., 2012). The agreement discrepancy may also explain why in three out of four times parents and preschool teachers disagree on children with a high severity of problems (Berg-Nielsen et al., 2012). Moreover, preschool teachers tend to perform worse than parents in discriminating between children with and without diagnosis (Sveen et al., 2013). This said, if a diagnosis or "gold standard" structured diagnostic interviews are based on parent-reports, parents would have a correlational advantage compared to preschool teachers. As early emotional and behavioral problems that manifest across multiple contexts are often more stable and serious, it is important to take a multi-informant approach to the identification of children at risk (Egger & Emde, 2011; Kerr et al., 2007).

Previous research has highlighted the lower rating scores provided by preschool teachers on children's problem behaviors than those of parents, especially for internalizing

problems. This has been explained as a lack of knowledge or that preschool teachers have a more normative perception of internalizing behaviors in preschool children, making them more reluctant to declare such behaviors as problematic (Berg-Nielsen et al., 2012; Heiervang et al., 2008). In contrast to previous studies with teachers in primary school (Loades & Mastroyannopoulou, 2010), Study 2 suggests that internalizing rather than externalizing problems make preschool teachers more likely to express concern (OR 1.55 vs. 1.21). Additionally, findings from Study 1 indicate that preschool teachers are somewhat more accurate in identifying externalizing problems compared to internalizing problems, which corresponds well with previous research (Dwyer et al., 2006; Heiervang et al., 2008). As shown in Study 1, preschool teachers are capable of identifying a considerable portion of children using only their subjective judgment to express whether they are concerned or not. However, preschool teachers are better at ruling out children within normal parameters than identifying children with a clinical level of symptoms.

It seems that it is not the type of problem behaviors per se that leads to inaccuracy, but rather the interaction of children's age and symptom expression. Preschool teachers have an especially hard time identifying the youngest preschool children (1-2 years old) with a clinical level of symptoms. Rather than being reluctant to declare internalizing behaviors as problematic, preschool teachers seem more reluctant to express concern about the youngest children. Findings from Study 2 show that the odds of preschool teachers expressing concern increases by 46% for each year the child ages. This is also reflected in Study 1, where almost twice as many children aged 3-6 years old were nominated compared to children aged 1-2 years old (20% vs 11% respectively). Even when applying the ASQ:SE (Study 3), the identification of children aged 1.5 to 2 years old with a clinical level of symptoms was unsatisfactory. It has been reported that preschool teachers perceive problem behaviors in older preschool children as more distressful for the child compared to younger preschool

children's problem behaviors (Gustafsson et al., 2017). Hence, knowledge of what constitutes age-appropriate and abnormal behavior may be the source of inaccuracy when examining the youngest preschoolers. Additionally, preschool teachers perceive older preschool children's problems to have a more negative impact on social interaction than younger preschool children's problems (Gustafsson et al., 2017). One explanation for this may be that children's social interaction with peers and adults increase in frequency and complexity as the child matures, which makes problem behaviors related to social interaction more observable for preschool teachers. This may also explain, in accordance with Study 1 and 2 in the present thesis, why preschool teachers are more likely to state concerns for older preschool children. Additionally, as children develop their language skills, they may more easily communicate to preschool teachers about how they feel. Consequently, identification of the youngest preschool children may require additional efforts in equipping preschool teachers with more knowledge relevant to identification of subtle symptom expression, as well as training preschool teachers in the use of sensitive observational methods and/or sensitive (or more thorough) screening instruments to increase the identification rate for this age group. Moreover, the younger the child, the more important is information from parents about a child's development and symptoms (Skovgaard et al., 2004), as emotional and behavioral problems in very young children are often closely related to their relationship to their parents (Szaniecki & Barnes, 2016).

As mentioned earlier, response biases may influence the accuracy and validity of checklist instruments, as informants would rate children's problem behaviors lower or higher than they normally would do, regardless of the children's actual symptom load. Parental depression has been shown to influence how parents rate children's problem behavior (Harvey et al., 2013; Salomonsson & Sleded, 2010), as does a conflictual teacher-child relationship (Berg-Nielsen et al., 2012). Positive reciprocal associations have been reported

between children's problem behaviors and conflictual teacher-child relationships, but not between problem behaviors and teacher-perceived closeness in the teacher-child relationship (Skalická et al., 2015; Zhang & Sun, 2011). This has been explained by teachers' perception of conflict being more child-driven, where the perception of closeness is more teacher-driven or driven by a combination of teacher and child characteristics (Choi & Dobbs-Oates, 2016; Silver et al, 2005).

Study 2 in the present thesis suggests that preschool teachers' concerns are child-driven rather than relational, as children's age and problem behaviors significantly increase the odds for preschool teachers expressing concern, while teacher-perceived conflicts in the relationship do not. Teacher-perceived closeness on the other hand, reduces the odds of preschool teachers expressing concern about children within normal parameters, indicating that preschool teachers' relational proximity is important when evaluating children. Another finding from Study 2 is the seemingly internal thresholds for preschool teachers to express concern. It seems that preschool teachers have a higher threshold for expressing concern about younger children, boys, and externalizing problems compared to older children, girls, and internalizing problems. In other words, they tolerate more from younger children, boys, and children with externalizing problems before expressing concern. This could be problematic, especially for the youngest children, as this is the age group where preschool teachers display the poorest accuracy. One explanation may be that preschool teachers perceive that problems are transient for the youngest children, giving development enough time to normalize before school entry. They may also perceive externalizing problems as more normative for preschool children, thus making them more likely to express concern when facing children with internalizing problems. It has also been reported that girls displaying externalizing symptoms and boys displaying internalizing symptoms are more likely to receive special educational support than it would be the other way around (boys with

externalizing symptoms and girls with internalizing symptoms) (Lekhal, 2020), indicating that preschool teachers attention may be drawn to what they perceive as non-stereotypical gender behavior. In a study of primary school teachers' perception of students' behavior (Liljequist & Renk, 2007), teachers found externalizing problems to be more bothersome than internalizing problems and attributed more responsibility to students displaying externalizing problems than those displaying internalizing problems. However, teachers' efficacy contributed significantly to the prediction of how bothersome teachers perceived students' internalizing problems compared to externalizing problems (i.e., teachers' perceived inability to induce changes in displaying internalizing problems). If preschool teachers are more familiar with addressing children displaying externalizing problems rather than internalizing problems, they may feel more incapable of dealing with the latter, thus making them more likely to express concern when actually reporting internalizing problems.

4.4 Strengths and limitations

The thesis addresses an important knowledge gap in the identification accuracy literature, as previous studies have mainly focused on primary school teachers rather than preschool teachers. As preschool teachers play an important role in the identification and referral of children with mental health problems, there is a necessity to understand their accuracy in performing these tasks beyond the mere symptom ratings on checklist tools and comparisons with other informants. This thesis has shed light on preschool teachers' accuracy when working with and without standardized instruments, which contributes important information regarding screening in the preschool arena. An additional strength is the large sample from a normal population included the full age range of preschool children, as well as a relatively high participation rate. The PPVs and NPVs reported from the studies in this thesis could easily be calculated with other prevalence estimates, which could provide an indication of the

accuracy in other cultures given the same sensitivity and specificity. Another strength is the application of a well-documented and extensively used “gold standard”, the C-TRF, when information from clinical interviews is unavailable.

There are several limitations as well. As most participating children were of Norwegian origin, the generalizability to more heterogeneous populations may be limited. Additionally, it was specified that the preschool teacher who knew the child best should be the informant. At least in Norway, a considerable amount of the preschool teachers’ time is taken up in meeting activities and other office-related tasks. Thus, it is unclear whether the preschool teacher who participated was the one who knew the child best (rather than an other preschool staff member). It would also have improved the confidence in the findings if the sample of the youngest preschool children had been larger, as misclassifications for this age group would yield a larger impact compared to older preschoolers.

Another limitation concerns the “gold standard”. As discussed previously, even though the C-TRF is considered a “gold standard”, it is in essence an imperfect “gold standard”. Coupled together with the increased correlation when using only one informant, the accuracy estimates may be rather optimistic in some instances and represent the upper limit of what to expect. Thus, overconfidence in preschool teachers’ ability to identify children with mental health problems without the use of validated screening instruments should be avoided, as there remains a considerable number of misclassifications that need to be remedied. In addition, a very low portion of children in the present sample are under the lower age limit (18 months) of the comparator, thus for these children the comparator may be less than optimal. This said, as the present sample has demonstrated few age effects (Drugli & Stensen, 2019), the effects of an age-inappropriate comparator for a low portion of children may not pose a problem for the accuracy estimates. Moreover, it is quite common to use the C-TRF or the CBCL as one of the comparators when investigating screening instruments

psychometric properties, even for children younger than 18 months (e.g., Briggs-Gowan et al., 2004; Karabekiroglu et al., 2010; Squires et al., 2002; de Wolff et al., 2013). Another limitation regarding the lack of diagnostic “gold standard” is the uncertainty of the actual sample prevalence of problems. However, as diagnostic interviews in large studies may be impractical or too resource demanding, applying the best information available is a commonly used choice. Finally, even though the clinical range in our studies constitutes the children with most symptoms rated by preschool teachers, we have no information on children’s function level, which would be a beneficial parameter to establish caseness.

4.5 Clinical implications

The present thesis has investigated preschool teachers’ role in identifying children at risk for mental health problems, both by unstandardized means (simple nomination) and standardized means (the ASQ:SE). Their recognition of children’s mental health problems is of great importance, as preschool teachers and parents constitutes children’s first line of defense in the early identification of problems, getting help in a timely manner, and thus increasing the probability of healthy development. As misclassifications most likely will continue to be a part of screening in the foreseeable future, debate regarding the appropriateness and ethical issues of screening will persist, especially for children as young as preschoolers.

Consequently, bringing available information about the accuracies of screening methods would be helpful when identifying children in need of help and ruling out those within normal developmental parameters.

Findings from the present thesis indicate that preschool teachers can identify a considerable portion of children with a clinical level of symptoms by simple nomination, especially for the oldest preschoolers. However, the high rates of false positives suggest that

preschool teachers overidentify, which is common when screening in the normal population where the prevalence of mental health problems usually is low. The finding that children classified as false positives by nominations have more mental health and relational problems than those not nominated supports the notion that preschool teachers' concerns should be taken seriously and not discarded. Those concerns should be scrutinized in collaboration with parents before referring to the Educational and Psychological Counselling Service if deemed necessary. In other words, among many of the children preschool teachers have concerns are children for which preschool teachers perceive to have a clinical or elevated level of symptoms. It may be that most false positive cases do not need clinical attention, but these children may need some minor help and support to ensure healthy development. These findings coupled with the Pyramid model framework proposed by Fox and colleagues (2009) may potentially improve the identification rate of children in need of clinical or sub-clinical support, as the first level of the Pyramid model emphasize universal screening and continuous monitoring of children with problem behaviors, as well as the collaboration between preschool teachers, parents, and other mental health professionals to provide support for those children in need in a timely matter.

Preschool teachers are also highly accurate in ruling out the children with normal parameters such that very few children with a clinical level of symptoms are missed. This said, behind the low rate of false negatives are individual children who preschool teachers have rated with a clinical level of symptoms, yet about whom they do not express any developmentally concerns. As long as a high rate of false positives is dealt with appropriately with a support system that has the capacity and resources to handle the extra workload, false positive cases may be perceived as more tolerable than false negatives.

Similar to preschool teacher nominations, the ASQ:SE performs worse when identifying younger preschoolers with a clinical level of symptoms compared to older

preschoolers. Thus, relying solely on preschool teachers' concerns and the younger age forms of the ASQ:SE may not be advisable due to the questionable accuracy. However, the ASQ:SE, in contrast to preschool teacher nominations, produces few false positives. Additionally, the older age ASQ:SE forms (forms for children at 2.5 years and older) have high sensitivity and specificity. Thus, if preschool teacher nominations are regarded as a universal pre-screener rather than a screener per se, preschool teachers' concerns can guide attention towards the children for whom the preschool teachers are concerned, thus leading to a more targeted screening. According to Fox and colleagues Pyramid Model as a Response to Intervention (2009), a positive screening test should be followed by a tailored intervention that should be evaluated after a while before the child is re-screened. Consequently, this procedure can monitor children's development as well as give an indication of the intervention effect. Such *sequential screening* may also reduce the high rates of false positives associated with screening in populations with low base rates, as the first stage sends the positive cases to the next stage, which would yield a higher prevalence and consequently a higher PPV for the latter stage (Lavigne et al., 2016a). The drawback of such an approach, however, is an increased rate of false negatives. Thus, to gain certainty that positive cases do indeed have a clinical level of symptoms and would benefit from follow-up assessment and eventual intervention, one must accept that more children in need of help would be missed. All these aspects would be important to consider before implementing a screening procedure.

If the ASQ:SE confirms a preschool teacher's concern, a follow-up assessment should be conducted. Moreover, a preschool teacher's concern should be scrutinized regardless of the result from the second stage of screening. However, the children with a clinical level of symptoms for whom preschool teachers had no concern would never make it to the second stage in a sequential screening procedure, and thus remaining unidentified. This emphasizes the importance of including more than one informant in the screening procedure, as preschool

teachers only offer one perspective from a childcare center setting. As previous research has indicated that parental concern can increase the pediatrician identification rate significantly (Glascoe, 1997; Glascoe, 2003; Glascoe & Marks, 2011), a combination of parent and preschool teachers concerns as pre-screener reporters seems to be a promising approach to increase the identification rate and reduce the misclassification. This would be the case especially when positive pre-screenings are followed by the use of psychometrically valid screening instrument. For the youngest preschoolers, however, it seems that additional efforts are needed. This could include increasing preschool teachers' knowledge of abnormal behavior for this age group, training them in developmentally sensitive observation methods, and/or providing them with validated developmentally sensitive checklist instruments, which may need to be more thorough than a brief screener.

Findings from this thesis may also be of value to the Educational and Psychological Counselling Service and special health service. Knowledge of preschool teachers' strengths and limitations in identifying children with more or less mental health problems may lead to more accurate referrals, and consequently, the administration of appropriate interventions if deemed necessary. The Educational and Psychological Counselling Service may also recommend or perform observation methods suitable for the youngest preschool children as part of the assessment, as this age group seem to need additional efforts beside just nominations and universal screening instruments to improve the identification rate of children with mental health problems.

4.6 Conclusions and suggestions for future research

When applying an unstandardized screener (or pre-screener), such as preschool teachers' nominations of children at risk for mental health problems, and a validated standardized screening instrument with optimized cutoff values for the targeted population, such as the ASQ:SE, both approaches can identify a considerable portion of children with a clinical level of symptoms and rule out those within normal parameters. However, similar to previous research (Sheldrick et al., 2011; Skovgaard et al., 2008), working without standardized instruments like the nomination method yields a high rate of misclassifications, especially in the form of false positive cases. As preschool teachers' concerns often indicate either a clinical or elevated level of symptoms for children, using their concerns as a pre-screener followed by a psychometrically valid screening instrument to confirm or disconfirm the initial concern might provide a promising framework for screening. This framework would also be a promising approach in a sequential screening procedure, as it would decrease the rate of false positives and increase the certainty that true positive cases are identified (at the cost of more false negatives). Ultimately, to reach a given screening goal, every aspect of the accuracy of a screening procedure needs to be discussed, as it involves deciding and prioritizing the trade-offs between identification, nonidentification, and misclassification. Future efforts should provide preschool teachers with enough knowledge and observational skills to discriminate between developmentally abnormal behaviors, especially for the youngest preschoolers, as well as provide them with developmentally sensitive checklist instruments to increase the identification rate and reduce the misclassification rate. In addition, such knowledge and instruments may prompt preschool teachers to act as soon as problems emerge and are recognized rather than waiting to see which course the development takes.

Future research

Future research should investigate the psychometric properties of other standardized screening instruments intended for preschool children, preferably in a Norwegian or Scandinavian context, as there is a need for an alternative to the youngest ASQ:SE forms (younger than 2.5 years old). For instance, the Brief Infant-Toddler Social and Emotional Assessment (BITSEA) (Briggs-Gowan et al., 2004), which is a relative short screener intended to be used with children aged 12 to 36 months, has demonstrated promising psychometric properties in several countries (Haapsamo et al., 2009; Karabekiroglu et al., 2009; Karabekiroglu et al., 2010; Kruizinga et al., 2012 ; Kruizinga et al., 2013; Kruizinga et al., 2015) and could be a viable alternative to the ASQ:SE. It would also be interesting to investigate how preschool teachers' perception of children's function level/impairment and distress influences which children they express concern about, as well as how it influences the way in which preschool teachers rate checklist instruments. It may be that preschool teachers are more accurate in evaluating a child's function level rather than recognizing symptoms of mental health problems. It would also be interesting to investigate whether preschool teachers' characteristics, such as age, gender, and years of experience, influence their perception and accuracy. Another important focus for future research is investigating whether or not *Differential Item Functioning* (DIFs) may be in play for standardized instruments (e.g., increased probability for some sub-groups to success/fail on a given criterion compared to others). This is especially important for instruments applied in decision-making processes, as DIFs may make some sub-groups of children more likely to receive follow-up assessment or referral, while overlooking the problems of others. Last, it would be interesting to investigate whether increased identification rates by using both preschool teacher and parent reports along with an informant unbiased "gold standard" or third-party expert observer would actually lead to more referrals and service use, as the end

goal would be to help children with their mental health problems and improve their quality of life.

5 References

- Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA Preschool Forms & Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.
- Akobeng, A. K. (2007). Understanding diagnostic tests 1: sensitivity, specificity, and predictive values. *Acta Paediatrica*, 96: 338-341
- Altman, D. G. (1991). *Practical statistics for medical research*. London: Chapman and Hall
- American Academy of Child and Adolescent Psychiatrists (2001). Psychiatric care of children in the foster care system. Retrieved from: <http://www.aacap.org/publications/policy/ps45.htm>.
- Angold, A., Costello, E. J., Farmer, E. M. Z., Burns, B. J., & Erkanli, A. (1999). Impaired but Undiagnosed. *Journal of the American Academy of Child & Adolescent Psychiatry*, 38: 129-137
- Angold, A., Erkanli, A., Copeland, W., Goodman, R., Fisher, P. W., & Costello, E. J. (2012). Psychiatric Diagnostic Interviews for Children and Adolescents: A Comparative Study. *Journal of the American Academy of Child & Adolescent Psychiatry*, 51: 506-517
- Anuniação, L., Squires, J., Clifford, J., & Landeira-Fernandez, J. (2019). Confirmatory analysis and normative tables for the Brazilian Ages and Stages Questionnaires: Social-Emotional. *Child: Care, Health, and Development*, 1-7. DOI: 10.1111/cch.12649
- ASEBA.org (2020). Achenbachs System of Empirically Based Assessment. Retrieved from: <https://aseba.org/>
- Axelrad, M. E., Butler, A. M., Dempsey, J., & Chapman, S. G. (2013). Treatment Effectiveness of a Brief Behavioral Intervention for Preschool Disruptive Behavior. *Journal of Clinical Psychology in Medical Settings*, 20: 323-332
- Basten, M., Tiemer, H., Althoff, R. R., van de Schoot, R., Jaddoe, V. W. V., Hofman, A., ..., & van der Ende, J. (2016). The Stability of Problem Behavior Across the Preschool years: An Empirical Approach in the General Population. *Journal of Abnormal Child Psychology*, 44: 393-404
- Belfer, M.L. (2008). Child and adolescent mental disorders: The magnitude of the problem across the globe. *Journal of Child Psychology and Psychiatry*, 49: 226-236

Berg-Nielsen, T.S., Solheim, E., Belsky, J., & Wichstrøm, L. (2012). Preschoolers' Psychosocial Problems: In the Eyes of the Beholder? Adding Teacher Characteristics as Determinants of Discrepant Parent-Teacher Reports. *Child Psychiatry and Human Development*. 43: 393-413

Berkhout, L., Hoekman, J., & Goorhuis-Brouwer, S. M. (2012). Psychosocial health of 4-year-old boys and girls, as observed by parents and teachers. *European Journal of Developmental Psychology*. 9: 500-507

Beyer, T., Postert, C., Müller, J. M., & Furniss, T. (2012). Prognosis and Continuity of Child Mental Health Problems from Preschool to Primary School: Results of a Four-Year Longitudinal Study. *Child Psychiatry and Human Development*. 43: 533-543

Bilancia, S. D. & Rescorla, L. (2010). Stability of Behavioral and Emotional Problems Over 6 Years in Children Ages 4 to 5 or 6 to 7 at Time I. *Journal of Emotional and Behavioral Disorders*. 18: 149-161

Bowlby, J. (1979). *The making and breaking of affectional bonds*. London: Tavistock

Briggs-Gowan, M. J., Carter, A. S., Irwin, J. R., Wachtel, K., & Cicchetti, D. V. (2004). The Brief Infant-Toddler Social and Emotional Assessment: Screening for Social-Emotional Problems and Delays in Competence. *Journal of Pediatric Psychology*. 29: 143-155

Briggs-Gowan, M. J., Carter, A. S., Bosson-Heenan, J., Guyer, A. E., & Horwitz, S. M. (2006). Are Infant-Toddler Social-Emotional and Behavioral Problems Transient? *American Academy of Child and Adolescent Psychiatry*. 45: 849-855

Briggs-Gowan, M. J. & Carter, A. S. (2008). Social-Emotional Screening Status in Early Childhood Predicts Elementary School Outcomes. *Pediatrics*. 121: 957-962

Brotman, L. M., Gouley, K. K., Huang, K., Rosenfelt, A., O'Neal, C., Klein, R. G., & Shrout, P. (2008). Preventive Intervention for Preschoolers at High Risk for Antisocial Behavior: Long-Term Effects on Child Physical Aggression and Parenting Practices. *Journal of Clinical Child & Adolescent Psychiatry*. 37: 386-396

Buffered, S. J., Dougherty, L. R., Carlson, G. A., Rose, S., & Klein, D. N. (2012). Psychiatric disorders in preschoolers: continuity from ages 3 to 6. *The American Journal of Psychiatry*. 169: 1157-1164

Carneiro, A., Soares, I., Rescorla, L., & Dias, P. (2020). Meta-Analysis on Parent-Teacher Agreement on Preschoolers' Emotional and Behavioural Problems. *Child Psychiatry & Human Development*. DOI: <https://doi.org/10.1007/s10578-020-01044-y>

Carter, A. S., Briggs-Gowan, M. J., & Davis, N. O. (2004). Assessment of young children's development and psychopathology: recent advances and recommendations for practice. *Journal of Child Psychology and Psychiatry*. 45: 109-134

Center on the Developing Child at Harvard University (2010). The Foundations of Lifelong Health Are Built in Early Childhood. Retrieved from <http://www.developingchild.harvard.edu>

Charach, A., Carson, P., Fox, S., Ali, M. U., Beckett, J., & Lim, C. G. (2013). Interventions for Preschool Children at High Risk for ADHD: A Comparative Effectiveness Review. *Pediatrics*. 131: 1584-1604

Charach, A., Bélanger, S. A., McLennan, J. D., & Nixon, M. K. (2017). Screening for disruptive behavior problems in preschool children in primary health care settings. *Paediatrics & Child Health*. 22: 478-484

Charach, A., Mohammadzadeh, F., Belanger, S. A., Easson, A., Lipman, E. L., McLennan, J. D., ..., & Szatmari, P. (2020). Identification of Preschool Children with Mental Health Problems in Primary Care: Systematic Review and Meta-analysis. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*. 29: 76-105

Chen, C., Squires, J., Heo, K. H., Bian, X., Chen, C-I., Filgueiras, A., ..., & Landeira-Fernandez, J. (2015). Cross Cultural Differences in Social-emotional Competence of Young Children: Comparisons with Brazil, China, South Korea, and the United States. *Mental Health in Family Medicine*. 11: 59-68

Chen, C., Squires, J., & Scalise, K. (2020). Evaluating the Dimensionality and Psychometric Properties of a Social-Emotional Screening Instrument for Young Children. *Infants & Young Children*. 33: 142-159

Choi, J. Y. & Dobbs-Oates, J. (2016). Teacher-Child Relationships: Contributions of Teacher and Child Characteristics. *Journal of Research in Childhood Education*. 30: 15-28

Cicchetti, D. & Rogosch, F. A. (1996). Equifinality and multifinality in developmental psychopathology. *Development and Psychopathology*. 8: 597-600

Coghill, D. & Sonuga-Barke, E. J. S. (2012). Annual Research Review: Categories versus dimensions in the classification and conceptualisation of child and adolescent mental disorders – implications of recent empirical study. *Journal of Child Psychology and Psychiatry*. 53: 469-489

Cortazar, A. & Herreros, F. (2010). Early Attachment Relationships and the Early Childhood Curriculum. *Contemporary Issues in Early Childhood*. 11: 192-202

Costello, H. & Bouras, N. (2006). Assessment of Mental Health Problems in People with Intellectual Disabilities. *Israel Journal of Psychiatry and Related Sciences*. 43: 241-251

Council on Children With Disabilities (2006). Identifying Infants and Young Children With Developmental Disorders in the Medical Home: An Algorithm for Developmental Surveillance and Screening. *Pediatrics*. 118: 405-420

Cox, M. J., Mills-Koonce, R., Propper, C., & Gariépy, J. L. (2010). Systems theory and cascades in developmental psychopathology. *Development and Psychopathology*. 22: 497-506

Dadds, M. R., Spence, S. H., Holland, D. E., Barrett, P. M., & Laurens, K. R. (1997). Prevention and early intervention for anxiety disorders: A controlled trial. *Journal of Consulting and Clinical Psychology*. 65: 627-635

Daley, D., Jones, K., Hutchings, J., & Thompson, M. (2009). Attention deficit hyperactivity disorder in pre-school children: current findings, recommended interventions and future directions. *Child: care, health, and development*. 35: 754-766

Davidson, R. J. & McEwen, B. S. (2012). Social influences on neuroplasticity: stress and interventions to promote well-being. *Nature Neuroscience*. 15: 689-695

Dirks, M. A., De Los Reyes, A., Briggs-Gowan, M., Cella, D., & Wakschlag, L. S. (2012). Annual Research Review: Embracing not erasing contextual variability in children's behavior- theory and utility in the selection and use of methods and informants in developmental psychopathology. *Journal of Child Psychology and Psychiatry*. 53: 558-574

Dougherty, L. R., Leppert, K. A., Merwin, S. M., Smith, V. C., Bufferd, S. J., & Kushner, M. R. (2015). Advances and Directions in Preschool Mental Health Research. *Child Development Perspectives*. 9: 14-19

- Doyle, O., Harmon, C. P., Heckman, J. J., & Tremblay, R. E. (2009). Investing in early human development: Timing and economic efficiency. *Economics and Human Biology*. 7: 1-6
- Drugli, M. B. & Hjemdal, O. (2013). Factor Structure of the Student-Teacher Relationship Scale for Norwegian School-Aged Children Explored with Confirmatory Factor Analysis. *Journal of Educational Research*. 57: 457-466
- Drugli, M. B. & Stensen, K. (2019). Psykiske vansker hos norske barnehagebarn. *Første steg*. 1: 46-49
- Duncan, G. J. & Magnuson, K. (2013). Investing in Preschool Programs. *Journal of Economic Perspectives*. 27: 109-132
- Dwyer, S. B., Nicholson, J. M., & Battistutta, D. (2006). Parent and teacher identification of children at risk of developing internalizing or externalizing mental health problems: A comparison of screening methods. *Prevention Science*. 7: 343-357
- Egger, H. L. & Angold, A. (2006). Common emotional and behavioral disorders in preschool children: presentation, nosology, and epidemiology. *Journal of Child Psychology and Psychiatry*. 47: 313-337
- Egger, H. L. & Emde, R. N. (2011). Developmentally Sensitive Criteria for Mental Health Disorders in Early Childhood. The Diagnostic and Statistical Manual of Mental Disorders-IV, the Research Diagnostic Criteria-Preschool Age, and the Diagnostic Classification of Mental Health and Developmental Disorders of Infancy and Early Childhood-Revised. *American Psychologist*. 66: 95-106
- Eklund, K., Renshaw, T. L., Dowdy, E., Jimerson, S. R., Hart, S. R., Jones, C. N., & Earhart, J. (2009). Early Identification of Behavioral and Emotional Problems in Youth: Universal Screening versus Teacher-Referral Identification. *The California School Psychologist*. 14: 89-95
- Ellingson, K. D., Briggs-Gowan, M. J., Carter, A. S., & Horwitz, S. M. (2004). *Archives of Pediatric and Adolescent Medicine*. 158: 766-772
- Essex, M. J., Kraemer, H. C., Slattery, M. J., Burk, L. R., Boyce, W. T., Woodward, H. R., & Kupfer, D. J. (2009). Screening for childhood mental health problems: outcomes and early identification. *Journal of Child Psychology and Psychiatry*. 50: 562-570
- Fanti, K. A. & Henrich, C. C. (2010). Trajectories of Pure and Co-Occurring Internalizing and Externalizing Problems From Age 2 to Age 12: Findings From the National Institute of Child Health and Human Development Study of Early Child Care. *Developmental Psychology*. 46: 1159-1175

- Fayers, P. M., & Machin, D. (2007). *Quality of Life*. 2.utg. John Wiley & Sons, Ltd.
- Feeney-Kettler, K. A., Kratochwill, T. R., Kaiser, A. P., Hemmeter, M. L., & Kettler R.J. (2010). Screening Young Children's Risk for Mental Health Problems: A Review of Four Measures. *Assessment of Effective Interventions*. 35: 218-230
- Feiring, C. & Lewis, M. (1987). Equifinality and Multifinality: Diversity in Development from Infancy into Childhood. *The Biennial Meeting of the Society for Research in Child Development, Baltimore*. Retrieved from: <https://files.eric.ed.gov/fulltext/ED283608.pdf>
- Finsaas, M. C., Bufferd, S. J., Dougherty, L. R., Carlson, G. A., & Klein, D. N. (2018a). Preschool Psychiatric Disorders: Homotypic and Heterotypic Continuity through Middle Childhood and Early Adolescence. *Psychological Medicine*. 48: 2159-2168
- Finsaas, M. C., Kessel, E. M., Dougherty, L. R., Bufferd, S. J., Danzig, A. P., Davila, J., Carlson, G. A., & Klein, D. N. (2018b). Early Childhood Psychopathology Prospectively Predicts Social Functioning in Early Adolescence. *Journal of Clinical Child & Adolescence Psychology*. DOI: <https://doi.org/10.1080/15374416.2018.1504298>
- Fox, L., Carta, J., Strain, P., Dunlap, G., & Hemmeter, M. L. (2009). *Response to Intervention and the Pyramid Model*. Tampa, Florida: University of South Florida, Technical Assistance Center on Social Emotional Intervention for Young Children.
- Glascoe, F. P. (1997). Parents' concerns about children's development: prescreening technique or screening test? *Pediatrics*. 99: 522-528
- Glascoe, F. P. (2001). Are Overreferrals on Developmental Screening Tests Really a Problem? *Archives of Pediatric and Adolescent Medicine*. 155: 54-59
- Glascoe, F. P. (2003). Parents' Evaluation of Developmental Status: How Well Do Parents' Concerns Identify Children With Behavioral and Emotional Problems? *Clinical Pediatrics*. 42: 133-138
- Glascoe, F. P. (2005). Screening for Developmental and Behavioral Problems. *Mental Retardation and Developmental Disabilities*. 11: 173-179
- Glascoe, F. P. & Marks, K. P. (2011). Detecting children with developmental-behavioral problems: The value of collaborating with parents. *Psychological Test and Assessment Modeling*. 53: 258-279

- Gokiert, R. J., Georgis, R., Tremblay, M., Krishnan, V., Vandenberghe, V., & Lee, C. (2014). Evaluating the Adequacy of Social-Emotional Measures in Early Childhood. *Journal of Psychoeducational Assessment*. 32: 441-454
- Gold, R., Reichman, M., Greenberg, E., Ivanidze, J., Elias, E., Tsiouris, A. J., ..., & Sanelli, P. C. (2010). Developing a New Reference Standard: Is Validation Necessary? *Academic Radiology*. 17: 1079-1082
- Gordon, M., Antshel, K., Faraone, S., Barkley, R., Lewandowski, L., Hudziak, J. J., ..., & Cunningham, C. (2006). Symptoms Versus Impairment. *Journal of Attention Disorders*. 9: 465-475
- de Groot, A., Koot, H. M., & Verhulst, F. C. (1994). Cross-Cultural Generalizability of the Child Behavior Checklist Cross-Informant Syndromes. *Psychological Assessment*. 6: 225-230
- Gross, D., Fogg, L., Garvey, C., & Julion, W. (2004). Behavior Problems in Young Children: An Analysis of Cross-Informant Agreements and Disagreements. *Research in Nursing & Health*. 27: 413-425
- Gustafsson, B. M., Proczkowska-Björklund, M., & Gustafsson, P. A. (2017). Emotional and behavioral problems in Swedish preschool children rated by preschool teachers with the Strengths and Difficulties Questionnaire (SDQ). *BMC Pediatrics*. 17. DOI: 10.1186/s12887-017-0864-2
- Haapsamo, H., Ebeling, H., Soini, H., Juskitt, L., Larinen, K., Penninkilampi-Kerola, V., ..., & Moilanen, I. (2009). Screening Infants with Social and Emotional Problems: A Pilot Study of the Brief Infant Toddler Social and Emotional Assessment (BITSEA) in Northern Finland. *International Journal of Circumpolar Health*. 68: 386-393
- Halperin, J. M., Bédard, A. V., & Curchack-Lichtin, J. T. (2012). Preventive Interventions for ADHD: A Neurodevelopmental Perspective. *Neurotherapeutics*. 9: 531-541
- Harvey, E. A., Fischer, C., Weieneth, J. L., Hurwitz, S. D., & Sayer, A. G. (2013). Predictors of discrepancies between informants' ratings of preschool-aged children's behavior: An examination of ethnicity, child characteristics, and family functioning. *Early Childhood Research Quarterly*. 28: 668-682
- Heckman, J. J. (2006). Skill Formation and the Economics of Investing in Disadvantaged Children. *Science*. 312: 1900-1902

Heiervang, E., Goodman, A., & Goodman, R. (2008). The Nordic advantage in child mental health: separating health differences from reporting style in a cross-cultural comparison of psychopathology. *Journal of Child Psychology and Psychiatry*. 49: 678-685

Hewlett, J. & Waisbren, S. E. (2006). A review of the psychosocial effects of false-positive results on parents and current communication practices in newborn screening. *Journal of Inherited Metabolic Disease*. 29: 677-682

Heo, K. H., & Squires, J. (2012). Cultural adaptation of a parent completed social emotional screening instrument for young children: Ages and stages questionnaire-social emotional. *Early Human Development*. 88: 151-158

Horwitz, S. M., Gary, L. C., Briggs-Gowan, M. J., & Carter, A. S. (2003). Do Needs Drive Service Use in Young Children? (2003). *Pediatrics*. 112: 1373-1378

Horwitz, S. M., Kelleher, K. J., Stein, R. E. K., Storfer-Isser, A., Youngstrom, E. A., Park, E. R., ..., & Hoagwood, K. E. (2007). Barriers to the Identification and Management of Psychosocial Issues in Children and Maternal Depression. *Pediatrics*. 119: 208-218

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2. ed.). New York: John Wiley & Sons Inc. 2000

Hui, S. L. & Zhou, X. H. (1998). Evaluation of diagnostic tests without gold standards. *Statistical Methods in Medical Research*. 7: 354-370

Ivanova, M. Y., Achenbach, T. M., Rescorla, L. A., Dumenci, L., Almqvist, F., Bathiche, M., ..., & Verhulst, F. (2007). Testing the Teacher's Report Form syndromes in 20 societies. *School Psychology Review*. 36: 468-483

Ivanova, M. Y., Achenbach, T. M., Rescorla, L. A., Harder, V. S., Ang, R. P., Bilenberg, N., ... & Verhulst, F. C. (2010). Preschool Psychopathology Reported by Parents in 23 Societies: Testing the Seven-Syndrome Model of the Child Behavior Checklist for Ages 1,5-5. *New Research*. 49: 1215-1224

Ivanova, M. Y., Achenbach, T. M., Rescorla, L. A., Bilenberg, N., Bjarnadottir, G., Denner, S., ..., & Verhulst, F. C. (2011). Syndromes of Preschool Psychopathology Reported by Teachers and Caregivers in 14 Societies Using the Caregiver-Teacher Report Form (C-TRF). *Journal of Early Childhood and Infant Psychology*. 7: 87-103

Jensen, P. S. & Watanabe, H. (1999). Sherlock Holmes and Child Psychopathology Assessment Approaches: The Case of the False-Positive. *Journal of American Academy of Child and Adolescent Psychiatry*. 38: 138-146

Karabekiroglu, K., Rodopman-Arman, A., Ay, P., Ozkesen, M., Akbas, S., Tasdemir, G. N., Boke, O., & Peksen, Y. (2009). The reliability and validity of the Turkish version of the brief infant-toddler social emotional assessment (BITSEA). *Infant Behavior and Development*. 32: 291-297

Karabekiroglu, K., Briggs-Gowan, M. J., Carter, A. S., Rodopman-Arman, A., & Akbas, S. (2010). The clinical validity and reliability of the Brief Infant-Toddler Social and Emotional Assessment (BITSEA). *Infant Behavior and Development*. 33: 503-509

Kauffman, J. M. (1999). How we prevent the prevention of emotional and behavioral disorders. *Exceptional Children*. 65: 448-468

Keenan, K., Shaw, D., Delliquadri, E., Giovannelli, J., & Walsh, B. (1998). Evidence for the Continuity of Early Problem Behaviors: Application of a Developmental Model. *Journal of Abnormal Child Psychology*. 26: 441-454

Kerr, D. C. R., Lunkenheimer, E. S., & Olson, S. L. (2007). Assessment of child problem behaviors by multiple informants: a longitudinal study from preschool to school entry. *Journal of Child Psychology and Psychiatry*. 48: 967-975

Koot, H., van den Oord, J., Verhulst, F. C., & Boomsma, D. (1997). Behavioral and emotional problems in young preschoolers: Cross-cultural testing of the validity of the Child Behavior Checklist/2-3. *Journal of Abnormal Child Psychology*. 25: 183-196

Kristensen, S., Henriksen T. B., & Bilenberg, N. (2010). The Child Behavior Checklist for Ages 1.5-5 (CBCL/1½-5): Assessment and analysis of parent and caregiver-reported problems in a population-based sample of Danish preschool children. *Nordic Journal of Psychiatry*. 64: 203-209

Kruizinga, I., Jansen, W., de Haan, C. L., van der Ende, J., Carter, A. S., & Raat, H. (2012). Reliability and Validity of the Dutch Version of the Brief Infant-Toddler Social and Emotional Assessment (BITSEA). *PLoS ONE*. 7: e38762. doi:10.1371/journal.pone.0038762

Kruizinga, I., Jansen, W., Mieloo, C. L., Carter, A. S., & Raat, H. (2013). Screening Accuracy and Clinical Application of the Brief Infant-Toddler Social and Emotional Assessment (BITSEA). *PLoS ONE*. 8: e72602. doi:10.1371/journal.pone.0072602

Kruizinga, I., Jansen, W., van Sprang, N. C., Carter, A. S., & Raat, H. (2015). The Effectiveness of the BITSEA as a Tool to Early Detect Psychosocial Problems in Toddlers, a Cluster Randomized Trial. *PLoS ONE*. 10: e0136488. doi:10.1371/journal.pone.0136488

Kucuker, S., Kapci, E. G., & Uslu, R. I. (2011). Evaluation of the Turkish Version of “the Ages and Stages Questionnaire; Social-Emotional” in Identifying Children with Social-Emotional Problems. *Infants & Young Children*. 24: 207-220

Larsson, B. & Drugli, M. B. (2011). School competence and emotional/behavioral problems among Norwegian school children as rated by teachers on the Teacher Report Form. *Scandinavian Journal of Psychology*. 52: 553-559

Lavigne, J. V., Arend, R., Rosenbaum, D., Binns, H. J., Christoffel, K. K., & Gibbons, R. D. (1998). Psychiatric Disorders With Onset in the Preschool Years: I. Stability of Diagnoses. *Journal of American Academy of Child and Adolescent Psychiatry*. 37: 1246-1254

Lavigne, J. V., Feldman, M., & Meyers, K. M. (2016a). Screening for Mental Health Problems: Addressing the Base Rate Fallacy for a Sustainable Screening Program in Primary Care. *Journal of Pediatric Psychology*. 41: 1081-1090

Lavigne, J. V., Meyers, K. M., & Feldman, M. (2016b). Systematic Review: Classification Accuracy of Behavioral Screening Measures for Use in Integrated Primary Care Settings. *Journal of Pediatric Psychology*. 5: 1-19

Layne, A. E., Bernstein, G. A., & March J. S. (2006). Teacher Awareness of Anxiety Symptoms in Children. *Child Psychiatry and Human Development*. 36: 383-392

Lekhal, R. (2020). Receiving extra support in Norwegian centre-based childcare: the role of children’s language and socioemotional development. *Early Child Development and Care*. 190: 580-593

Lenze, S. N., Pautsch, J., & Luby, J. (2011). Parent-Child Interaction Therapy Emotion Development: A Novel Treatment For Depression in Preschool Children. *Depression and Anxiety*. 28: 153-159

Liljequist, L. & Renk, K. (2007). The Relationships among Teachers' Perceptions of Student Behaviour, Teachers' Characteristics, and Ratings of Students' Emotional and Behavioural Problems. *Educational Psychology*. 27: 557-571

Loades, M. E. & Mastroyannopoulou, K. (2010). Teachers' Recognition of Children's Mental Health Problems. *Child and Adolescent Mental Health*. 15: 150-156

Luby, J. L. (2010). Preschool Depression: The Importance of Identification of Depression Early in Development. *Current Directions in Psychological Science*. 19: 91-95

Luby, J., Lenze, S., & Tillman, R. (2012). A novel early intervention for preschool depression: findings from a pilot randomized controlled trial. *Journal of Child Psychology and Psychiatry*. 53: 313-322

Luby, J. L. (2013). Treatment of Anxiety and Depression in the Preschool Period. *Journal of the American Academy of Child and Adolescent Psychiatry*. 52: 346-358

Lydersen, S. (2012). Diagnostic tests, ROC curves, and measures of agreement. In M. B. Veierød, S. Lydersen, & P. Laake (Eds.). *Medical statistics in clinical and epidemiological research* (p. 462-492). Oslo: Gyldendal Norsk Forlag

Masten, A. S. & Cicchetti, D. (2010). Developmental cascades. *Development and Psychopathology*. 22: 491-495

Mathiesen, K. S. & Sanson, A. (2000). Dimensions of early childhood behavior problems: Stability and predictors of change from 18 to 30 months. *Journal of Abnormal Child Psychology*. 28: 15-31

McCabe, P. C. & Altamura, M. (2011). Empirically Valid Strategies to Improve Social and Emotional Competence of Preschool Children. *Psychology in the Schools*. 48: 513-540

McKnight, P. E. & Kashdan, T. B. (2009). The importance of functional impairment to mental health outcomes: A case for reassessing our goals in depression treatment research. *Clinical Psychology Review*. 29: 243-259

McKnight, P. E., Monfort, S. S., Kashdan, T. B., Blalock, D. V., & Calton, J. M. (2016). Anxiety symptoms and functional impairment: A systematic review of the correlation between the two measures. *Clinical Psychology Review*. 45: 115-130

Moor, S., Sharrock, G., Scorr, J., McQueen, H., Wrate, R., Cowan, J., & Blair, C. (2000). Evaluation of a teaching package designed to improve teachers' recognition of depressed pupils - a pilot study. *Journal of Adolescence*. 23: 331-342

Myers, S. S. & Pianta, R. C. (2008). Developmental Commentary: Individual and Contextual Influences on Student-Teacher Relationships and Children's Early Problem Behaviors. *Journal of Clinical Child & Adolescent Psychology*. 37: 600-608

Neyman, J. M., Heron, M. A., Hayatbakhsh, M. R., Dingle, K., Jamrozik, K., Bor, W., O'Callaghan, M. J., & Williams, G. M. (2008). Screening in early childhood for risk of later mental health problems: A longitudinal study. *Journal of Psychiatric Research*. 42: 694-700

Njoroge, W. F. M. & Bernhart, K. P. (2011). Assessment of Behavioral Disorders in Preschool-Aged Children. *Current Psychiatry Reports*. 13: 84-92

Nores, M. & Barnett, W. S. (2010). Benefits of early childhood interventions across the world: (Under) Investing in the very young. *Economics of Education Review*. 29: 271-282

Norwegian Kindergarten Act (2017). Framework Plan for Kindergartens. Retrieved from: <https://www.udir.no/globalassets/filer/barnehage/rammeplan/framework-plan-for-kindergartens2-2017.pdf>

Nøvik, T. S. (2000). Child Behavior Checklist item scores in Norwegian children. *European Child & Adolescent Psychiatry*. 9: 54-60

Ollendick, T. H., Greene, R. W., Weist, M. D., & Oswald, D. P. (1990). The Predictive Validity of Teacher Nominations; A Five-Year Follow-up of At-Risk Youth. *Journal of Abnormal Child Psychology*. 18: 699-713

Pianta, R. C. (2001). *STRS Student-Teacher Relationship Scale (professional manual)*. Lutz, FL: Psychological Assessment Resources, Inc.

Pihlakoski, L., Sourander, A., Aromaa, M., Rautava, P., Helenius, H., & Sillanpää, M. (2006). The continuity of psychopathology from early childhood to preadolescence. *European Child & Adolescent Psychiatry*. 15: 409-417

Polanczyk, G. V., Salum, G. A., Sugaya, L. S., Caye, A., & Rohde, L. A. (2015). Annual Research Review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *Journal of Child Psychology and Psychiatry*. doi:10.1111/jcpp.12381

- Poulou, M. S. (2015). Emotional and Behavioural Difficulties in Preschool. *Journal of Child and Family Studies*. 24: 225-236
- Powell, D., Dunlap, G., and Fox, L. (2006). Prevention and Intervention for the Challenging Behaviors of Toddlers and Preschoolers. *Infants & Young Children*. 19: 25-35
- Radecki, L., Sand-Loud, N., O'Connor, K. G., Sharp, S., & Olson, L. M. (2011). Trends in the Use of Standardized Tools for Developmental Screening in Early Childhood: 2002-2009. *Pediatrics*. 128: 14-19
- Rapee, R. M. (2013). The preventive effects of a brief, early intervention for preschool-aged children at risk for internalising: follow-up into middle adolescence. *Journal of Child Psychology and Psychiatry*. 54: 780-788
- Raver, C. C., Jones, S. M., Li-Grining, C., Zhai, F., Metzger, M. W., & Solomon, B. (2009). Targeting Children's Behavior Problems in Preschool Classrooms: A Cluster-Randomized Controlled Trial. *Journal of Consulting and Clinical Psychology*. 77: 302-316
- Rescorla, L. A. (2005). Assessment of Young Children Using the Achenbach System of Empirically Based Assessment (ASEBA). *Mental Retardation and Developmental Disabilities Research Review*. 11: 226-237
- Rescorla, L. A., Achenbach, T. M., Ivanova, M. Y., Bilenber, N., Bjarnadottir, G., Denner, S., ... & Verhulst, F. C. (2012). Behavioral/Emotional Problems of Preschoolers: Caregiver/Teacher Reports From 15 Societies. *Journal of Emotional and Behavioral Disorders*. 20: 68-81
- Rescorla, L. A., Boichichio, L., Achenbach, T. M., Ivanova, M. Y., Almqvist, F., Begovac, I., ... & Verhulst, F. C. (2014). Parent-Teacher Agreement on Children's Problems in 21 Societies. *Journal of Clinical Child and Adolescent Psychology*. 43: 627-642
- Roeser, R. W., & Midgley, C. (1997). Teachers' views of issues involving students' mental health. *The Elementary School Journal*. 2: 115-133
- Sabol, T. J. & Pianta, R. C. (2012). Recent trends in research on teacher-child relationships. *Attachment and Human Development*. 14: 213-231
- Salomonsson, B., & Sleded, M. (2010). The Ages & Stages Questionnaire: Social-Emotional: A Validation Study of a Mother-Report Questionnaire on a Clinical Mother-Infant Sample. *Infant Mental Health Journal*. 31: 412-431

- Sameroff, A. J. (2000). Developmental systems and psychopathology. *Development and Psychopathology*. 12: 297-312
- Sameroff, A. J. (2009). The Transactional Model. In A. J. Sameroff (Ed.), *The Transactional Model of Development* (p. 3-21). Washington DC: American Psychological Association
- Sameroff, A. J. & Mackenzie, M. J. (2003). Research strategies for capturing transactional models of development: The limits of the possible. *Development and Psychopathology*. 15: 613-640
- Sawyer, A. C. P., Chittleborough, C. R., Lynch, J. W., Baghurst, P., Mittinty, M. N., Kaim, A. L. E., & Sawyer, M. G. (2013). Can screening 4-5 year olds accurately identify children who will have teacher reported mental health problems when children are aged 6-7 years? *Australian & New Zealand Journal of Psychiatry*. 00: 1-10
- Shankman, S. A., Lewinsohn, P. M., Klein, D. N., Small, J. W., Seeley, J. R., & Altman, S. E. (2009). Subthreshold conditions as precursors for full syndrome disorders: a 15-year longitudinal study of multiple diagnostic classes. *Journal of Child Psychology and Psychiatry*. 50: 1485-1494
- Sheldrick, R. C., Merchant, S., & Perrin, E. C. (2011). Identification of Developmental-Behavioral Problems in Primary Care: A Systematic Review. *Pediatrics*. 128: 356-363
- Sheldrick, R. C., Benneyan, J. C., Kiss, I. G., Briggs-Gowan, M. J., Copeland, W., & Carter, A. S. (2015). Threshold and accuracy in screening tools for early detection of psychopathology. *The Journal of Child Psychology and Psychiatry*. 56: 936-948
- Silver, R. B., Measelle, J. R., Armstrong, J. M., & Essex, M. J. (2005). Trajectories of classroom externalizing behavior: Contributions of child characteristics, family characteristics, and the teacher-child relationship during the school transition. *Journal of School Psychology*. 43:39-60
- Skalická, V., Stenseng, F., & Wichstrøm, L. (2015). Reciprocal relations between student-teacher conflict, children's social skills and externalizing behavior: A three wave longitudinal study from preschool to third grade. *International Journal of Behavioral Development*. 39: 413-425
- Skogen, J. C., Kjeldsen, A., Knudsen, A. K., Myklestad, I., Nesvåg, R., Reneflot, A., & Major, E. (2014). *Psykisk helse hos barn og unge* (Folkehelse rapporten 2014:4). Oslo: Nasjonalt Folkehelseinstitutt

Skovgaard, A. M., Houmann, T., Landorph, S. L., & Christiansen, E. (2004). Assessment and classification of psychopathology in epidemiological research of children 0-3 years of age. A review of the literature. *European Child & Adolescent Psychiatry*. 13: 337-346

Skovgaard, A. M., Houmann, T., Christiansen, E., Olsen, E. M., Landorph, S. L., Lichtenberg, A., & Jorgensen, T. (2008). Can a general health surveillance between birth and 10 months identify children with mental disorder at 1 ½ year? A case-control study nested in cohort CCC 2000. *European Child and Adolescent Psychiatry*. 17: 290-298

Soles, T., Bloom, E. L., Heath, N. L., & Karagiannakis, A. (2008). An exploration of teachers' current perceptions of children with emotional and behavioural difficulties. *Emotional and Behavioural Difficulties*. 13: 275-290

Solheim, E., Berg-Nielsen, T. S., & Wichstrøm, L. (2012). The Three Dimensions of the Student-Teacher Relationship Scale: CFA Validation in a Preschool Sample. *Journal of Psychoeducational Assessment*. 30: 250-263

Squires, J., Bricker, D., & Twombly, E. (2002). *The ASQ:SE User's Guide*. Paul H. Brookes Publishing Co.

Squires, J., Bricker, D., & Twombly, E. (2015). *Ages & Stages Questionnaires: social-emotional, second edition*. Baltimore, MD : Paul H. Brookes Publishing

Sroufe, L. A. (2005). Attachment and development: A prospective, longitudinal study from birth to adulthood. *Attachment and Human Development*. 7: 349-367

Statistics Norway (2020). Kindergartens. Retrieved from <https://www.ssb.no/barnehager>

Strickland, J., Hopkins, J., & Keenan, K. (2012) Mother-Teacher Agreement on Preschoolers' Symptoms of ODD and CD: Does Context Matter? *Journal of Abnormal Child Psychology*. 40: 933-943

Sveen, T. H., Berg-Nielsen, T.S., Lydersen, S., & Wichstrøm, L. (2013). Detecting Psychiatric Disorders in Preschoolers: Screening With the Strengths and Difficulties Questionnaire. *Journal of American Academy of Child and Adolescent Psychiatry*. 52: 728-736

Szaniecki, E. & Barnes, J. (2016). Measurement issues: measures of infant mental health. *Child and Adolescent Mental Health*. 21: 64-74

- Theunissen, M. H. C., Vogels, A. G. C., de Wolff, M. S., Crone, M. R., & Reijneveld, S. A. (2015). Comparing three short questionnaires to detect psychosocial problems among 3 to 4- year olds. *BMC Pediatrics*. 15: 1-8
- Thompson, R. A. (2000). The Legacy of Early Attachments. *Child Development*. 71: 145-152
- Tveit, H. H., Drugli, M. B., Fossum, S., Handegård, B. H., & Stenseng, F. (2019). Does the Incredible Years Teacher Classroom Management programme improve child-teacher relationships in childcare centres? A 1-year universal intervention in a Norwegian community sample. *European Child & Adolescent Psychiatry*. DOI: <https://doi.org/10.1007/s00787-019-01387-5>
- Vaezghasemi, M., Eurenus, E., Ivarsson, A., Sundberg, L. R., Silfverdal, S., & Lindkvist, M. (2020). Social-emotional problems among Swedish three-year-olds: an Item Response Theory analysis of the Ages and Stages Questionnaire: Social-Emotional. *BMC Pediatrics*. DOI: <https://doi.org/10.1186/s12887-020-2000-y>
- Vasileva, M., Graf, R. K., Reinelt, T., Petermann, T., & Petermann, F. (2020). Research review: A meta-analysis of the international prevalence and comorbidity of mental disorders in children between 1 and 7 years. *Journal of Child Psychology and Psychiatry*. DOI: 10.1111/jcpp.13261
- Verhulst, F. C., & Koot, H. M. (1992). *Child psychiatric epidemiology: concepts, methods and findings*. Beverly Hills: Sage Publications
- Wichstrøm, L., Berg-Nielsen, T.S., Angold, A., Egger, H.L., Solheim, E. & Sveen, T.H. (2012). Prevalence of psychiatric disorders in preschoolers, *Journal of Child Psychology and Psychiatry*. 53: 695–705
- Wichstrøm, L., Belsky, J., Jozefiak, T., Sourander, A. & Berg-Nielsen T. S. (2014). Predicting Service Use for Mental Health Problems Among Young Children. *Pediatrics*. 133: 1054-1060
- de Wolff, M. S., Theunissen, M. H., Vogels, A. G. C., & Reijneveld, S. A. R. (2013). Three Questionnaires to Detect Psychosocial Problems in Toddlers: A Comparison of the BITSEA, ASQ:SE, and KIPPPI. *Academic Pediatrics*. 13: 587-592
- Xie, H., Bian, X., Chen, C., Squires, J., & Lu, P. (2019). Examining the Convergent Evidence of a Parent-Completed, Social-Emotional Screening Tool in China. *Journal of Child and Family Studies*. DOI: 10.1007/s10826-019-01362-3

Young, N. D. & Takala, C. R. (2018). Sequential Screening to Improve Behavioral Health Needs in Primary Care. *Journal of American Child and Adolescent Psychiatry*. 57: 603-609

Zeanah, C. H., Berlin, L. J., & Boris, N. W. (2011). Practitioner Review: Clinical applications of attachment theory and research for infants and young children. *Journal of Child Psychology and Psychiatry*. 52: 819-833

Zhang, X. & Sun, J. (2011). The Reciprocal Relations Between Teachers' Perception of Children's Behavior Problems and Teacher-Child Relationships on the First Preschool Year. *The Journal of Genetic Psychology*. 172: 176-198

6 Appendix

Paper I



Childcare Providers' Nominations of Preschool Children at Risk for Mental Health Problems: Does it Discriminate Well Compared to the Caregiver-Teacher Report Form (C-TRF)?

Kenneth Stensen , Stian Lydersen , Frode Stenseng , Jan Lance Wallander , Håvard Tveit & May Britt Drugli

To cite this article: Kenneth Stensen , Stian Lydersen , Frode Stenseng , Jan Lance Wallander , Håvard Tveit & May Britt Drugli (2021): Childcare Providers' Nominations of Preschool Children at Risk for Mental Health Problems: Does it Discriminate Well Compared to the Caregiver-Teacher Report Form (C-TRF)?, Scandinavian Journal of Educational Research

To link to this article: <https://doi.org/10.1080/00313831.2020.1869089>



Published online: 11 Jan 2021.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

Full Terms & Conditions of access and use can be found at
<https://www.tandfonline.com/action/journalInformation?journalCode=csje20>

Paper II

Teacher Nominations of Preschool Children at Risk for Mental Health Problems:

How False is a False Positive Nomination and What Make Teachers Concerned?

Kenneth Stensen¹, Stian Lydersen¹, Frode Stenseng², Jan Lance Wallander^{1,3}, May Britt Drugli^{1,4}

¹ Regional Centre for Child and Youth Mental Health and Child Welfare, Faculty of Medicine and Health Science, Department of Mental Health, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

² Department of Education and Lifelong Learning, Faculty of Social and Educational Sciences, Norwegian University of Science and Technology (NTNU) and Queen Maud University College, Trondheim, Norway

³ University of California, Merced, California, USA

⁴ Centre of the Study of Educational Practice, Inland University College, Hamar, Norway

Correspondence regarding this article should be addressed to:

Kenneth Stensen

Email: kenneth.stensen@ntnu.no

Postal address: Regional Centre for Child and Youth Mental Health and Child Welfare, P.O.

Box 8905, MTFS, N-7491 Trondheim

Abstract

Objective: Identification attempts in populations with a low prevalence of problems usually result in a considerable number of false positives. Thus, the aim of the current study was to investigate the false positive rate following nomination of developmental concerns by preschool teachers and the reasons for which teachers raise developmental concerns about children who display non-clinical levels of mental health problems.

Methods: A total of 1430 children aged 1 to 6 years in Norwegian childcare centers were classified as true positive, false positive, true negative, or false negative by comparing preschool teachers' nomination with their ratings on the Caregiver-Teacher Report Form, resulting in 127 (9%) false positives and 1142 (80%) true negatives.

Results: Compared to the true negative group, the false positive group received significantly higher scores on internalizing problems, externalizing problems than true negatives, conflict and significantly lower scores on closeness. Children's internalizing and externalizing problems and age were the main factors that increased the likelihood of teachers raising concerns, while increased closeness in the teacher-child relationship reduced the likelihood of being nominated. Children's gender and conflict level were not significant when adjusting for other factors.

Conclusion: These findings suggest that preschool teachers' concerns about children's development should not be discarded as the false positive group did show elevated levels of problem behavior and poorer teacher-child relationship compared to the true negative group. Scrutinizing concerns in collaboration with parents and other mental health professionals may be beneficial to ensure healthy development for children with elevated problem levels.

Keywords: Preschool children; Teacher nominations; Mental health problems; False positives

Introduction

Globally, approximately 20% of all children experience mental health difficulties (Belfer, 2008) and 13% meet the criteria for a psychiatric disorder (Polanczyk et al., 2015). In Norway, the prevalence rates are slightly lower with 15% to 20% of the preschool children exhibiting some mental health problems (Skogen et al., 2014) and 7% showing a symptom load that would qualify for a psychiatric disorder (Wichstrøm et al., 2012). However, very few preschoolers who meet diagnostic criteria are referred for mental health evaluation or receive treatment (Egger & Angold, 2006; Horwitz et al., 2003; Horwitz et al., 2007) with approximately half of all children with behavioral disabilities not being identified before school entry (Glascoe & Marks, 2011). Additionally, in Norway, only 10% of four-year-old children with emotional and behavioral difficulties have received help (Wichstrøm et al., 2014).

At the community level, parents and other caregivers, such as preschool teachers, are the only viable source of information regarding young children's development (Sveen et al., 2013). Parents are the primary initiators of contact with health services when there are concerns about a child's development (Ellingson et al., 2004). Usually, the task of identifying children with emotional and behavioral difficulties has been carried out by pediatric practitioners in collaboration with parents. However, many cases may go undetected if responsibility is placed solely on parents and pediatric practitioners (Lavigne et al., 2016a). Alternatively, preschool may provide a valuable context to screen for early developmental concerns. Thus, more attention should be directed towards preschool teachers' perceptions of children's difficulties (Poulou, 2015) as their observational accuracy may be an important factor in connecting children in need of help with relevant mental health services (Berkhout et al., 2012; Eklund et al., 2009). However, surprisingly little research has been carried out on

the ability of preschool teachers to identify children with emotional and behavioral difficulties.

There is an increasing awareness of the great importance of early identification of emotional and behavioral difficulties (Council on Children with Disabilities et al., 2006; Essex et al., 2009; Glascoe & Marks, 2011; Njoroge & Bernhart, 2011; Radecki et al., 2011). The rapid development occurring in preschool-aged children may place them at risk of developing emotional/internalized problems and behavioral/externalized problems. While some children overcome these difficulties, others struggle to return to a normal developmental trajectory (Essex et al., 2009; Lavigne et al., 1998; Poulou, 2015). Early emotional and behavioral difficulties are predictors of later maladjustment, underlining the importance of identifying those children with high levels of internalized and externalized problems or those with continuity in their problem behavior over time (Basten et al., 2016; Briggs-Gowan et al., 2008; Essex et al., 2009; Fanti & Henrich, 2010; Gilliom & Shaw, 2004). Well-timed and targeted interventions may disrupt these negative trajectories and enhance the probability of better adjustment (Masten & Cicchetti, 2010). This presumes that children in need of intervention are identified at an early stage; however, such identification may be challenging in a period during which children's development proceeds rapidly (Keenan et al., 1998). Given the potentially serious consequences of early difficulties in children's development and lifelong health (Center on the Developing Child Harvard University, 2010), developing procedures to identify those in need of help should be a public health priority (Essex et al., 2009; Sawyer et al., 2013).

In contrast to diagnostic assessment, the main purpose of identifying children at risk through screening is to detect which children are in need of further assessment to decide whether treatment is necessary or not. In other words, screening may be viewed as a first step that indicates the need for a more thorough diagnostic assessment and identifies children who

do not meet diagnostic criteria but may still be at risk developmentally. Screening procedures are usually shorter and less costly to administer than diagnostic assessments, and the briefest and most low-cost screening procedure is the nomination method. Simply, this approach involves informants nominating the children who they perceive to meet a given criteria. This can be in the form of either a certain number of risk factors being present or the informant's perception of developmental concerns. The nomination method may be regarded as pre-screening or as a subjective judgment call as it can direct attention towards subsets of children in need of further screening.

Preschool teachers are a logical source for the developmental screening of children during early development. However, accuracy in identifying children with difficulties has predominantly been investigated for teachers of school-aged children rather than teachers of preschool children. Several studies of children in elementary school have shown that teachers are more likely to elicit concern and exhibit higher precision regarding externalizing symptoms compared with internalizing symptoms (Dwyer et al., 2006; Loades & Mastroyannopoulou, 2010; Soles et al., 2008). Furthermore, children identified as needing mental health services show significantly more adjustment problems than their peers (Layne et al., 2006; Roeser & Midgley, 1997) and children nominated by a teacher as at-risk for problems differ significantly from non-nominees with respect to academic grades, sociometric status, and social behavior five years after nomination (Ollendick et al., 1990). However, previous studies have reported low to moderate accuracy when teachers are asked to identify children with anxiety and depression difficulties (Dadds et al., 1997; Moor et al., 2000). Similar to elementary school teachers, preschool teachers also tend to under-report internalizing symptoms (e.g., anxiety, depression, and withdrawal) compared with parent reports. In addition, if preschool teachers perceive their relationship with a child as conflictual, they tend to over-report both internalizing and externalizing symptoms (Berg-

Nielsen et al., 2012). That said, it has been reported that preschool teachers can identify a considerable portion of children rated with a clinical symptom load (true positives), especially among the oldest preschool children, and leave very few false negative cases (children rated with a clinical symptom load for which teachers have no concerns). However, preschool teachers' nominations can also produce a high rate of false positive cases with about every other nomination identifying a child rated with a non-clinical symptom load (Stensen et al., 2021). Being identified as a false negative, that is having a negative screening test with positive follow-up or presence of problems that are not identified in a timely manner, may deprive children of appropriate help through non-referral for a more thorough assessment. Misclassifications in the form of false positives, that is a positive screening test with negative follow-up results, indicating an absence of problems, may have a stigmatizing effect for the child and create unnecessary anxiety for the child and parents. This will also result in a waste of clinical resources, which ideally should be allocated to those clearly needing services.

The high rates of false positive cases associated with screening in a population with a low base rate of difficulties, such as a normal population of preschool children, remain problematic (Lavigne et al., 2016a). The ability of screening procedures to classify cases correctly tends to decline as the base rate of the target problem declines, thereby leading to a higher misclassification rate (Lavigne et al., 2016a; Young & Takala, 2018). However, some findings suggest that the validity of a false positive screening result can be debated. By drawing from various screening instrument validation studies, Glascoe (2001) recruited 512 parents and their children, age 7 months to 8 years, to undergo screening followed by diagnostic assessment for all. The results showed that the children classified as false positives performed significantly worse on diagnostic measures than the children classified as true negatives (i.e., those who showed negative screening results and an absence of problems). In

another study, Jensen and Watanbe (1999) used DSM criteria and symptom checklists, as well as other survey measures, to compare true positive, false positive, true negative, and false negative cases. They found that the false positive cases exhibited higher levels of a range of risk factors than the true negative cases.

Even though reports of estimated accuracy are somewhat mixed, the nomination method can be used to identify a considerable proportion of children with emotional and behavioral problems. Children classified as false positive cases may still present problems, although not necessarily at a clinical level. Indeed, these children have been reported to receive poorer outcome scores on various measures than children classified as true negative (Glascoe, 2001; Jensen & Watanbe, 1999). In addition, it has been reported that children displaying symptoms of psychopathology experience considerable impairment even when they do not meet DSM criteria for disorders (Angold et al., 1999). Subthreshold conditions may be effective targets for preventive interventions as these can be precursors for disorders later in life (Shankman et al., 2009). Because preschool teachers can potentially play an important role in identifying and helping children with emotional and behavioral problems, the validity of false positive teacher nominations needs to be investigated by comparing the characteristics of those classified as false positive with those classified as true negative. The current study thus sought to test the following hypothesis in a sample of preschool children:

(H1) Children with false positive teacher nominations will receive higher scores for problem behaviors and lower scores for teacher-child relationship.

(H2) Children with an elevated but non-clinical, symptom load will have higher odds of a false positive classification.

(H3) Negative teacher-child relationships, represented by either high levels of conflict or low levels of closeness, will increase the odds of a false positive classification.

Methods

Data were collected from 2012-2014 as part of the Children in Central Norway study, which aimed to enhance teacher competence in addressing preschool children's mental health and to improve the quality of relationships between teachers and children. The study was approved by the Regional Committee for Medical and Health Research Ethics.

Procedure and participants

Parents with children in childcare centers, serving children from age one to six years old, in three municipalities in Central Norway received recruitment letters with information regarding the project as well as an informed consent form. Information was also provided in parent meetings before the project started. The recruitment letter provided the option for parents to consent either by logging in with a personal invitation code or by returning the consent form to the childcare center. Participation was voluntary and parental consent could be withdrawn at any time until the participation registry was deleted without reprisal. Parental consent gave the teacher in the childcare center who was most familiar with their child permission to complete a survey regarding that parent's child. Children are usually enrolled in childcare centers in the autumn, and the data were collected in January the following year. Thus, most teachers would be expected to have known the child for at least a few months. The teachers provided consent electronically with their own invitation codes. Of the invited parents, 1631 (77%) consented to enroll their child in the study, and 169 teachers (7% male) reported on 1431 children (88% of eligible). The gender distribution of the children was 51% boys and 49% girls, and the mean age was 45 months. In the survey, the preschool teachers were first asked to decide whether or not to nominate the child as being subject of developmental concern. Next, the teachers were asked to respond to the Student-Teacher

Relationship Scale (STRS) and the Caregiver-Teacher Report Form (C-TRF). The teachers responded for all children in the same sitting.

Measures

Teacher nomination

Teachers were asked to make a global subjective judgment concerning each child's risk status by answering "yes" or "no" to indicate whether they perceived that the child had any developmental concerns. This question was located at the start of the survey before the standardized questionnaires were presented. If "yes" was answered, teachers could specify their nomination by checking one or more reasons for the nomination, including aggression, attention, emotional, social, motoric, language, and home situation. However, only those nominated with specifications of aggression, attention, emotional, or social concerns were considered in the analyses to have been nominated at being at risk to match the types of problems addressed in the criterion (see below).

Student-Teacher Relationship Scale (S-TRS)

The S-TRS (Pianta, 2001) is a teacher-report form developed to measure teachers' perception of their relationship to a child or student. It contains the subscales *closeness*, *conflict*, and *dependency* with item responses ranging from 1 ("definitely does not apply") to 5 ("definitely applies"). In the current study, only the *closeness* (11 items) and *conflict* (12 items) subscales were used. A total score for each subscale is obtained by summing the individual items, where higher scores in *closeness* indicate a higher degree of warmth in the relationship and higher scores in *conflict* indicate a higher degree of problem interactions in the relationship. The *closeness* and *conflict* subscales have been demonstrated to have high internal consistency ($\alpha = .86$ for *closeness* and $.92$ for *conflict*) and test-retest reliability (4-week $r = .88$ for *closeness* and $.92$ for *conflict*) (Pianta, 2001). In addition, these two subscales have

been shown to have good concurrent and discriminant validity in Norway and have been shown to have good factor validity in a slightly modified version (Drugli & Hjemdal, 2013; Solheim et al., 2012).

Caregiver-Teacher Report Form (C-TRF)

Teachers also completed the C-TRF (Achenbach & Rescorla, 2000), which contains 100 items describing problem behaviors for children who are between 1.5 and 5 years old. Each item has three response options: “not true (as far as you know)”, “somewhat or sometimes true”, and “very often or often true”. These answers correspond to scores from zero to two. The C-TRF contains the following subscales: *emotionally reactive* (7 items), *anxious/depressed* (8 items), *withdrawn* (10 items), *somatic complaints* (7 items), *attention problems* (9 items), and *aggressive behavior* (25 items). A total problem score (ranging from zero to 200) can be calculated by adding the scores across all items. In addition, two broadband scales can be calculated by adding certain subscales for *internalizing problems* (*emotionally reactive, anxious/depressed, withdrawn, and somatic complaints*) and *externalizing problems* (*attention problems and aggression problems*). The validity, reliability, and factor structure of the C-TRF have demonstrated to be excellent across cultures (de Groot et al., 1994; Ivanova et al., 2007; Ivanova et al., 2010; Ivanova et al., 2011; Koot et al., 1997; Rescorla et al., 2012; Rescorla et al., 2014; Verhulst & Koot, 1992). A score at or above the 90th percentile defines the clinical range on the C-TRF total problem score and has been shown to discriminate well between referred and non-referred children (Achenbach & Rescorla, 2000). In addition, the parent-reported counterpart to the C-TRF, the Child Behavior Checklist (CBCL), has been reported to show good correspondence and predictive validity to DSM diagnoses both for preschool children and older children (Bellina et al., 2013; Ebesutani et al., 2010; Krol et al., 2006; de la Osa et al., 2016).

We defined children with a score at or above the 90th percentile on the C-TRF's *Total Problem, Internalizing, or Externalizing* scale as having a clinical level of mental health problems. In addition, children in the top 2% on at least one C-TRF subscale (except *somatic complaints*, as this scale does not match the concern specification options in the nomination process and, thus, would have created mismatched data) but who were not rated in the clinical range (90th percentile) on the C-TRF broader scales were also considered to be at risk. Consistent with recommendations by Achenbach and Rescorla (2000), this was done because the subscales comprise a smaller and more homogeneous set of problems, which are believed to require more stringent cutoff values to indicate a need for professional help. Thus, we ensured that children scoring very high on a specific set of problems were classified in the clinical range and in need of mental health services, even though they might have scored below the cutoff value on a broader scale. Because teachers tend to score boys higher than girls on the C-TRF (Achenbach & Rescorla, 2000; Kristensen et al., 2010), separate norms for girls and boys were used to establish gender-specific cutoffs. The cutoff values used are shown in Table 1.

Insert Table 1 about here

Statistical analyses

Before establishing cutoff values defining the clinical range on the C-TRF, one child was excluded due to missing age information. Based on teacher nominations of children with developmental concerns (246/1430 = 17%) and the C-TRF cutoff values described above (161/1430 = 11% in the clinical range), children were placed in one of the following categories: true positive (119/1430 = 8%), false positive (127/1430 = 9%), true negative (1142/1430= 80%), and false negative (42/1430 = 3%) (Table 1).

Insert Table 2 about here

Independent sample t-tests (equal variance not assumed) were carried out to test for group differences between the true negative cases and the false positives cases, followed by two-level (children nested within preschool teachers) binominal logistic regression analyses to investigate the covariates of group membership for the false positive (target group, n=127) compared to the true negative (reference group, n=1142). None of the 1269 total children in these two groups had missing data, and all were thus included in the current study. Due to the relatively low number of false positive cases, the number of covariates were limited to children's age and gender, *S-TRS conflict* score, *S-TRS closeness* score, *C-TRF internalizing problems* score, and *C-TRF externalizing problems* score. The covariates were entered in the analytic model in three blocks to yield unadjusted (single covariate entry), adjusted (single covariate adjusted for children's age and gender), and fully adjusted (full model with all covariates) odds ratio (OR) estimates. The analyses were performed using SPSS25 and STATA16.

Results

The true negative group contained 50% (569/1142) boys, while the false positive group contained 60% (76/127) boys. The mean age for the true negatives was 3.70 years and the mean age for the false positives was 4.25 years. Both gender ($p=.03$) and age ($p<.001$) yielded significant group differences. In general, as seen in Table 3, the false positive group had a significantly higher mean score on all covariates except the closeness scale, which was significantly lower, all indicating more negative evaluations for this group, as well as larger variation. In addition, when comparing the false positive group's mean scores for internalizing ($M=5.76$) and externalizing ($M=8.54$) problems (Table 3) with the clinical cutoff values for the same scales (Table 1), the estimates indicate that, on average, the false positive cases would have to display approximately twice the symptom load to approach a clinical level on the C-TRF.

Insert Table 3 about here

Table 4 shows that all covariates in the unadjusted and age- and gender adjusted analyses were significantly associated with group membership, while male gender and conflict were non-significant covariates in the fully adjusted analyses. With the exception of closeness, all covariates (age, male gender, conflict, internalizing problems, and externalizing problems) showed ORs greater than one, indicating as these covariates increased, the chance of being classified as a false positive increased. For closeness, the OR was significant and less than one, indicating that as closeness increased the chance of being classified as a false positive decreased. More specifically, the fully adjusted analysis revealed that for each year of age, the risk of being classified as a false positive increased by 68%, and a one-unit increase in internalizing or externalizing problem score was associated with a 55% and 21% increased risk of being classified as false positive, respectively. A one unit increase on the closeness

scale was associated with an 8% decrease in the risk for false positive classification. Male gender was associated with a 1% decrease in the risk of being classified as a false positive, while a one-unit increase on the conflict scale was associated with a 2% increase in the risk of being classified as a false positive. Neither gender nor conflict were significant covariates in the fully adjusted analysis.

Insert Table 4 about here

Discussion

The current study aimed to investigate, in essence, how false a false positive teacher nomination is for preschool children at risk for mental health problems and to compare the characteristics of those classified as false positive with those classified as true negative. In support of our initial hypotheses, our findings indicate that children in the normal range of the C-TRF who were nominated by preschool teachers with developmental concerns (false positive) were reported to have significantly higher internalizing and externalizing problem scores compared with children who not were nominated. Children identified as false positive cases were also perceived by teachers to have poorer teacher-child relationships, have higher levels of conflict, and have lower levels of teacher-child closeness. Neither child's gender nor degree of conflictual teacher-child relationship was significantly associated with false positive classification when adjusting for other factors. Age, internalizing problems, and externalizing problems increased the risk of false positive classification, while increased closeness. Closeness in the teacher-child relationship reduced the risk of false positive classification.

Children's internalizing and externalizing problems and teacher-child relationships

The findings of the current study support results from previous research on school-aged children, showing that teacher-nominated children differ significantly from their non-nominated peers (Layne et al., 2006; Ollendick et al., 1990; Roeser & Midgley, 1997). Although not reaching the clinical cutoff, the teacher-nominated preschool children still received higher scores for internalizing and externalizing problems, and lower scores for teacher-child relationship quality compared to their non-nominated peers (Table 4). These results suggest that even when the teachers' concerns are classified as false positives and referrals for a more thorough assessment appear to be unnecessary, children may still be in

need of extra monitoring and support to ensure that development proceeds normally. Contrary to previous research (Loades & Mastroyannopoulou, 2010), internalizing rather than externalizing problems made teachers more likely to report concerns about children's development, at least for children in the non-clinical range of the C-TRF. Although preschool teachers tend to underreport internalizing problems (Berg-Nielsen et al., 2012), they may be more vigilant in raising concerns when asked specifically to report this type of problem. If they have more experience dealing with externalizing problems than internalizing problems, a lower threshold for deviancy from what they perceive as normal behavior in the internalizing domain may occur, thus causing them to raise concerns more readily when dealing with children with internalizing problems.

As seen in Table 4, characteristics of the children themselves (i.e., age, internalizing problems, externalizing problems) were the main factors that increased the odds of teacher nomination, while teachers' perception of teacher-child relationships did not have the same impact. This may be because positive relationships can and do develop even in the presence of problem behavior (Myers & Pianta, 2008). For each one-unit increase in the closeness scale, the odds of false positive classification were reduced significantly by 8%, indicating that teachers' proximity is important when teachers raise concerns. It seems plausible that with increased closeness, teachers are more able to accurately assess development and thus provide more reliable information. Children with externalizing problems are more likely to develop conflicting relationships with their teachers, which may lead to a maladaptive spiral (Sabol & Pianta, 2012). This reciprocal relationship has been found in several studies (e.g., Skalická et al., 2015; Zhang & Sun, 2011). Internalizing problems also exhibit the same reciprocal relationship with conflict, but neither externalizing nor internalizing problems show this bidirectional effect for teacher-child closeness (Zhang & Sun, 2011). These findings may indicate that two distinct mechanisms are involved in these two relationship

dimensions. It may be that preschool teachers' perception of conflict is child-driven, while the perception of closeness is more teacher-driven (Silver et al., 2005). This could explain the stronger link between problem behavior and conflict, while factors such as teachers' self-efficacy may play a more important role in teacher-child closeness. In this study, the false positive group received significantly higher scores in internalizing and externalizing problems and conflict as well as lower scores in closeness than the true negative group. Even though the reciprocity of children's problem behaviors and teacher-child conflict has been demonstrated in previous studies (Skalická et al., 2015; Zhang & Sun, 2011), conflictual relationships do not seem to be a source of concern for teachers, although closeness reduced slightly but significantly the odds for stating concern.

Children's age and gender

As seen in Table 4, even though boys are more prone to false positive classification, this association was not significant when adjusting for other covariates, such as internalizing and externalizing problems. Although preschool teachers tend to report boys with more problem behaviors than girls on the C-TRF (Achenbach & Rescorla, 2000; Kristensen et al., 2010), it does not seem to bias teacher concerns when children fall in the non-clinical range. As mentioned previously, different teacher thresholds for internalizing and externalizing problems may be in play when teachers raise developmental concerns and the same mechanism may also be at play regarding child gender. If teachers' perception of normal behavior differs for boys and girls and they operate with different thresholds for raising concern, it may result in girls and boys having the same odds for teacher nomination, even when boys display more symptoms. Future studies should investigate the interaction effect between gender and type of problems regarding teacher concerns and false positive classification.

Children's age was a significant predictor of teacher nomination for children in the non-clinical range of the C-TRF. Preschool teachers were more likely to raise concerns for older children than for younger children, making a false positive classification more likely with increased age. One explanation may be that teachers feel more capable of discriminating between normal and abnormal behavior for older children, as symptom expression in older children can be more distinguishable than the more subtle expressions in younger children. Thus, teachers may have a better reference base for normal rather than abnormal behavior when reporting concern. Another explanation may be that teachers perceive younger children to have more transient problems that are more likely to normalize before school entry. As children age and school entry approaches, teachers may grow increasingly concerned if there is a dissonance in the perception of developmental skills and school readiness. As the current study shows, false positive cases received significantly higher problem scores than the true negative cases, indicating that teachers' concerns should not be disregarded out-of-hand.

Some behaviors are more likely to be considered deviant or unusual as they raise concerns more easily. However, most behaviors are displayed on a continuum and depend on context, which makes the discrimination of normal and abnormal behavior more difficult to establish (Carter et al., 2004). In addition, internalizing and externalizing problems exhibit a low to modest correlation to functional impairment (Gordon et al., 2006; McKnight & Kashdan, 2009; McKnight et al., 2016). Thus, some children may display an elevated symptom load without significant impairment, while other children may display few symptoms and significant impairment. Consequently, blindly relying on categorical criteria may hinder the identification of children with emerging psychopathology. As children with sub-clinical levels of symptoms continue to display impairment later in life (Finsaas et al., 2018; Shankman et al., 2009), it is important to identify relevant cases below clinical cutoffs and among children not meeting diagnostic criteria. As symptoms of psychiatric disorders

reflect normal behaviors that change phenotypically as children develop, approaches that capture the full range of behaviors relevant to psychopathology are needed (Dougherty et al., 2015). A probabilistic approach to clinical cutoff values was proposed by Sheldrick and colleagues (2015) in which children with a very high symptom score are assumed to be more likely to have some psychopathology than children with a low score. In addition, children approaching a clinical cutoff would also be more likely to display some psychopathology than children with low scores. Thus, this approach assumes that increases in symptom scores indicate an increased probability of psychopathology. In support of Sheldrick et al. (2015), the current study reports that increases in internalizing and externalizing problems were associated with an increased likelihood of nomination by preschool teachers. Further, the nominated children were reported to have significantly more symptoms of problematic behavior compared to the children not nominated, indicating that children classified as false positive through preschool teachers' nominations should be developmentally monitored rather than regarded as completely behaviorally healthy. Our results and those of previous studies indicate that, although both categorical and dimensional approaches to psychopathology are capable of discriminating between normal and abnormal behaviors in preschoolers (Moreland & Dumas, 2008), dimensional approaches are better suited for monitoring developmental trajectories because of their flexibility (Coghill & Sonuga-Barke, 2012). Consequently, dimensional approaches to identify and intervene for sub-clinical problems should be initiated before more stable patterns of psychopathology emerge.

Clinical implications

It is reasonable to assume that a false positive classification may be more correct for some children and less correct for others. Further, better training for preschool teachers so that they can recognize what constitutes developmental concerns, might be expected to lower teacher nomination-associated misclassification rates. Findings from this study suggest that the term

false positive may in some cases be misleading or inaccurate, which could potentially hinder at-risk children getting appropriate help in a timely manner. When a preschool teacher raises concerns that a child may potentially have clinical problems, it is important and worthwhile to scrutinize the concern in a prompt manner even for children who are found to be in the non-clinical range following further tests. Further, preschool teachers should be encouraged to express their concerns once they arise, preferably in a forum which includes colleagues, parents, and other mental health professionals. The Norwegian Kindergarten Act (2017) states that preschool teachers are responsible for following up concerns they might have regarding a child's development. Thus, one approach may be to apply teacher concerns as a pre-screening method to identify children of interest, supplemented by a psychometrically sound screening instrument, which will either confirm or dismiss the concern. It may also be beneficial to monitor child development dimensionally by scrutinizing scores and/or establishing a symptom profile rather than blindly using categorical cutoffs. This may ensure that children with elevated (but non-clinical) problem scores are monitored and get help for their problems, which in turn may increase the likelihood of healthy development. The importance of this task is underlined by findings indicating that young children who display internalizing and externalizing problems at a sub-clinical level will continue to exhibit poorer functioning throughout childhood and adolescence (Finsaas et al., 2018). Since few preschool children are referred and receive treatment for existing mental health problems (Egger & Angold, 2006; Horwitz et al., 2003; Horwitz et al., 2007; Wichstrøm et al., 2014), scrutinizing preschool teachers' concerns may lead to an increase in referral rate. Currently, it is known that parents are a strong trigger for the initiation of contact with health services for young children not meeting developmental expectations within the family context (Ellingson et al., 2004). The results of the current study and previous studies suggest that preschool teachers may play a similar role within the context of childcare centers, as their concerns

seem to capture a considerable portion of children with a clinical symptom load (Stensen et al., 2021), as well as those with a sub-clinical symptom load.

Strength and limitations

To the best of our knowledge, this study is the first to investigate false positive classification rates and the factors that predict teachers' developmental concerns in a large sample which includes the full age range of preschool children. Although this study was strengthened by its large sample size, an even larger sample to obtain more false positive cases would have been beneficial, allowing more covariates to be investigated. Future studies should investigate the interaction of the covariates examined in the current study to further illuminate the factors leading to teachers being concerned about children without obvious clinical problems. One limitation of the current study is that the ordering of measures in the survey may have increased the susceptibility to a confirmation bias. Preschool teachers nominated children prior to completing the C-TRF, which could have primed them to respond to the survey differently than if the nomination item was dropped or maybe located elsewhere in the survey. Although the C-TRF might be regarded as a "gold standard" for measuring children's mental health problems based on its psychometric properties, it does not exhibit perfect accuracy compared with diagnostic interviews (Lavigne et al., 2016b). Future studies could investigate preschool teachers' concerns against classifications from a diagnostic interview, thus not solely relying on one informant. The inclusion of other informants, such as parents, would have been beneficial when investigating the nomination method. As preschool teachers provide one perspective, future studies should include other informants to investigate the psychometric indices of the nomination method, as the "gold standard" for clinical decisions in developmental psychopathology use multi-informant reporting. It would also be valuable to investigate the relationship between preschool teachers' concerns and measures of functional impairment, rather than symptom scales isolated. Finally, as this study is cross-

sectional and represents only a snapshot, an important focus for future research should be to longitudinally investigate whether scrutinizing preschool teachers' concerns, preferably in combination with a psychometrically sound screening instrument and in collaboration with parents and other mental health professionals, actually leads to increased support for children displaying an elevated level of mental health problems. As children's mental health problems may be a precursor for emerging psychopathology, early identification and support can be of great importance to ensure their healthy development.

Conclusion

The results of the current study emphasize the importance of seriously considering preschool teachers concerns about children in their care, as even children classified as false positive by teacher nomination displayed significantly poorer outcomes compared with true negative cases. Because an elevated level of mental health problems and a decreased quality of teacher-child relationship may be precursors to emerging psychopathology and later maladjustment, it is important to scrutinize concerns when they arise and supplement with a psychometrically sound screening instrument that may confirm or dismiss the initial concern. Even in cases where follow-up instruments dismiss the initial concern, scrutinizing the preschool teachers' concern in collaboration with others may still reveal that the child needs support to reduce the level of problem behavior and thus increase the likelihood of healthy development.

Compliance with Ethical Standards

Conflict of Interests: None of the authors have declared any competing or potential conflicts of interest.

Funding: No external funding was received as part of this work.

Ethical Approval: The study was approved by the Regional Committee for Medical and Health Research Ethics (REK) (reference number: 2011/2252; 2012/1030; 2013/1363)

Informed Consent: Informed consent was obtained from all parents and teachers of included children.

Experiment Participants: All procedures performed in studies involving human participants were conducted in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

References

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA Preschool Forms & Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Angold, A., Costello, E. J., Farmer, E. M. Z., Burns, B. J., & Erkanli, A. (1999). Impaired but Undiagnosed. *Journal of the American Academy of Child & Adolescent Psychiatry*. 38: 129-137

Basten, M., Tiemer, H., Althoff, R. R., van de Schoot, R., Jaddoe, V. W. V., Hofman, A., ..., & van der Ende, J. (2016). The Stability of Problem Behavior Across the Preschool years: An Empirical Approach in the General Population. *Journal of Abnormal Child Psychology*. 44: 393-404

Belfer, M.L. (2008). Child and adolescent mental disorders: The magnitude of the problem across the globe. *Journal of Child Psychology and Psychiatry*. 49: 226-236

Bellina, M., Brambilla, P., Garzitto, M., Negri, G. A. L., Molteni, M., & Nobile, M. (2013). The ability of CBCL DSM-oriented scales to predict DSM-IV diagnoses in a referred sample of children and adolescent. *European Child and Adolescent Psychiatry*. 22: 235-246

Berg-Nielsen, T.S., Solheim, E., Belsky, J., & Wichstrøm, L. (2012). Preschoolers' Psychosocial Problems: In the Eyes of the Beholder? Adding Teacher Characteristics as Determinants of Discrepant Parent-Teacher Reports. *Child Psychiatry and Human Development*. 43: 393-413

Berkhout, L., Hoekman, J., & Goorhuis-Brouwer, S. M. (2012). Psychosocial health of 4-year-old boys and girls, as observed by parents and teachers. *European Journal of Developmental Psychology*. 9: 500-507

Briggs-Gowan, M. J. & Carter, A. S. (2008). Social-Emotional Screening Status in Early Childhood Predicts Elementary School Outcomes. *Pediatrics*. 121: 957-962

Carter, A. S., Briggs-Gowan, M. J., & Davis, N. O. (2004). Assessment of young children's development and psychopathology: recent advances and recommendations for practice. *Journal of Child Psychology and Psychiatry*. 45: 109-134

Center on the Developing Child at Harvard University (2010). The Foundations of Lifelong Health Are Built in Early Childhood. Retrieved from <http://www.developingchild.harvard.edu>

Coghill, D. & Sonuga-Barke, E. J. S. (2012). Annual Research Review: Categories versus dimensions in the classification and conceptualisation of child and adolescent mental disorders – implications of recent empirical study. *Journal of Child Psychology and Psychiatry*. 53: 469-489

Council on Children With Disabilities (2006). Identifying Infants and Young Children With Developmental Disorders in the Medical Home: An Algorithm for Developmental Surveillance and Screening. *Pediatrics*. 118: 405-420

Dadds, M. R., Spence, S. H., Holland, D. E., Barrett, P. M., & Laurens, K. R. (1997). Prevention and early intervention for anxiety disorders: A controlled trial. *Journal of Consulting and Clinical Psychology*. 65: 627-635

Dougherty, L. R., Leppert, K. A., Merwin, S. M., Smith, V. C., Bufferd, S. J., & Kushner, M. R. (2015). Advances and Directions in Preschool Mental Health Research. *Child Development Perspectives*. 9: 14-19

Drugli, M. B. & Hjemdal, O. (2013). Factor Structure of the Student-Teacher Relationship Scale for Norwegian School-Aged Children Explored with Confirmatory Factor Analysis. *Journal of Educational Research*. 57: 457-466

Dwyer, S. B., Nicholson, J. M., & Battistutta, D. (2006). Parent and teacher identification of children at risk of developing internalizing or externalizing mental health problems: A comparison of screening methods. *Prevention Science*. 7: 343-357

Ebesutani, C., Bernstein, A., Nakamura, B. J., Chorpita, B. F., Higa-McMillan, C. K., Weisz, J. R., & the Research Network on Youth Mental Health (2010). Concurrent Validity of the Child Behavior Checklist DSM-Oriented Scales: Correspondence with DSM Diagnoses and Comparison to Syndrome Scales. *Journal of Psychopathology and Behavioral Assessment*. 32: 373-384

Egger, H. L. & Angold, A. (2006). Common emotional and behavioral disorders in preschool children: presentation, nosology, and epidemiology. *Journal of Child Psychology and Psychiatry*. 47: 313-337

Eklund, K., Renshaw, T. L., Dowdy, E., Jimerson, S. R., Hart, S. R., Jones, C. N., & Earhart, J. (2009). Early Identification of Behavioral and Emotional Problems in Youth: Universal Screening versus Teacher-Referral Identification. *The California School Psychologist*. 14: 89-95

Ellingson, K. D., Briggs-Gowan, M. J., Carter, A. S., & Horwitz, S. M. (2004). *Archives of Pediatric and Adolescent Medicine*. 158: 766-772

Essex, M. J., Kraemer, H. C., Slattery, M. J., Burk, L. R., Boyce, W. T., Woodward, H. R., & Kupfer, D. J. (2009). Screening for childhood mental health problems: outcomes and early identification. *Journal of Child Psychology and Psychiatry*. 50: 562-570

Fanti, K. A. & Henrich, C. C. (2010). Trajectories of Pure and Co-Occurring Internalizing and Externalizing Problems From Age 2 to Age 12: Findings From the National Institute of Child Health and Human Development Study of Early Child Care. *Developmental Psychology*. 46: 1159-1175

Finsaas, M. C., Kessel, E. M., Dougherty, L. R., Bufferd, S. J., Danzig, A. P, Davila, J., Carlson, G. A., & Klein, D. N. (2018). Early Childhood Psychopathology Prospectively Predicts Social Functioning in Early Adolescence. *Journal of Clinical Child & Adolescence Psychology*. DOI: <https://doi.org/10.1080/15374416.2018.1504298>

Gilliom, M. & Shaw, D. S. (2004). Codevelopment of externalizing and internalizing problems in early childhood. *Development and Psychopathology*. 16: 313-333

Glascoe, F. P. (2001). Are Overreferrals on Developmental Screening Tests Really a Problem? *Archives of Pediatric and Adolescent Medicine*. 155: 54-59

Glascoe, F. P. & Marks, K. P. (2011). Detecting children with developmental-behavioral problems: The value of collaborating with parents. *Psychological Test and Assessment Modeling*. 53: 258-279

Gordon, M., Antshel, K., Faraone, S., Barkley, R., Lewandowski, L., Hudziak, J. J., ..., & Cunningham, C. (2006). Symptoms Versus Impairment: The Case of Respecting DSM-IV's Criterion D. *Journal of Attention Disorders*. 9: 465-475

de Groot, A., Koot, H. M., & Verhulst, F. C. (1994). Cross-Cultural Generalizability of the Child Behavior Checklist Cross-Informant Syndromes. *Psychological Assessment*. 6: 225-230

Horwitz, S. M., Gary, L. C., Briggs-Gowan, M. J., & Carter, A. S. (2003). Do Needs Drive Service Use in Young Children? (2003). *Pediatrics*. 112: 1373-1378

Horwitz, S. M., Kelleher, K. J., Stein, R. E. K., Storfer-Isser, A., Youngstrom, E. A., Park, E. R., ..., & Hoagwood, K. E. (2007). Barriers to the Identification and Management of Psychosocial Issues in Children and Maternal Depression. *Pediatrics*. 119: 208-218

Ivanova, M. Y., Achenbach, T. M., Rescorla, L. A., Dumenci, L., Almqvist, F., Bathiche, M., ..., & Verhulst, F. (2007). Testing the Teacher's Report Form syndromes in 20 societies. *School Psychology Review*. 36: 468-483

Ivanova, M. Y., Achenbach, T. M., Rescorla, L. A., Harder, V. S., Ang, R. P., Bilenberg, N, ... & Verhulst, F. C. (2010). Preschool Psychopathology Reported by Parents in 23 Societies: Testing the Seven-Syndrome Model of the Child Behavior Checklist for Ages 1,5-5. *New Research*. 49: 1215-1224

Ivanova, M. Y., Achenbach, T. M., Rescorla, L. A., Bilenberg, N., Bjarnadottir, G., Denner, S., ..., & Verhulst, F. C. (2011). Syndromes of Preschool Psychopathology Reported by Teachers and Caregivers in 14 Societies Using the Caregiver-Teacher Report Form (C-TRF). *Journal of Early Childhood and Infant Psychology*. 7: 87-103

Jensen, P. S. & Watanabe, H. (1999). Sherlock Holmes and Child Psychopathology Assessment Approaches: The Case of the False-Positive. *Journal of American Academy of Child and Adolescent Psychiatry*. 38: 138-146

Keenan, K., Shaw, D., Delliquadri, E., Giovannelli, J., & Walsh, B. (1998). Evidence for the Continuity of Early Problem Behaviors: Application of a Developmental Model. *Journal of Abnormal Child Psychology*. 26: 441-454

Koot, H., van den Oord, J., Verhulst, F. C., & Boomsma, D. (1997). Behavioral and emotional problems in young preschoolers: Cross-cultural testing of the validity of the Child Behavior Checklist/2-3. *Journal of Abnormal Child Psychology*. 25: 183-196

Kristensen, S., Henriksen T. B., & Bilenberg, N. (2010). The Child Behavior Checklist for Ages 1.5-5 (CBCL/1½-5): Assessment and analysis of parent and caregiver-reported problems in a population-based sample of Danish preschool children. *Nordic Journal of Psychiatry*, 64: 203-209

Krol, N. P. C. M., De Bruyn, E. E. J., Coolen, J. C., & van Aarle, E. J. M. (2006). From CBCL to *DSM*: A Comparison of Two Methods to Screen for *DSM-IV* Diagnoses Using CBCL Data. *Journal of Clinical Child and Adolescent Psychology*. 35: 127-135

Lavigne, J. V., Feldman, M., & Meyers, K. M. (2016a). Screening for Mental Health Problems: Addressing the Base Rate Fallacy for a Sustainable Screening Program in Integrated Primary Care. *Journal of Pediatric Psychology*. 41: 1081-1090

Lavigne, J. V., Meyers, K. M., & Feldman, M. (2016b). Systematic Review: Classification Accuracy of Behavioral Screening Measures for Use in Integrated Primary Care Settings. *Journal of Pediatric Psychology*. 41: 1091-1109

Lavigne, J. V., Arend, R., Rosenbaum, D., Binns, H. J., Christoffel, K. K., & Gibbons, R. D. (1998). Psychiatric Disorders With Onset in the Preschool Years: I. Stability of Diagnoses. *Journal of American Academy of Child and Adolescent Psychiatry*. 37: 1246-1254

Layne, A. E., Bernstein, G. A., & March J. S. (2006). Teacher Awareness of Anxiety Symptoms in Children. *Child Psychiatry and Human Development*. 36: 383-392

Loades, M. E. & Mastroyannopoulou, K. (2010). Teachers' Recognition of Children's Mental Health Problems. *Child and Adolescent Mental Health*. 15: 150-156

Masten, A. S. & Cicchetti, D. (2010). Developmental cascades. *Development and Psychopathology*. 22: 491-495

McKnight, P. E. & Kashdan, T. B. (2009). The importance of functional impairment to mental health outcomes: A case for reassessing our goals in depression treatment research. *Clinical Psychology Review*. 29: 243-259

McKnight, P. E., Monfort, S. S., Kashdan, T. B., Blalock, D. V., & Calton, J. M. (2016). Anxiety symptoms and functional impairment: A systematic review of the correlation between the two measures. *Clinical Psychology Review*. 45: 115-130

Moor, S., Sharrock, G., Scorr, J., McQueen, H., Wrate, R., Cowan, J., & Blair, C. (2000). Evaluation of a teaching package designed to improve teachers' recognition of depressed pupils - a pilot study. *Journal of Adolescence*. 23: 331-342

Moreland, A. D. & Dumas, J. E. (2008). Categorical and dimensional approaches to the measurement of disruptive behavior in the preschool years: A meta-analysis. *Clinical Psychology Review*. 28: 1059-1070

Myers, S. S. & Pianta, R. C. (2008). Developmental Commentary: Individual and Contextual Influences on Student-Teacher Relationships and Children's Early Problem Behaviors. *Journal of Clinical Child & Adolescent Psychology*. 37: 600-608

Njoroge, W. F. M. & Bernhart, K. P. (2011). Assessment of Behavioral Disorders in Preschool-Aged Children. *Current Psychiatry Reports*. 13: 84-92

Norwegian Kindergarten Act (2017). Framework Plan for Kindergartens. Retrieved from: <https://www.udir.no/globalassets/filer/barnehage/rammeplan/framework-plan-for-kindergartens2-2017.pdf>

Ollendick, T. H., Greene, R. W., Weist, M. D., & Oswald, D. P. (1990). The Predictive Validity of Teacher Nominations; A Five-Year Followup of At-Risk Youth. *Journal of Abnormal Child Psychology*. 18: 699-713

de la Osa, N., Granero, R., Trepát, E., Domenech, J. M., & Ezpeleta, L. (2016). The discriminative capacity of CBCL/1½-5-DSM5 scales to identify disruptive and internalizing disorders in preschool children. *European Child and Adolescent Psychiatry*. 25: 17-23

Pianta, R. C. (2001). *STRS Student-Teacher Relationship Scale (professional manual)*. Lutz, FL: Psychological Assessment Resources, Inc.

Polanczyk, G. V., Salum, G. A., Sugaya, L. S., Caye, A., & Rohde, L. A. (2015). Annual Research Review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *Journal of Child Psychology and Psychiatry*. doi:10.1111/jcpp.12381

Poulou, M. S. (2015). Emotional and Behavioural Difficulties in Preschool. *Journal of Child and Family Studies*. 24: 225-236

Radecki, L., Sand-Loud, N., O'Connor, K. G., Sharp, S., & Olson, L. M. (2011). Trends in the Use of Standardized Tools for Developmental Screening in Early Childhood: 2002-2009. *Pediatrics*. 128: 14-19

Rescorla, L. A., Achenbach, T. M., Ivanova, M. Y., Bilenberg, N., Bjarnadottir, G., Denner, S., ... & Verhulst, F. C. (2012). Behavioral/Emotional Problems of Preschoolers: Caregiver/Teacher Reports From 15 Societies. *Journal of Emotional and Behavioral Disorders*. 20: 68-81

Rescorla, L. A., Boichichio, L., Achenbach, T. M., Ivanova, M. Y., Almqvist, F., Begovac, I., ... & Verhulst, F. C. (2014). Parent-Teacher Agreement on Children's Problems in 21 Societies. *Journal of Clinical Child and Adolescent Psychology*. 43: 627-642

Roeser, R. W., & Midgley, C. (1997). Teachers' views of issues involving students' mental health. *The Elementary School Journal*, 2, 115-133

Sabol, T. J. & Pianta, R. C. (2012). Recent trends in research on teacher-child relationships. *Attachment and Human Development*. 14: 213-231

Sawyer, A. C. P., Chittleborough, C. R., Lynch, J. W., Baghurst, P., Mittinty, M. N., Kaim, A. L. E., & Sawyer, M. G. (2013). Can screening 4-5 year olds accurately identify children who will have teacher reported mental health problems when children are aged 6-7 years? *Australian & New Zealand Journal of Psychiatry*. 00: 1-10

Shankman, S. A., Lewinsohn, P. M., Klein, D. N., Small, J. W., Seeley, J. R., & Altman, S. E. (2009). Subthreshold conditions as precursors for full syndrome disorders: a 15-year longitudinal study of multiple diagnostic classes. *Journal of Child Psychology and Psychiatry*. 50: 1485-1494

Sheldrick, R. C., Benneyan, J. C., Kiss, I. G., Briggs-Gowan, M. J., Copeland, W., & Carter, A. S. (2015). Threshold and accuracy in screening tools for early detection of psychopathology. *The Journal of Child Psychology and Psychiatry*. 56: 936-948

Silver, R. B., Measelle, J. R., Armstrong, J. M., & Essex, M. J. (2005). Trajectories of classroom externalizing behavior: Contributions of child characteristics, family characteristics, and the teacher-child relationship during the school transition. *Journal of School Psychology*. 43:39-60

Skalická, V., Stenseng, F., & Wichstrøm, L. (2015). Reciprocal relations between student-teacher conflict, children's social skills and externalizing behavior: A three wave longitudinal study from preschool to third grade. *International Journal of Behavioral Development*. 39: 413-425

Skogen, J. C., Kjeldsen, A., Knudsen, A. K., Myklestad, I., Nesvåg, R., Reneflot, A., & Major, E. (2014). *Psykisk helse hos barn og unge* (Folkehelse rapporten 2014:4). Oslo: Norwegian Institute of Public Health

Soles, T., Bloom, E. L., Heath, N. L., & Karagiannakis, A. (2008). An exploration of teachers' current perceptions of children with emotional and behavioural difficulties.

Emotional and Behavioural Difficulties. 13: 275-290

Solheim, E., Berg-Nielsen, T. S., & Wichstrøm, L. (2012). The Three Dimensions of the Student-Teacher Relationship Scale: CFA Validation in a Preschool Sample. *Journal of Psychoeducational Assessment*. 30: 250-263

Stensen, K., Lydersen, S., Stenseng, F., Wallander, J. L., Tveit, H., & Drugli M. B. (2021). Childcare Providers' Nominations of Preschool Children At Risk for Mental Health Problems: Does it Discriminate Well Compared to the Caregiver-Teacher Report Form (C-TRF)? *Scandinavian Journal of Educational Research*.

DOI: 10.1080/00313831.2020.1869089

Sveen, T. H., Berg-Nielsen, T.S., Lydersen, S., & Wichstrøm, L. (2013). Detecting Psychiatric Disorders in Preschoolers: Screening With the Strengths and Difficulties Questionnaire. *Journal of American Academy of Child and Adolescent Psychiatry*. 52: 728-736

Verhulst, F. C., & Koot, H. M. (1992). *Child psychiatric epidemiology: concepts, methods and findings*. Beverly Hills: Sage Publications

Wichstrøm, L., Berg-Nielsen, T.S., Angold, A., Egger, H.L., Solheim, E. & Sveen, T.H. (2012). Prevalence of psychiatric disorders in preschoolers, *Journal of Child Psychology and Psychiatry*. 53: 695–705

Wichstrøm, L., Belsky, J., Jozefiak, T., Sourander, A. & Berg-Nielsen T. S. (2014). Predicting Service Use for Mental Health Problems Among Young Children. *Pediatrics*. 133: 1054-1060

Young, N. D. & Takala, C. R. (2018). Sequential Screening to Improve Behavioral Health Needs in Primary Care. *Journal of American Child and Adolescent Psychiatry*. 57: 603-609

Zhang, X. & Sun, J. (2011). The Reciprocal Relations Between Teachers' Perception of Children's Behavior Problems and Teacher-Child Relationships in the First Preschool Year. *The Journal of Genetic Psychology*. 172: 176-198

Table 1. Sample-specific cutoff values used on the C-TRF to define the clinical range (N=1430)

Scale (Percentile cutoff)	Girls	Boys
Total Problems (90%)	27	33
Internalizing (90%)	9	10
Externalizing (90%)	14	18
Emotionally Reactive (98%)	5	6
Anxious/Depressed (98%)	6	6
Withdrawn (98%)	7	8
Attention (98%)	10	12
Aggression (98%)	20	26

Table 2. Frequency overview of children’s risk status and teacher nominations (N=1430)

		Teacher Nomination Screening	
Criterion		Positive	Negative
C-TRF Clinical Range	Positive	TP, 119 (8%)	FN, 42 (3%)
	Negative	FP, 127 (9%)	TN, 1142 (80%)

Note: TP=true positive, FP=false positive, FN=false negative, TN=true negative

Table 3. Continuous covariates

Covariate (min–max)	True negative (n=1142)	False positive (n=127)
Conflict (8–44)	16.35 (4.17)	20.17 (6.48)
Closeness (22–55)	42.71 (5.29)	39.17 (5.95)
Internalizing problems (0–17)	2.00 (2.30)	5.76 (3.60)
Externalizing problems (0–23)	3.50 (4.33)	8.54 (5.71)

All values are mean (SD); the group difference was significant at $p < .001$ for all covariates.

Table 4. Associations of covariates with false positive vs. true negative classification based on teacher nominations of children with developmental concerns

Covariate	Unadjusted OR (CI)	Age- and gender-adjusted OR (CI)	Fully adjusted OR (CI) ^a
Age (years)	1.39 (1.18–1.64) <i>p</i> <.001	1.39 (1.19–1.64) <i>p</i> <.001	1.68 (1.30–2.17) <i>p</i> <.001
Male gender	1.49 (1.01–2.20) <i>p</i> =.04	1.52 (1.03–2.24) <i>p</i> =.04	.99 (.58–1.71) <i>p</i> =.98
Conflict	1.17 (1.12–1.22) <i>p</i> <.001	1.18 (1.13–1.23) <i>p</i> <.001	1.02 (.96–1.09) <i>p</i> =.55
Closeness	.87 (.84–.91) <i>p</i> <.001	.87 (.84–.91) <i>p</i> <.001	.92 (.87–.97) <i>P</i> =.002
Internalizing problems	1.64 (1.49–1.79) <i>p</i> <.001	1.63 (1.49–1.79) <i>p</i> <.001	1.55 (1.40–1.72) <i>p</i> <.001
Externalizing problems	1.22 (1.17–1.27) <i>p</i> <.001	1.24 (1.19–1.30) <i>p</i> <.001	1.21 (1.14–1.29) <i>p</i> <.001

Note: OR=odds ratio (significant associations in bold); CI=95% confidence interval

^aAdjusted for all covariates in first column

Paper III

Screening for mental health problems in a Norwegian preschool population. A validation of the ages and stages questionnaire: Social-emotional (ASQ:SE)

Kenneth Stensen¹, Frode Stenseng¹, Stian Lydersen¹, Marianne Suzanne de Wolff², Jan Wallander^{1,3} & May Britt Drugli^{1,4}

¹Regional Centre for Child and Youth Mental Health and Child Welfare, Faculty of Medicine and Health Science, Department of Mental Health, RKBU Midt-Norge, Norwegian University of Science and Technology (NTNU), Postboks 8905 MTF5, 7491, Trondheim, Norway. E-mail: kenneth.stensen@ntnu.no

²The Netherlands Organization for Applied Scientific Research (TNO), Quality of Life, Leiden, The Netherlands

³Merced - Psychological Sciences, School of Social Sciences, Humanities and Arts, and Health Sciences Research Institute, University of California, Merced, CA, USA

⁴Centre of the Study of Educational Practice, Hedmark University College, Elverum, Norway

Background: Early detection of mental health problems in childhood is important. However, studies on screening instruments for preschool children are rare. The aim of this study was to validate the Ages and Stages Questionnaire: Social-Emotional (ASQ:SE) with teacher reports and examine its screening accuracy in a preschool population. **Methods:** A total of 1428 children, aged 18 months – 5 years, attending child-care centers were recruited in Norway. Their teachers completed a survey including the ASQ:SE and the Caregiver-Teacher Report Form (C-TRF). The Spearman's correlation was calculated for the convergence between the ASQ:SE and the C-TRF and the screening accuracy of the ASQ:SE was assessed using receiver operating characteristic (ROC) analysis with the criterion of a score at or above the 90th percentile for the C-TRF total problem score. **Results:** The Spearman's correlation between the total scores for the ASQ:SE and the C-TRF were from .49 to .72. The ROC analyses demonstrated that the ASQ:SE had a promising ability to classify children at risk based on the C-TRF criterion with AUC ranging from .87 to .96 for the different forms. The ASQ:SE generally demonstrated high specificity across all forms and some forms (from age 30 months upwards) produced both high sensitivity and high specificity using the selected cutoff values. **Conclusions:** The ASQ:SE could serve as a good starting point for screening for social-emotional problems among children in child-care centers. The 30- to 60-month ASQ:SE forms exhibit promising psychometric properties and may prove useful for early detection. The 18- to 24-month ASQ:SE forms demonstrate more limited efficacy in detecting children at risk.

Key Practitioner Message

- Psychometrically sound screening instruments can contribute to early detection of children with mental health problems, however, validation and calibration should be performed within the same population as it is intended to be used.
- The short screening instrument ASQ:SE has previously only been validated with maternal reports. This study adds to the knowledge how it works with teacher reports in preschool settings.
- The ASQ:SE does capture much of the same social-emotional problems as the longer well-established C-TRF.
- The ASQ:SE forms 30 to 60 months shows good screening accuracy in detecting children at risk. However, the 18- and 24-month forms exhibit more limited efficacy and appear to be less useful.
- For this sample in general, a lowering of the cutoff values compared to those in the ASQ:SE manual would be beneficial to increase the rate of true positives.
- Findings from this study can guide practitioners and researchers on the use of the ASQ:SE in a preschool population.

Keywords: Screening; validity; mental health; psychometrics; preschool children

Introduction

Prevalence rates of mental disorder among preschool children have been estimated to range between 7% and 16% (Egger & Angold, 2006; Wichstrøm et al., 2012)

with mental health problems thought to be present in approximately 20% of preschool children (Belfer, 2008; Essex et al., 2009). Many childhood mental health problems are transient, but these problems are unlikely to remit for a portion of children. For some children, early

mental health problems have serious consequences for early learning, social development and even lifelong health (Center on the Developing Child at Harvard University, 2010). The development of screening procedures able to identify those in need of intervention is a major public health concern (Essex et al., 2009; Sawyer et al., 2013).

When mental health problems in the preschool period are left unidentified and untreated, they can negatively impact children's development and evolve into disorders (Dougherty et al., 2015; Feeney-Kettler, Kratochwill, Kaiser, Hemmeter, & Kettler, 2010). There is broad consensus that the early years are the optimal period for identifying children at risk for later serious mental health problems (Dougherty et al., 2015; Doyle, Harmon, Heckman, & Tremblay, 2009; Heckman, 2006; Heo & Squires, 2011; Kauffman, 1999; Poulou, 2015; de Wolff, Theunissen, Vogels, & Reijneveld, 2013). The enduring nature of untreated emotional and behavioral problems renders identification at an early stage critical to increase the probability of successful treatment (Dowdy, Chin, & Quirk, 2013). However, without psychometrically valid screening tools, children in need of early intervention may not be identified, referred and treated (Dougherty et al., 2015; Feeney-Kettler et al., 2010). Moreover, establishing norms for normal and abnormal development would permit a more thorough screening by capturing those who fall short of diagnostic criteria in the preschool period, but who still may be at risk for developmental impediments. Therefore, we require tools that detect the full spectrum of behaviors relevant to psychopathology and are applicable in different settings for the purpose of early detection (Dougherty et al., 2015; Sawyer et al., 2013).

A screening tool requires a well-established psychometric foundation, so that practitioners and researchers are certain about what they are measuring. Validity and reliability are important for the accurate interpretation of psychometric testing. Criterion reference validity is the degree of agreement between two instruments, in which one of the instruments is considered 'gold standard' based on well-established documentation. This may be a reasonable approach if the objective of the targeted instrument is to produce a shorter or simpler assessment and if the established instrument sets the standard to achieve (Fayers & Machin, 2007). In addition, it is important that a screening tool at community level exhibits good sensitivity and specificity. Sensitivity refers to an instrument's accuracy in identifying children at risk, whereas specificity refers to an instrument's accuracy in identifying children who are not at risk (Salomonsson & Sleed, 2010). The overidentification of false positives can result in wasted resources and possible stigmatization, whereas the overidentification of false negatives may deprive children of appropriate help (Sawyer et al., 2013).

Several instruments exist for measuring preschool children's behavioral development and mental health. The Achenbach System of Empirically Based Assessment (ASEBA) is the most widely used instrument internationally for assessing child psychological attributes and behavior both clinically and in research (Achenbach & Rescorla, 2000). The different instruments constituting this system have well documented psychometric properties across cultures (de Groot, Koot, & Verhulst,

1994; Ivanova et al., 2007; Ivanova et al., 2010; Koot, van den Oord, Verhulst, & Boomsma, 1997; Rescorla et al., 2012, 2014; Verhulst & Koot, 1992). The ASEBA instrument for children one and one-half to 5 years of age is administered to parents (Child Behavior Checklist:CBCL) and teachers (Caregiver-Teacher Report Form:C-TRF). However, the ASEBA is too long with 100 problem-related items to be used as a routine screening tool at the community level (de Wolff et al., 2013). The Ages and Stages Questionnaire: Social-Emotional (ASQ:SE) (Squires, Bricker, & Twombly, 2002) is an instrument developed for measuring social-emotional development in children aged six to 60 months and is frequently used in childcare centers. Thus far, the psychometric properties of ASQ:SE and its utility for screening has received less attention in the research field. With only 19-33 items, the ASQ:SE is considerably shorter than the ASEBA, it assesses children's development and behavior repertoire more broadly, and it is applicable to younger children.

The psychometric information about the ASQ:SE stems mainly from the developed manual based on a study of 3000 U.S. children aged three to 66 months old (Squires, Bricker, Heo, & Twombly, 2001; Squires et al., 2002). The ASQ:SE was found to have lower sensitivity in a Dutch toddler population than in the population reported in the ASQ:SE manual (de Wolff et al., 2013). This led to the conclusion that the ASQ:SE does not exhibit acceptable discrimination between children at risk and nonrisk at six and 14 months of age, although it displays somewhat better discrimination at 24 months of age. The CBCL was used as criterion in this study with the 90th percentile as cutoff to allocate whether children were within normal range or an elevated or clinical range. A validation of the ASQ:SE on a Swedish clinical sample displayed that the ASQ:SE mean scores were reflective of the clinical measurements (i.e. elevated risk), but the study also observed problems regarding concurrent validity, mainly that the mothers' own problems strongly predicted their responses on the ASQ:SE (Salomonsson & Sleed, 2010). In an adaptation of the ASQ:SE to Korean, the ASQ:SE exhibited adequate internal consistency and convergent validity against the Kongju Early Developmental Assessment System and the CBCL (Heo & Squires, 2011). A Dutch study (Theunissen, Vogels, de Wolff, Crone, & Reijneveld, 2015) reported weaker psychometric properties among a population of 1650 children aged three to 4 years old than those reported in the ASQ:SE manual. However, these authors reported better properties than those observed in the Korean study. These evaluations of the ASQ:SE highlight the need for further research into the instruments psychometric properties.

A review of classification accuracy (Lavigne, Meyers, & Feldman, 2016) identified four studies of the ASQ:SE that used the CBCL as criterion. The reported sensitivity (SE) and specificity (SP) values were: (a) children age 12-36 months (SE .93, SP .78), (b) children age 14 months (SE .56, SP .91) and 24 months (SE .84, SP .91), (c) children age 24 months (SE .95, SP .90) and (d) children age 24-48 months for clinical cutoff on the CBCL (85th percentile, SE .96, SP .87) and for the concerned cutoff (75th percentile, SE .80, SP .75). However, the age of the samples and the cutoff values applied varied between these studies. Another review by Velikonja et al. (2016)

concluded that the psychometric properties of the original versions of the ASQ:SE for children two to two and a half years generally exhibited good reliability, sensitivity and specificity against the CBCL comparator, whereas the adapted or translated versions of these age forms exhibited more mixed results. However, the cutoff values applied on the CBCL are not reported.

The aim of this study was to validate the ASQ:SE teacher report with a sample of children aged 18 months to 5 years in Norwegian child-care centers. Our research questions are as follows: (a) How do the different ASQ:SE forms correspond with the C-TRF? (b) How efficient are the different ASQ:SE forms in classifying children at risk for problem behavior and those who are not? Consequently, criterion reference validity and screening accuracy of the ASQ:SE will be investigated.

Methods

The data are from the Children in Central Norway (CCN) intervention study conducted to improve mental health among children in child-care centers. The data used in this study were collected from 2012 to 2014 before the intervention commenced.

Participants

In Norway, children typically begin at child-care centers when they are one to 2 years old. In 2016, 91% of the Norwegian children attended child-care centers (SSB, 2017). Of 2108 eligible children, a total of 1486 were recruited for wave 1 of the CCN study (consent 70.5%). The sample contained 51% boys and 49% girls. Children about whom the 12-month ASQ:SE was administered (3% of the sample) were excluded from this study because the criterion measure (the C-TRF) is appropriate for children from 18 months to 5 years of age. Fourteen children with age inappropriate administered ASQ:SE form were excluded, and one response was incomplete and removed from the data set, leaving $n = 1428$ of which 1395 children had both complete ASQ:SE (teacher report) and C-TRF data. No information was collected from those who did not consent.

Procedure

Parents with children in child-care centers from 18 months to 5 years of age in three municipalities in Central Norway received recruitment letters with information about the CCN study and a consent form. The recruitment letter also contained an invitation code to an online survey. Parents could provide consent for the study either by logging into the survey with their invitation code or by returning the consent form to the child's child-care center. Parental consent also gave the teacher in the child-care center who was most familiar with the child permission to complete a survey. Teachers provided consent electronically via the survey with their own invite codes. Participation was voluntary and parental consent could be withdrawn at any time without reprisal until the participation registry was deleted. The study was approved by the Regional Committee for Medical and Health Research Ethics.

Measures

Ages and Stages Questionnaire: Social-Emotional (ASQ:SE). The ASQ:SE is a brief parent- or teacher-reported instrument designed to assist in identifying developmental delays in children aged six to 60 months (Squires et al., 2002). Different forms are used depending on the child's age, and the number of scored items range from 19 (6 months) to 33 (48 and 60 months). The following age intervals are covered by the different forms: ASQ:SE 18 (15 to 20 months), ASQ:SE 24 (21 to 26 months), ASQ:SE 30 (27 to 32 months), ASQ:SE 36 (33 to 41 months), ASQ:SE 48 (42 to 53 months) and ASQ:SE 60 (54 to 65 months). There are three response options (rarely or

never, sometimes, most of the time) for each item, which are scored zero, five and ten with a possible additional five points if this specific behavior worries the informant. A total difficulty score is calculated by adding the points from all the items and the items related to expressed concerns. The cutoff scores provided by the manual vary by age and the alpha coefficients reported ranges from .80 to .91 for the 18- to 60-months forms. The following sensitivity (SE) and specificity (SP) pairs are reported: 18 months (SE 75%, SP 96.6%), 24 months (SE 70.8%, SP 93%), 30 months (SE 80%, SP 89.5%), 36 months (SE 77.8%, SP 93%), 48 months (SE 76.9%, SP 94.6%), 60 months (SE 84.6%, SP 95.8%) (Squires et al., 2002). As noted, other studies have reported somewhat lower sensitivity than the manual (Theunissen et al., 2015; de Wolff et al., 2013).

Caregiver-Teacher Report Form (C-TRF)

The C-TRF (Achenbach & Rescorla, 2000) contains 100 items describing problem behavior for children from 18 months to 5 years of age. Each item has three response options (not true (as far as you know), somewhat or sometimes true, very often or often true) that are scored from zero to two. A total problem score can be calculated by adding the item scores, which range from zero to 200. The validity, reliability and factor structure of C-TRF have been extensively tested across cultures with excellent psychometric properties (de Groot et al., 1994; Ivanova et al., 2007, 2010; Koot et al., 1997; Rescorla et al., 2012, 2014; Verhulst & Koot, 1992)

The sample means and standard deviations for total scores on the ASQ:SE and the C-TRF are presented in Table 1 and the percentiles in Table 2.

Statistical analysis

Caregiver-Teacher Report Form data on 1428 children were used to establish risk status. Subsequent analyses were performed with those who had both complete ASQ:SE and C-TRF data ($n = 1395$). The Spearman's correlation was used to evaluate the criterion reference validity of the ASQ:SE against the C-TRF as the criterion. The sensitivity, specificity and positive and negative predictive values (PPV and NPV) were calculated with receiver operating characteristic (ROC) analysis for the ASQ:SE using the criterion of a score at or above the 90th percentile for the C-TRF total problem score. PPV and NPV were calculated for the prevalence of 10%. To identify an appropriate cutoff value on the ASQ:SE, a criterion of specificity of at least 90% was established. The analyses were performed using SPSS 21.

Results

We initially present the correlation between the total scores of the ASQ:SE forms and the C-TRF, followed by the screening accuracy of the ASQ:SE forms on the C-TRF criterion.

Table 1. Descriptive information of the sample for Ages and Stages Questionnaire: Social-Emotional (ASQ:SE) and Caregiver Teacher Report Form (C-TRF) total problem score separated by age group (n ASQ:SE = 1395, n C-TRF = 1428)

	ASQ:SE		C-TRF	
	<i>n</i>	Mean (<i>SD</i>)	<i>n</i>	Mean (<i>SD</i>)
ASQ:SE 18	101	21.24 (22.02)	104	12.45 (13.72)
ASQ:SE 24	114	19.61 (17.69)	115	14.63 (15.33)
ASQ:SE 30	128	28.16 (25.64)	132	13.58 (14.79)
ASQ:SE 36	298	35.15 (34.04)	300	14.34 (15.36)
ASQ:SE 48	337	25.70 (27.84)	351	11.87 (12.95)
ASQ:SE 60	417	28.39 (31.53)	426	12.59 (14.10)

SD, standard deviation.

Table 2. Percentiles of the Ages and Stages Questionnaire: Social-Emotional (ASQ:SE) forms and the 90th percentile on the Caregiver Teacher Report Form (C-TRF)

Form	ASQ:SE					C-TRF 90th %
	n items	25th %	50th %	75th %	90th %	
ASQ:SE 18	26	10	15	25	49	29.5
ASQ:SE 24	26	5	15	30	45	34.2
ASQ:SE 30	29	10	20	40	55	26
ASQ:SE 36	31	10	25	45	90	37.9
ASQ:SE 48	33	5	20	35	60	27
ASQ:SE 60	33	5	20	35	70	30

Table 3. The Spearman’s correlation between Ages and Stages Questionnaire: Social-Emotional (ASQ:SE) and Caregiver Teacher Report Form (C-TRF) total problem scores

ASQ:SE form	n	The Spearman’s correlation	95% CI
18	101	.53	.38 to .66
24	114	.49	.33 to .63
30	128	.59	.46 to .70
36	298	.69	.61 to .76
48	337	.66	.58 to .72
60	417	.72	.66 to .77

Confidence intervals (CIs) are based on bootstrapping 10,000 samples (bias-corrected and accelerated).

Validity and screening accuracy

Table 3 presents the Spearman’s correlation between the total score on each ASQ:SE age form and the C-TRF, ranging from .49 to .72. Table 4 demonstrates that the area under curve (AUC) ranged from .87 to .96. The sensitivity and specificity pairs for each age group can also be seen in Table 4, showing the lowest sensitivity (50%) at 18 months and highest at 48 and 60 months (85%). The specificity for all the forms was equal to or above 90% (Figure 1).

Discussion

To our knowledge, this is the first study to investigate the validity and screening accuracy of the ASQ:SE based on teacher reports. Parents and teachers are the only viable source of information at the community level (Sveen, Berg-Nielsen, Lydersen, & Wichstrøm, 2013) and teacher reports are therefore important. Our main findings are that the ASQ:SE generally shows a good ability to discriminate between children who are at risk for mental health problems and those who are not, based on a well-established, widely used criterion.

Psychometric performance

The ASQ:SE exhibit good criterion reference validity against the C-TRF, especially for the 30- to 60-month forms with strong positive correlations ranging from .59 to .72 and a narrow confidence interval. The wider confidence interval for the 18- to 24-month forms is due to the smaller sample size. The lower correlation for younger ages may be explained by children at this age having a more limited behavior repertoire and teachers having less knowledge of normal and abnormal social-emotional development for younger children. Another

reason could be that the ASQ:SE forms for younger children have fewer items, which could reduce the correlations with the criterion, other things being equal.

Based on Hosmer and Lemeshow’s (2000) strength of discrimination (0.5 = no discrimination, 0.7 to 0.79 = acceptable, 0.8 to 0.89 = excellent, 0.9 to 1 = outstanding), Table 4 demonstrates that the AUCs of the different ASQ:SE forms have an excellent to outstanding ability to discriminate between at risk and low-risk based on the C-TRF criterion. Because the ASQ:SE produces high specificity across the ASQ:SE forms, it performs well in identifying the low-risk children, only producing about 1/10 false positives. However, the different forms exhibited mixed findings regarding sensitivity. Caution is particularly warranted for the use of the 18- and 24- month forms (50% and 64%, respectively). These two forms failed to identify children at risk at chance level or slightly above. The ASQ:SE forms for 30, 36, 48 and 60 months indicate that these forms are able to produce high sensitivity and specificity simultaneously using the given cutoff values. The positive predictive values are generally low for all the forms, as approximately half the children who were above cut-offs on the ASQ:SE were actually at risk at the 10% prevalence level. However, the negative predictive values are very high, indicating that a negative test on the ASQ:SE reflects development within normal parameters. It should be noted that predictive values are strongly influenced by prevalence; thus, more prevalent problems produce higher predictive values than less prevalent problems. Here, we predicted the top 10%, a relatively low prevalence. Populations with a higher prevalence of mental health problems would have obtained a higher PPV given the same sensitivity and specificity.

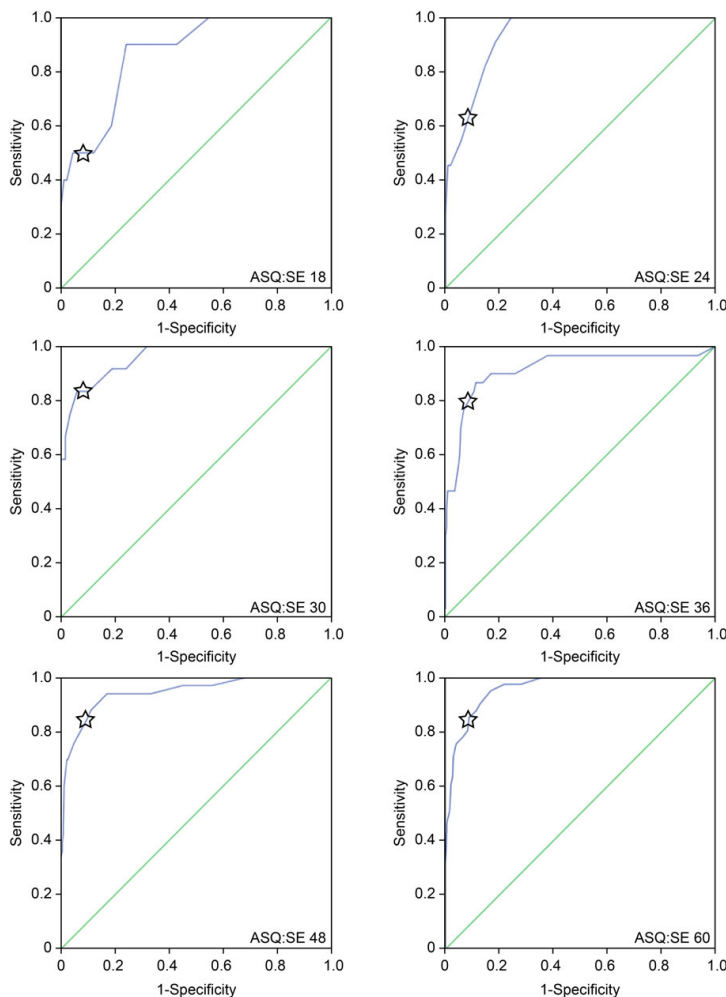
The criterion used in this study, the C-TRF, does not in itself provide a diagnosis. It is a questionnaire measurement tool that is simply longer and more detailed than the ASQ:SE. However, both C-TRF and CBCL are commonly included in the assessment battery by clinicians. The CBCL is often used as criterion for other screening instruments, but it does not exhibit a perfect 100% sensitivity and specificity against structured interviews/diagnosis (Lavigne et al., 2016). Future research should investigate the classification accuracy of the ASQ:SE compared to structured interviews and clinical diagnoses as well.

Another issue is the predictive validity. To our knowledge, no studies have been conducted for the ASQ:SE regarding individual stability over time, which is a crucial aspect of any screening instrument. However, some findings on the CBCL and the C-TRF exist. Kerr,

Table 4. Ages and Stages Questionnaire: Social-Emotional (ASQ:SE) screening performance against scores ≥ 90 percentile on the Caregiver Teacher Report Form (C-TRF)

ASQ:SE	<i>n</i>	AUC	Cutoff	Positives	Sens	Spec	Prevalence 10%	
							PPV	NPV
18	101	.87	37.5	14% (14/101)	50% (5/10)	90% (82/91)	36% (5/14)	94% (82/87)
24	114	.93	37.5	14% (16/114)	64% (7/11)	91% (93/102)	44% (7/16)	96% (93/97)
30	128	.96	47.5	13% (17/128)	83% (10/12)	94% (109/116)	59% (10/17)	98% (109/111)
36	298	.91	67.5	16% (48/298)	80% (24/30)	91% (244/268)	50% (24/48)	98% (244/250)
48	337	.94	47.5	17% (58/337)	85% (28/33)	90% (274/304)	48% (28/58)	98% (274/279)
60	417	.96	52.5	16% (68/417)	85% (35/41)	91% (343/376)	51% (35/68)	98% (343/349)

Sens, sensitivity; Spec, specificity; AUC, area under curve; positives, rate of positive identifications by ASQ:SE; PPV, positive predictive value; NPV, negative predictive value.

**Figure 1.** Receiver operating characteristic (ROC) curves for each Ages and Stages Questionnaire: Social-Emotional (ASQ:SE) form marked with selected cutoff value. Note. Green line = reference line, blue line = test score, star = selected cutoff value

Lunkenheimer, and Olson (2007) measured externalizing and internalizing problems with the CBCL and the C-TRF in an at-risk sample of children at age three and

5 years. They found that parent and teacher reports predicted 9% to 33% of the variance in the latent problem factor at early school age. Basten et al. (2016) report

from their general population study of 12-18-month-old children with elevated problem scores (based on the ASEBA manuals T-scores), were at increased risk of elevated problem scores at age three and six. On the other hand, the problem profiles were hard to predict, indicating a heterotypic stability. Future research would benefit to investigate the ASQ:SE' and the C-TRF' predictive validity from early childhood through school age.

Cutoff values

Sensitivity and specificity depends on how the cutoff value on the comparator is defined and the selection of cutoff values will always involve a tradeoff between sensitivity and specificity. The choice depends largely on the context in which the instrument is intended to be used. If the priority is to reduce the rate of false-negative cases, a cutoff with high specificity would be considered acceptable. If high sensitivity is required, a lower cutoff value should be chosen. In other words, lowering the cutoff value increases the sensitivity leading to a higher probability of correctly identifying children at risk, but at the cost of a higher rate of false positives (low-risk children testing positive for risk). The consequence of such an approach could be unnecessary referrals and follow-up evaluations, as well as stress and worry for falsely classified children and their parents. However, if screening is not undertaken, the chance of early detection and intervention may be lost for children actually at risk. Factors such as the expense of intervention or treatment, available resources, intrusiveness and possible stigmatizing effects must also be considered before screening (Sawyer et al., 2013). A positive test on a brief screening instrument used in child-care centers primarily to stimulate dialog between teachers and parents, should prompt further investigation (conferring with parents, conferring or collaborating with others, referral, further testing, etc.). Training users of the instrument how to score, interpret and make informed decisions, preferably in collaboration with other actors in the mental health field, could provide a good framework for screening.

It is important to note that a screening instrument such as the ASQ:SE is not sufficient to establish a diagnosis (APA Practice Central, 2014). The ASQ:SE was not developed for this goal. It can only provide a snapshot of children's social-emotional development at the time of screening. All those who perform screening for children have an ethical responsibility to ensure that the appropriate next steps are taken if a child's test results are positive (American Academy of Child and Adolescent Psychiatrists, 2001). In Norway, most preventive interventions are performed at the child-care centers by the staff, for example by supporting in everyday activities. Often the staff members receive supervision from other mental health or educational professionals. Around 4% of Norwegian preschoolers at age four use mental health services and 1/10 of children with a symptom load that qualifies for a psychiatric disorder have received help (Wichstrøm, Belsky, Jozefiak, Sourander, & Berg-Nielsen, 2014). Given the free and easily accessible health care in Norway, one might argue that these rates are low and that procedures that facilitate early identification of mental health problems are warranted. This study suggests that the ASQ:SE could serve as a good starting point for teachers in child-care centers who are uncertain whether a child is developing normally or

those who seek a brief screening instrument to test the overall social-emotional development among a group of children.

Strengths and limitations

Previous studies have solely used parent reports, mainly those of mothers, when investigating the psychometric properties of the ASQ:SE. This study adds to the knowledge of how this instrument performs using teachers as informants. Another strength of the study is the large sample size and the age span of the sample. However, employing a larger sample within the ASQ:SE forms for the youngest ages would have been beneficial in testing the accuracy of the ASQ:SE. A limitation of this study is the multiple responses from the same teacher for different children. Consequently, different biases could have been introduced through the procedure or context (mood, priming, etc.) and different confounders may be in play. For example the teacher-child relationship may influence the scoring of an instrument if the teacher rates children with whom they have a poor relationship worse regardless of the children's actual problem status. Another possible bias could be introduced through the lack of information about those who chose not to participate.

Future research should also investigate the psychometric properties of the ASQ:SE 12 form using an appropriate criterion measure. The PPV and NPV reported in this study cannot be generalized to other countries, as the prevalence of mental health problems may differ. The standardizations of norms and the development of cutoff values should be conducted with samples drawn from the same population to which they will be applied (Velikonja et al., 2016). It should also be mentioned that ASQ:SE2 was launched in 2015 with new forms for younger and older children. The developers have also added more items to the new/revised forms. ASQ:SE2 has not yet been translated to Norwegian and no distribution plan for the Norwegian market exist at present. However, future research should investigate the psychometrics qualities of ASQ:SE2 as well.

Clinical implications

This study suggests that the use of the ASQ:SE in child-care centers may be efficient in identifying concerns about children at risk. A reduction in the cutoff values in Norwegian child-care centers from the original cutoff values in the ASQ:SE manual (table A9, page 89, Squires et al., 2002) would be beneficial to increase the detection rate of children with social-emotional problems (true positives), with the exception of ASQ:SE 36 where the optimal cutoff value was found to be higher compared to the manual. Given the generally lower mean scores for social-emotional problems in this Norwegian sample, this approach seems reasonable. The low sensitivity observed for the 18- and 24- month ASQ:SE forms suggests that these forms should be avoided and may need to be accompanied by additional screening instruments or observational methods, if used, to increase their detection rate of children at risk. The 30- to 60- month forms exhibit promising psychometric properties and could be recommended as a first-step screening instrument in Norwegian child-care centers. These forms may prove helpful in early detection of children at risk and could facilitate early intervention.

Conclusion

The ASQ:SE could serve as a good starting point for screening for social-emotional problems at a child-care center community level, but it should be used in a reflective manner based on what teachers wish to accomplish. Similar to other screening instrument, the results of the ASQ:SE depend on the informants' knowledge of normal and abnormal development and their observational skills, as well as the instruments psychometric properties. The six ASQ:SE forms investigated in this study have exhibited promising sensitivity and specificity overall, however, the two youngest age forms should be avoided or complemented by other measures to increase the detection rate of children at risk.

Acknowledgements

No external funding was received as part of this work. The authors have declared that they have no competing or potential conflicts of interest.

References

- Achenbach, T.M., & Rescorla, L.A. (2000). *Manual for the ASEBA preschool forms & profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.
- American Academy of Child and Adolescent Psychiatrists (2001). Psychiatric care of children in the foster care system. Available from: <http://www.aacap.org/publications/policy/ps45.htm> [last accessed 15 June 2016].
- APA Practice Central (2014). Distinguishing Between Screening and Assessment for Mental and Behavioral Health Problems. Available from: <http://www.apapracticentral.org/reimbursement/billing/assessment-screening.aspx> [last accessed 15 June 2016].
- Basten, M., Tiemeer, H., Althoff, R.R., van de Schoot, R., Jad-doe, V.W.V., Hofman, A., ... & van der Ende, J. (2016). The stability of problem behavior across the preschool years: An empirical approach in the general population. *Journal of Abnormal Psychology, 44*, 393–404.
- Belfer, M.L. (2008). Child and adolescent mental disorders: The magnitude of the problem across the globe. *Journal of Child Psychology and Psychiatry, 49*, 226–236.
- Center on the Developing Child at Harvard University (2010). The Foundations of Lifelong Health Are Built in Early Childhood. Available from: <http://www.developingchild.harvard.edu> [last accessed 15 June 2016].
- Dougherty, L.R., Leppert, K.A., Merwin, S.M., Smith, V.C., Bufferd, S.J., & Kushner, M.R. (2015). Advances and directions in preschool mental health research. *Child Development Perspectives, 9*, 14–19.
- Dowdy, E., Chin, J.K., & Quirk, M.P. (2013). Preschool screening: An examination of the Behavioral and Emotional Screening System Preschool teacher form (BESS Preschool). *Journal of Psychoeducational Assessment, 31*, 578–684.
- Doyle, O., Harmon, C.P., Heckman, J.J., & Tremblay, R.E. (2009). Investing in early human development: Timing and economic efficiency. *Economics and Human Biology, 7*, 1–6.
- Egger, H.L., & Angold, A. (2006). Common emotional and behavioral disorders in preschool children: Presentation, nosology, and epidemiology. *Journal of Child Psychology and Psychiatry, 47*, 313–337.
- Essex, M.J., Kraemer, H.C., Slattery, M.J., Burk, L.R., Boyce, W.T., Woodward, H.R., & Kupfer, D.J. (2009). Screening for childhood mental health problems: Outcomes and early identification. *Journal of Child Psychology and Psychiatry, 50*, 562–570.
- Fayers, P.M., & Machin, D. (2007). *Quality of life* (2nd edn). Chichester: John Wiley & Sons, Ltd.
- Feeney-Kettler, K.A., Kratochwill, T.R., Kaiser, A.P., Hemmeter, M.L., & Kettler, R.J. (2010). Screening young children's risk for mental health problems: A review of four measures. *Assessment of Effective Interventions, 35*, 218–230.
- de Groot, A., Koot, H.M., & Verhulst, F.C. (1994). Cross-cultural generalizability of the child behavior checklist cross-informant syndromes. *Psychological Assessment, 6*, 225–230.
- Heckman, J.J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science, 312*, 1900–1902.
- Heo, K.H., & Squires, J. (2011). Cultural adaption of a parent completed social emotional screening instrument for young children: Ages and stages questionnaire-social emotional. *Early Human Development, 88*, 151–158.
- Hosmer, D.W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd edn, pp. 2000). New York: John Wiley & Sons Inc.
- Ivanova, M.Y., Achenbach, T.M., Rescorla, L.A., Dumenci, L., Almqvist, F., Bathiche, M., ... & Verhulst, F. (2007). Testing the teacher's report form syndromes in 20 societies. *School Psychology Review, 36*, 468–483.
- Ivanova, M.Y., Achenbach, T.M., Rescorla, L.A., Harder, V.S., Ang, R.P., Bilenberg, N., ... & Verhulst, F.C. (2010). Preschool psychopathology reported by parents in 23 societies: Testing the seven-syndrome model of the child behavior checklist for ages 1,5-5. *New Research, 49*, 1215–1224.
- Kauffman, J.M. (1999). How we prevent the prevention of emotional and behavioral disorders. *Exceptional Children, 65*, 448–468.
- Kerr, D.C.R., Lunkenheimer, E.S., & Olson, S.L. (2007). Assessment of child problem behaviors by multiple informants: A longitudinal study from preschool to school entry. *Journal of Child Psychology and Psychiatry, 48*, 967–975.
- Koot, H., van den Oord, J., Verhulst, F.C., & Boomsma, D. (1997). Behavioral and emotional problems in young preschoolers: Cross-cultural testing of the validity of the child behavior checklist/2-3. *Journal of Abnormal Child Psychology, 25*, 183–196.
- Lavigne, J.V., Meyers, K.M., & Feldman, M. (2016). Systematic review: Classification accuracy of behavioral screening measures for use in integrated primary care settings. *Journal of Pediatric Psychology, 5*, 1–19.
- Poulou, M.S. (2015). Emotional and behavioural difficulties in preschool. *Journal of Child and Family Studies, 24*, 225–236.
- Rescorla, L.A., Achenbach, T.M., Ivanova, M.Y., Bilenberg, N., Bjarnadottir, G., Denner, S., ... & Verhulst, F.C. (2012). Behavioral/emotional problems of preschoolers: Caregiver/teacher reports from 15 societies. *Journal of Emotional and Behavioral Disorders, 20*, 68–81.
- Rescorla, L.A., Bochicchio, L., Achenbach, T.M., Ivanova, M.Y., Almqvist, F., Begovac, I., ... & Verhulst, F.C. (2014). Parent-teacher agreement on children's problems in 21 societies. *Journal of Clinical Child and Adolescent Psychology, 43*, 627–642.
- Salomonsson, B., & Sled, M. (2010). The ages & stages questionnaire: Social-emotional: A validation study of a mother-report questionnaire on a clinical mother-infant sample. *Infant Mental Health Journal, 31*, 412–431.
- Sawyer, A.C.P., Chittleborough, C.R., Lynch, J.W., Baghurst, P., Mittinty, M.N., Kaim, A.L.E., & Sawyer, M.G. (2013). Can screening 4-5 year olds accurately identify children who will have teacher reported mental health problems when children are aged 6-7 years? *Australian and New Zealand Journal of Psychiatry, 48*, 1–10.
- Squires, J., Bricker, D., Heo, K., & Twombly, E. (2001). Identification of social-emotional problems in young children using a parent-complete screening measure. *Early Childhood Research Quarterly, 16*, 405–419.
- Squires, J., Bricker, D., & Twombly, E. (2002). *The ASQ:SE user's guide*. Baltimore, MD: Paul H. Brookes Publishing Co.
- SSB (2017). Kindergartens, 2016, final figures. Available from <https://www.ssb.no/en/utdanning/statistikker/barnehager/aar-enderlige/2017-03-21> [last accessed 1 September 2017].

- Sveen, T.H., Berg-Nielsen, T.S., Lydersen, S., & Wichstrøm, L. (2013). Detecting psychiatric disorders in preschoolers: Screening with the strengths and difficulties questionnaire. *Journal of American Academy of Child and Adolescent Psychiatry*, 52, 728–736.
- Theunissen, M.H.C., Vogels, A.G.C., de Wolff, M.S., Crone, M.R., & Reijneveld, S.A. (2015). Comparing three short questionnaires to detect psychosocial problems among 3 to 4-year olds. *BMC Pediatrics*, 15, 1–8.
- Velikonja, T., Edbrooke-Childs, J., Calderon, A., Slead, M., Brown, A., & Deighton, J. (2016). The psychometric properties of the ages & stages questionnaires for ages 2-2.5: A systematic review. *Child: Care, Health, and Development*, 43, 1–17.
- Verhulst, F.C., & Koot, H.M. (1992). *Child psychiatric epidemiology: Concepts, methods and findings*. Beverly Hills, CA: Sage Publications.
- Wichstrøm, L., Belsky, J., Jozefiak, T., Sourander, A., & Berg-Nielsen, T.S. (2014). Predicting service use for mental health problems among young children. *Pediatrics*, 133, 1054–1060.
- Wichstrøm, L., Berg-Nielsen, T.S., Angold, A., Egger, H.L., Solheim, E., & Sveen, T.H. (2012). Prevalence of psychiatric disorders in preschoolers. *Journal of Child Psychology and Psychiatry*, 53, 695–705.
- de Wolff, M.S., Theunissen, M.H., Vogels, A.G.C., & Reijneveld, S.A.R. (2013). Three questionnaires to detect psychosocial problems in toddlers: A comparison of the BITSEA, ASQ:SE, and KIPPP. *Academic Pediatrics*, 13, 587–592.

Accepted for publication: 13 December 2017

ISBN 978-82-326-5351-5 (printed ver.)
ISBN 978-82-326-5460-4 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)



NTNU

Norwegian University of
Science and Technology