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

K. Stensen*, F. Stenseng*, S. Lydersen*, M. S. de Wolff, J. Wallander***, and M. B. Drugli*, ****

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

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

*** University of California, Merced

**** Centre of the Study of Educational Practice, Hedmark University College

Correspondence regarding this article should be addressed to Kenneth Stensen, Regional Centre for Child and Youth Mental Health and Child Welfare, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. 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

Background: Early detection of mental health problems in childhood is important. However, studies on screening instruments for preschool children are rare. The aim of the present 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 childcare centers were recruited in Norway. Their teachers completed a survey including the ASQ:SE and the Caregiver-Teacher Report Form (C-TRF). The Spearman' 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' correlations 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.

Conclusion: The ASQ:SE could serve as a good starting point for screening for social-emotional problems among children in childcare 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%-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 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 (Feeney-Kettler et al., 2010; Dougherty et al., 2015). There is broad consensus that the early years are the optimal period for identifying children at risk for later serious mental health problems (Kauffman, 1999; Heckman, 2006; Doyle et al., 2009; Heo & Squires, 2011; de Wolff et al., 2013; Poulou, 2015; Dougherty et al., 2015). 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 (Feeney-Kettler et al., 2010; Dougherty et al., 2015). 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 (Sawyer et al., 2013; Dougherty et al., 2015).

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). Additionally, 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 (Verhulst & Koot, 1992; de Groot, Koot, & Verhulst, 1994; Koot et al., 1997; Ivanova et al., 2007a; Ivanova et al., 2010; Rescorla et al., 2012; Rescorla et al., 2014). The ASEBA instrument for children one and one-half to five 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 US children aged three to 66 months old (Squires et al., 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 non-risk 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 et al., 2015) reported weaker psychometric properties among a population of 1650 children aged three to four 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: (1) children age 12-36 months (SE .93, SP .78), (2) children age 14 months (SE .56, SP .91) and 24 months (SE .84, SP .91), (3) children age 24 months (SE .95, SP .90), and (4) 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 one 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 the present study is to validate the ASQ:SE teacher report with a sample of children aged 18 months to five years in Norwegian childcare centers. Our research questions are as follows: 1) How do the different ASQ:SE forms correspond with the C-TRF? 2) 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 childcare centers. The data used in the present study were collected from 2012 to 2014 before the intervention commenced.

Participants

In Norway, children typically begin at childcare centers when they are one to two years old. In 2016, 91% of the Norwegian children attended childcare 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 five 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 childcare centers from 18 months to five 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 childcare center. Parental consent also gave the teacher in the childcare 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 (six 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 (de Wolff et al., 2013; Theunissen et al., 2015).

Caregiver-Teacher Report Form (C-TRF).

The C-TRF (Achenbach & Rescorla, 2000) contains 100 items describing problem behavior for children from 18 months to five 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 (Verhulst & Koot, 1992; de Groot et al., 1994; Koot et al., 1997; Ivanova et al., 2007a; Ivanova et al., 2010; Rescorla et al., 2012; Rescorla et al., 2014)

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.

Insert Table 1& 2 here

Statistical analysis

C-TRF 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' 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.

Validity and screening accuracy

Table 3 presents the Spearman' 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%.

Insert Table 3 here

Insert Figure 1 here

Insert Table 4 here

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 et al., 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, 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 five 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 childcare centers primarily to stimulate dialogue 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 childcare centers by the staff, for example by supporting in everyday activities. Often the staff 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 et al., 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 childcare 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 by 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 childcare centers may be efficient in identifying concerns about children at risk. A reduction in the cutoff values in Norwegian childcare 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 childcare 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 childcare 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.

Acknowledgment

No external funding was received as part of this work. None of the authors have declared any 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. Retrieved from: http://www.aacap.org/publications/policy/ps45.htm.

APA Practice Central (2014). Distinguishing Between Screening and Assessment for Mental and Behavioral Health Problems. Retrieved from: http://www.apapracticecentral.org/reimbursement/billing/assessment-screening.aspx

Basten, M., Tiemeier, 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 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. Retrieved from *http://www.developingchild.harvard.edu*

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

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

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*. 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

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* (2. ed.). New York: John Wiley & Sons Inc. 2000

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 Pedriatic 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., & Sleed, 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 & New Zealand Journal of Psychiatry*. 00: 1-10

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

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

SSB (2017). Retrieved from https://www.ssb.no/en/utdanning/statistikker/barnehager/aar-endelige/2017-03-21

Sveen, T. H., Berg-Nielsen, T.S., Lydersen, S., & Wichstrøm, L. (2013). Detecting Psychiatric Disorders in Preschoolers: Screening With the Strenghts 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

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

Velikonja, T., Edbrooke-Childs, J., Calderon, A., Sleed, 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*. 1-17

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

Table 1. Descriptive information of the sample for ASQ:SE and C-TRF total problem score separated by age group (n ASQ:SE= 1395, n C-TRF= 1428)

	ASQ:SE		C-TRF	
	n	Mean (SD)	n	Mean (SD)
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)

Note. SD = standard deviation

Table 2. Percentiles of the ASQ:SE forms and the 90th percentile on the C-TRF

	ASQ:SE					C-TRF
Form	n items	25th %	50th %	75th %	90th %	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 ASQ:SE and C-TRF total problem scores

ASQ:SE form	n	The Spearman' 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

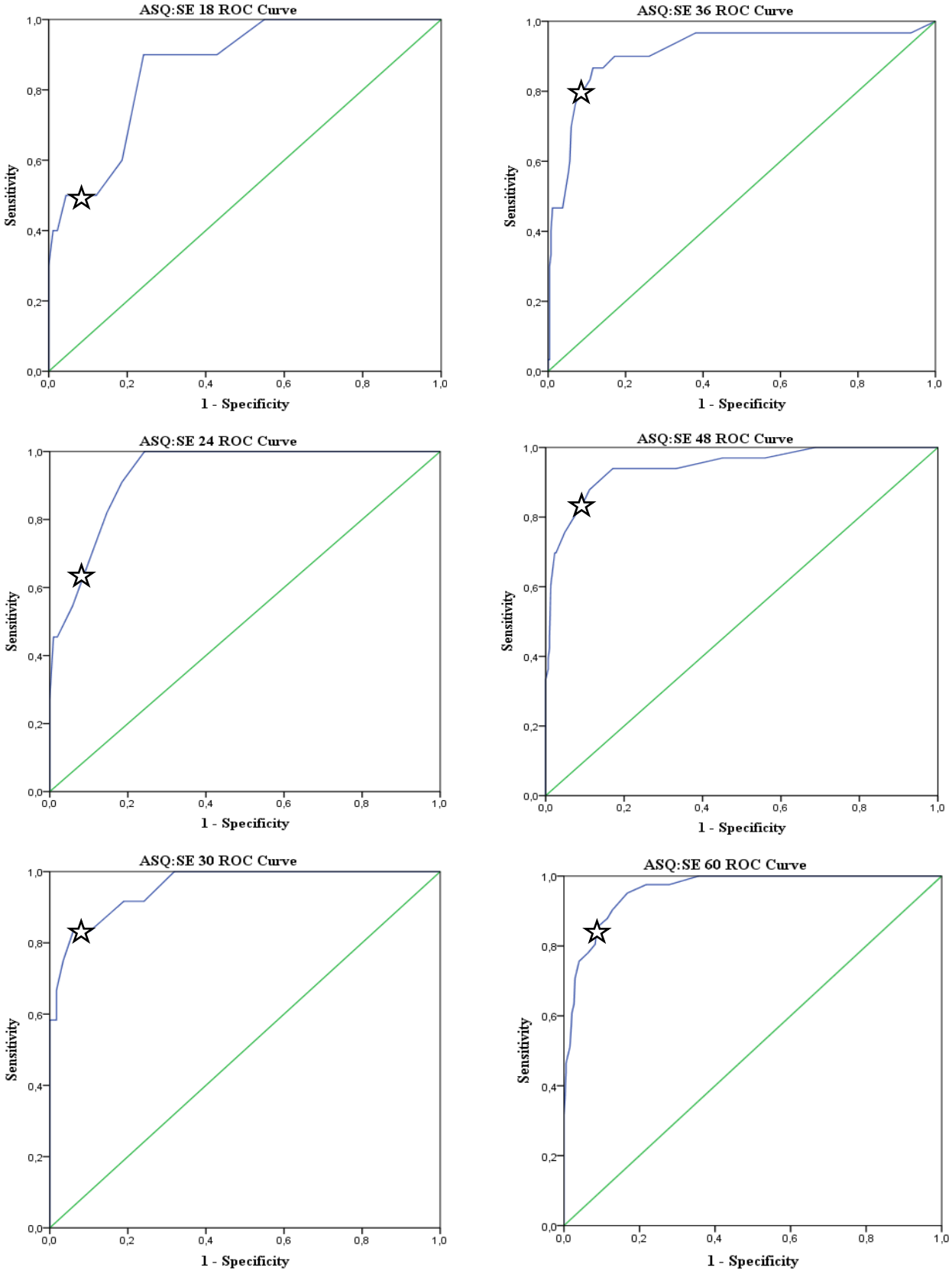
Note. Confidence intervals (CIs) are based on bootstrapping 10,000 samples (bias-corrected and accelerated)

Table 4. ASQ:SE screening performance against scores ≥ 90 percentile on the C-TRF

							Prevalence 10%	
ASQ:SE	n	AUC	Cutoff	Positives	Sens	Spec	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)

Note. 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. ROC curves for each ASQ:SE form marked with selected cutoff value



Note. Green line= reference line, blue line= test score, star= selected cutoff value