



# Investigating the Measurement Invariance of the Caregiver-Teacher Report Form (C-TRF) Factors in a Norwegian Early Childhood Education and Care Context

Kenneth Stensen<sup>1</sup> · Stian Lydersen<sup>1</sup> · Ingunn Ranøyen<sup>1</sup> · Ratib Lekhal<sup>2,4</sup> · May Britt Drugli<sup>1,5</sup> · Elisabet Solheim Buøen<sup>3</sup>

Accepted: 18 July 2023  
© The Author(s) 2023

## Abstract

The Caregiver-Teacher Report Form (C-TRF) is an instrument frequently used by researchers and clinicians to assess psychopathology in preschool children based on information from early childhood education and care professionals. However, the measurement invariance of this instrument has not been investigated. Thus, the aim of the current study was to investigate the measurement invariance of the C-TRF for (1) girls vs. boys, and (2) for children below the age of three years old vs. three years or older, as measurement invariance is a prerequisite for a meaningful comparison of latent means between groups. Based on reports of 2904 children from 353 early childhood education and care professionals, our results showed full scalar invariance for an adapted C-TRF model, which supports the applicability of the instrument. However, awareness of the factor *Somatic complaints* and items with out-of-range parameters is warranted.

**Keywords** Childhood development · Mental health · Preschool · Psychiatric disorder · Psychometric properties · Scalar invariance

## Introduction

The preschool period is characterized by children's rapid development, making it an important window for shaping healthy social, emotional, and cognitive functioning throughout their lifespans, as well as identifying and preventing

mental health problems and intervening before problems evolve into disorders (Dougherty et al., 2015; Pool & Hourcade, 2011; Poulou, 2015; Wakschlag et al., 2019; Zaim and Harrison, 2020). Worldwide prevalence estimates indicate that every fifth child experiences mental health problems, and 13% of children meet the criteria for a psychiatric disorder (Belfer, 2008; Polanczyk et al., 2015). Additionally, comorbidity is common, as one third of children aged 1 to 7 years old who meet the criteria for a psychiatric disorder also meet the criteria for at least one additional psychiatric disorder (Vasileva et al., 2021). For some children the problems are transient, while for others the problems are more enduring, where children may struggle to return to a normal developmental trajectory (Blok et al., 2022; Powell et al., 2006). Approximately half the preschoolers with a psychiatric diagnosis still meet the criteria for a psychiatric disorder in middle childhood or early adolescence (Finsaas et al., 2018a). Moreover, psychopathology in early childhood predicts long-lasting deleterious effects on children's functioning, even in children with subthreshold psychiatric problems (Finsaas et al., 2018b).

---

✉ Kenneth Stensen  
kenneth.stensen@ntnu.no

- <sup>1</sup> 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), P.O. Box 8905, Trondheim N- 7491, Norway
- <sup>2</sup> Department of Communication and Culture, BI Norwegian Business School, Oslo, Norway
- <sup>3</sup> Regional Centre for Child and Adolescent Mental Health Eastern and Southern Norway, Oslo, Norway
- <sup>4</sup> Department of Education, University of Oslo, Oslo, Norway
- <sup>5</sup> Centre of the Study of Educational Practice, Inland Norway University of Applied Sciences, Hamar, Norway

Mental health problems in preschool children remain under-identified, under-referred, and under-treated, even though they show the same prevalence estimates as older children (Bufferd et al., 2012; Egger & Angold, 2006; Horwitz et al., 2003, 2007; Wichstrøm et al., 2012). Identifying preschool children with mental health problems can be challenging due to the rapid development that occurs in this period (Keenan et al., 1998), where some behaviors may be considered appropriate at one age or in one context and inappropriate at another age or in another context. Many children attend early childhood education and care (ECEC) centers, where they spend prolonged periods with adults outside the family, therefore, the ECEC context represents a promising arena for early identification of and intervention for mental health problems. Thus, attention should be directed towards ECEC professionals' perceptions of problem behaviors among children (Poulou, 2015). However, attempts by ECEC professionals to identify problem behaviors without the use of standardized instruments can lead to misclassification. Such misclassification may hinder children who need support from getting help or unnecessarily encumber those who do not need help (Stensen et al., 2022a). Developing psychometrically valid identification procedures and instruments to ensure that children in need of help receive that help is essential for their healthy development (Feeney-Kettler et al., 2010; Ivanova et al., 2011).

The Caregiver-Teacher Report Form (C-TRF) (Achenbach & Rescorla, 2000) is a frequently used instrument to gather information on psychopathology in preschool children from other caregivers than parents, such as ECEC professionals, and is applied by both researchers and clinicians. The C-TRF and its parent-reported equal, the Child Behavior Checklist, have shown promising psychometric properties and are often used as "gold standards" for testing other assessment or screening instruments against (Lavigne et al., 2016). The six factors constituting the C-TRF, measuring *Emotionally reactive*, *Anxious/depressed*, *Withdrawn*, *Somatic complaints*, *Attention problems*, and *Aggressive behavior*, have found psychometric support regarding factorial validity and internal consistency in several countries outside the United States where the instrument was developed (Ivanova et al., 2011). However, factorial issues with the *Emotionally reactive* and *Anxious/depressed* factors have also been reported due to the high correlation between these factors, and the factor *Somatic complaints* has shown sub-optimal psychometric properties (Ivanova et al., 2011; Stensen et al., 2022b).

Measurement invariance (MI) is an important psychometric property of any instrument and a requirement for meaningful comparisons of latent means between groups (such as children's sex and age) (Putnick & Bornstein, 2016; van de Schoot et al., 2012). In other words, MI relates to

how raters interpret items. If raters attribute different meanings to the same items across groups, it may jeopardize the validity of the instrument when comparing the latent means because the factors hold different meanings. For instance, it has been reported from studies using the C-TRF that boys are scored significantly higher than girls on externalizing problems (*Attention problems* and *Aggressive behavior*), while younger preschool children are scored significantly higher on both internalizing (*Emotionally reactive*, *Anxious/depressed*, *Withdrawn*, *Somatic complaints*) and externalizing problems than older preschool children (Marković et al., 2016; Rescorla et al., 2012). However, without information of the instrument's MI it is uncertain if the C-TRF measures the same latent factors based on children's sex and age or if it measures more or less similar factors due to informant's interpretations of the items, making comparison of latent means more or less appropriate. For example, if higher aggressive behavior scores for younger preschool children compared to older preschool children are caused by the ECEC professionals attributing other behavioral expressions, such as frustration and anger, as aggressive behavior to a larger degree in younger children compared to older children, the latent means may not be directly comparable as aggressive behavior holds different meanings based on children's age. Consequently, an awareness of invariant and non-invariant items is important for an instrument's interpretability and validity.

The MI of the C-TRF has not yet been investigated. Thus, the aim of the current study was to examine the MI of the C-TRF in the context of Norwegian early childhood education and care, which may inform how ECEC professionals interpret the items included in the instrument. More specifically, based on previous studies (Marković et al., 2016; Rescorla et al., 2012) reporting significant sex and age differences in C-TRF scores, the current study examined the MI of the C-TRF for (1) girls vs. boys, and (2) for children below the age of three years old vs. three years or older as it is important for the interpretability of the instrument's reported mean scores. As knowledge about the psychometric properties of assessment instruments applied in Nordic countries is scarce or lacking, especially for very young children (Peltonen et al., 2022), efforts to inform researchers and clinicians about the applicability and accuracy of instruments are important.

## Methods

This study is based on combined baseline data from two ECEC projects conducted in central and south eastern Norway. Data from the project *Children in Central Norway*

were collected from 2012 to 2014 and data from *Thrive by Three* were collected during 2018.

## Procedure and Participants

### Children in Central Norway

The parents of children enrolled in ECEC centers in three municipalities in central Norway received information regarding the project via letters and parent meetings before the project commenced. The letters provided parents with the option to enroll their child in the project either by returning a signed consent form to their ECEC center or by consenting digitally with their unique invitation code to the project's online survey. Parental consent was obtained to allow the ECEC professionals who knew the child best to complete the online survey regarding the child. ECEC professional provided written consent using their own unique invitation code for the online survey. A total of 1,631 (77%) of the invited parents consented to enroll their child in the project, and 169 ECEC professionals reported on 1,430 children between 1 and 6 years of age (mean age = 44 months; 51% boys).

### Thrive by Three

Data was obtained from *Thrive by Three* for 1,474 children (mean age = 21 months; 51% boys) and 184 units/groups from 78 ECEC centers from seven municipalities or city parts from central and south eastern Norway. An ECEC professional within the unit/group who knew the child best answered the C-TRF. When the sample from *Thrive by Three* was combined with the sample from *Children in Central Norway*, the total sample for this study included 2,904 children (mean age = 33 months; 51% boys) and 353 ECEC professionals. On average, ECEC professionals reported on 8.2 children each. For both projects, participation was voluntary and consent could be withdrawn without reprisal until the participation registry was anonymized.

## Measurements

### The C-TRF

ECEC professionals completed the C-TRF (Achenbach & Rescorla, 2000), which contains 99 items and one open-ended item describing symptoms of mental health problems in children aged from 1.5 to 5 years old. Each item has three response options: 0 = "not true (as far as you know)", 1 = "somewhat or sometimes true" and 2 = "very often or often true". The scores can be added up to obtain a *Total problem* score, ranging from 0 to 200, where a higher score indicates

a higher symptom load. The C-TRF contains the following six latent factors: *Emotionally reactive* (seven items), *Anxious/depressed* (eight items), *Withdrawn* (10 items), *Somatic complaints* (seven items), *Attention problems* (nine items), and *Aggressive behavior* (25 items). The 66 items constituting these six factors were the focus of the current study, whereas the remaining 34 items constituting *Other problems* were excluded as they do not form a latent factor in the C-TRF.

## Statistical Analyses

ECEC children were dichotomized by age into younger ECEC children (younger than 36 months) and older (36 months and older) children to reflect the organizational structure of ECEC centers in Norway. First, as children were nested within ECEC professionals, the internal consistency of the six C-TRF factors for the full sample was investigated to obtain the multilevel omega ( $\omega$ ) coefficient. Unless reliability is identical at each level of analysis, reporting only single-level estimates does not reflect the scales actual internal consistency (Geldhof et al., 2014). The  $\omega$  coefficient was preferred over Cronbach's alpha because the latter depends on rather strict assumptions, such as tau-equivalence and normally distributed scores, which can lead to biased estimates if violated (Dunn et al., 2014; McNeish, 2018; Peters, 2014; Sijtsma, 2009; Yang & Green, 2011). The multilevel  $\omega$  with a 95% confidence interval was computed with the package "multilevelTools" (Wiley, 2022) in RStudio. The multilevel  $\omega$  estimates are interpreted in the same way as the alpha, where estimates  $\geq 0.70$  are considered to indicate satisfactory internal consistency (Taber, 2018).

We then investigated the MI of the C-TRF using a series of multilevel multigroup confirmatory factor analyses (MGCFAs) grouped by (1) girls vs. boys, and (2) for children below the age of three years old vs. three years or older. The MGCFAs are carried out to determine whether the respondents attribute the same meaning to the latent factors and whether the means and scores can be interpreted similarly across groups (van de Schoot et al., 2012). This is accomplished by investigating the model fit indices while adding additional constraints to the models that follow a hierarchical structure ranging from configural (weak invariance) to scalar (strong invariance).

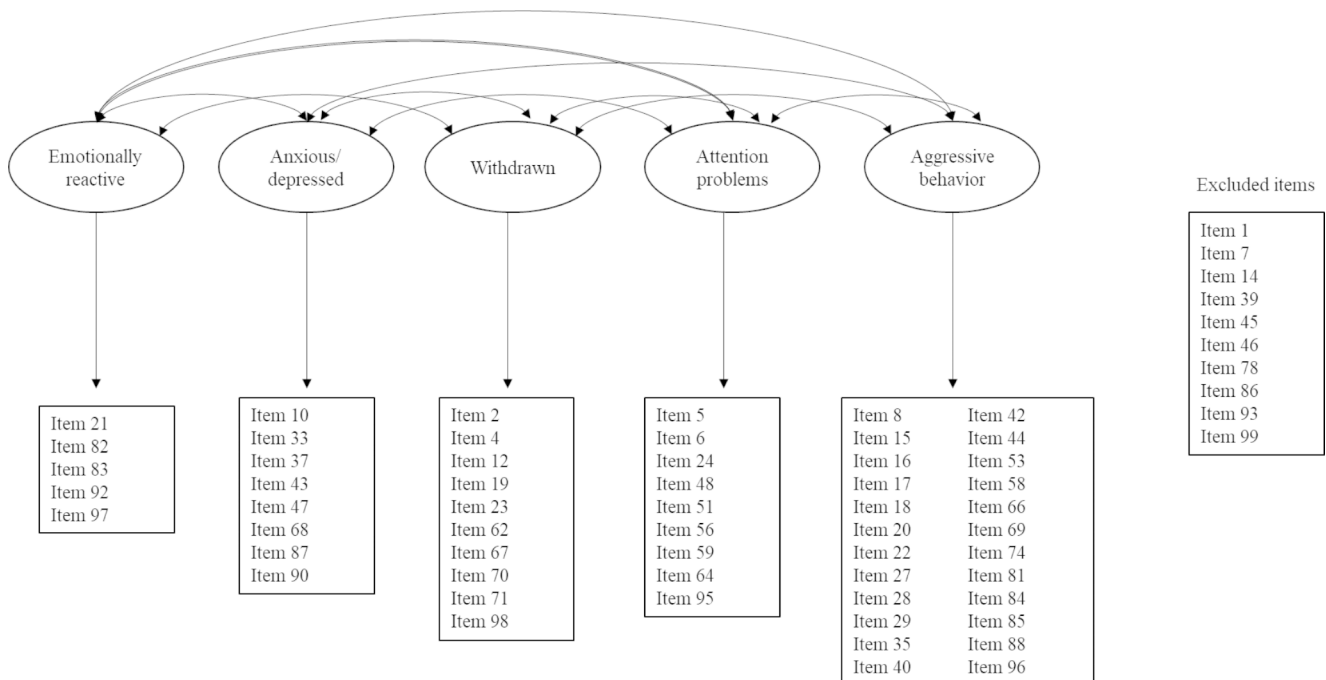
1. Configural invariance (equal factor structure across groups).
2. Metric invariance (equal factor loadings across groups).
3. Scalar invariance (equal thresholds across groups as the variables are ordered categorical).

Before examining the MI of the C-TRF, 10 items were removed from the analyses as Mplus warned that the variable variance/covariance matrix was non-positive definite. The warning possibly arose due to out-of-range parameters (i.e., negative variance or residual variance for a latent variable, a correlation  $\geq 1.0$  between two latent variables), and a linear dependency between more than two latent variables (Muthén & Muthén, 2017) (Fig. 1). Nine items were excluded due to empty cells in the bivariate frequency tables when these items were combined with certain other variables, leading to a correlation of 1. This problem concerned the factor *Somatic complaints* and the items “cruel to animals” (*Aggressive behavior*), and “nervous movements or twitching” (*Emotionally reactive*). In addition, the item “worries” (*Emotionally reactive*) loaded on several factors and was excluded from further analyses. When rerunning the analyses without these 10 items, the warning regarding non-positive definiteness was resolved.

Step 1 in testing the measurement invariance involved investigating the C-TRF-adapted five-factor baseline model across children’s sex and then across age (configural variance), where all parameters could vary freely. Step 2 tested a model in which the factor loadings were constrained between groups, while the thresholds could vary freely (metric variance). In Step 3, a model was tested in which both loadings and thresholds were constrained to be equal between groups (scalar invariance) was tested. Configural invariance was determined if the five-factor model showed a good fit across the groups tested. Metric invariance is

established if the more constrained model still shows a good fit compared to the configural baseline model, whereas scalar invariance is established if the even more constrained model still shows a good fit compared to the metric invariance model (Hirschfeld & von Brachel, 2014).

The model fit was evaluated by inspecting the root mean square error of approximation (RMSEA), comparative fit index (CFI), and Tucker-Lewis index (TLI). RMSEA values of  $\leq 0.05$  indicate a good fit and values between 0.05 and 0.10 indicate an acceptable fit (MacCallum et al., 1996). For the CFI and TLI, values of  $\geq 0.95$  are commonly used to indicate a good model fit (Hu & Bentler, 1999), however, Browne and Cudeck (1993) argue that these thresholds are too strict and recommend a threshold of  $> 0.90$  to determine a good model fit and values of 0.80 to 0.90 determine an acceptable model fit. To evaluate invariance, Cheung and Rensvold (2002) recommended that a CFI reduction of  $\leq 0.01$  when adding additional constraints to the model indicate that the null hypothesis of invariance should not be rejected. The CFI difference between models was preferred as an indicator of invariance, as it is less sensitive to sample size and more sensitive to lack of variance than chi-square ( $\chi^2$ ) statistics (Meade et al., 2008). Multilevel MGCEFA analyses were performed with Mplus v.8.4 (Muthén & Muthén, 2017) using the weighted least square mean variance (WLSMV) estimator to account for the non-normal distribution of data. The WLSMV estimator is appropriate for ordered categorical data and produces accurate parameter estimates (DiStefano & Morgan, 2014).



**Fig. 1** Test model for multigroup confirmatory factor analyses  
Note: An item overview can be found in Achenbach and Rescorla (2000)

**Table 1** The multilevel omega coefficient with 95% confidence interval for the factors in the Caregiver-Teacher Report Form with all the 66 items included

Multilevel omega (n items)	Within	Between
Emotionally reactive (7)	0.65 [0.63, 0.67]	0.89 [0.83, 0.94]
Anxious/depressed (8)	0.69 [0.67, 0.71]	0.90 [0.86, 0.94]
Somatic complaints (7)	0.15 [0.12, 0.18]	0.96 [0.94, 0.99]
Withdrawn (10)	0.77 [0.76, 0.79]	0.86 [0.83, 0.94]
Attention problems (9)	0.87 [0.86, 0.88]	0.93 [0.90, 0.95]
Aggressive behavior (25)	0.92 [0.91, 0.92]	0.96 [0.94, 0.97]

**Table 2** Baseline (configural) model fit indices grouped by the children's sex and age

	Chi-Square	Degrees of freedom	CFI	TLI	RMSEA
Children's sex	8851.355	2948	0.902	0.897	0.037
Children's age	7775.083	2948	0.909	0.905	0.033

CFI, comparative fit index; TLI, Tucker-Lewis index; RMSEA, root mean square error of approximation

Lastly, if scalar invariance was not found, an inspection of the modification indices ( $\chi^2$ ) to locate non-invariant items was carried out. We then relaxed the constraints for the non-invariant items one by one, starting with the item with the greatest expected parameter change (EPC), to determine if this improved the model fit. If the less constrained scalar invariant model showed a CFI estimate within the threshold of  $\leq 0.01$ , compared to the metric model, partial scalar invariance was observed. No data were missing.

## Results

The internal consistency of the C-TRF factors was highly consistent among ECEC professionals, while more variation was found at the within level (Table 1). *Emotionally reactive* almost reached the threshold of  $\geq 0.70$  for satisfactory internal consistency, while the upper confidence interval of the *Anxious/depressed* factor surpassed this threshold. *Withdrawn*, *Attention problems*, and *Aggressive behavior* were above the threshold of  $\geq 0.70$ , with the highest estimate found for *Aggressive behavior* (0.92 at the within level). The lowest internal consistency was found at the within level for *Somatic complaints* (0.15).

Results from the multilevel MGCFAs using the adapted five-factor structure indicated good model fit when the model fit indices were inspected (Table 2). The TLI estimate for the children's sex group (0.897) was an exception as it fell within the acceptable model fit range, slightly below the threshold of  $> 0.90$  for a good fit. The investigation of the measurement invariance of the C-TRF exhibited full scalar invariance for both sex and age (Table 3), as adding additional constraints to the models did not result in a

**Table 3** Measurement invariance indices of the Caregiver-Teacher Report Form using the comparative fit index estimates grouped by children's sex and age

	Configural	Metric	Scalar
Children's sex	0.902	0.904	0.908
Children's age	0.909	0.911	0.913

significant drop ( $\leq 0.01$ ) in the CFI, thus, indicating that the latent means could be meaningfully compared. Because full scalar invariance was found, further investigation to locate non-invariant items was unnecessary.

## Discussion

The current study aimed to examine the MI of the C-TRF across children's sex and ages, which is an important psychometric property for meaningful comparison of latent means across groups. Previous studies using this instrument have reported significant age and sex differences (Marković et al., 2016; Rescorla et al., 2012); therefore, the findings from the current study can inform whether the C-TRF mean scores are based on actual differences or on different perceptions of the same items across groups. The results of this study indicate that the factor *Somatic complaints* is problematic both in terms of internal consistency and out-of-range parameters due to many zero cells when it is combined with other items. Consequently, this factor, along with three other items, were excluded. The adapted five-factor model showed full scalar invariance across children's sex and also across age. Thus, ECEC professionals attributed the same meaning to these latent factors independent of children's sex and age and the latent means were comparable across these groups.

The internal consistency of the six latent factors in the current sample showed estimates similar to the alpha and omega coefficients reported previously (Rescorla et al., 2012; Stensen et al., 2022b). That said, the lower mean age of children in the current study may have impacted the internal consistency for *Somatic complaints* compared to Rescorla et al. (2012); Stensen et al. (2022b). The estimates from Table 1 indicate that this factor is consistently inconsistent between ECEC professionals (omega coefficient of 0.15 within and 0.96 between), suggesting that the applicability of this factor is highly questionable and removing it from the instrument or aggregated scores used for research and clinical purposes should be considered. Given the limited capacity that younger children have to communicate, ECEC professionals may have problems distinguishing somatic problems due to psychogenic causes from actual illnesses. For instance, an ECEC professional may attribute the inability of a child with stomachache to sit still to attention rather than to somatic problems. In

addition to the *Somatic complaints* factor, three items were problematic. As with *Somatic complaints*, ECEC professionals may have difficulty capturing for ECEC very young children's concerns because of the limited communicative capacity of these children. The item "worries" loaded on more than one latent factor and a previous study (Rescorla et al., 2012) has shown that the factors *Emotionally reactive* and *Anxious/depressed* are highly correlated. Consequently, ECEC professionals may misattribute this type of behavior. "Cruelty to animals" may be explained by low prevalence and lack of opportunity for ECEC professionals' to observe such behavior, as animals (e.g., pets) are rarely found in the ECEC setting. Thus, it may be advisable to ask parents for information regarding this type of behavior as it may be more prevalent in the home environment. The last item, "nervous movements or twitching" may also be challenging for ECEC professionals to capture, as the motor skills of children develop rapidly in the preschool period. ECEC professionals may consider this type of behavior to be normal to motor problems rather than internalizing problems.

The adapted five-structure model showed full scalar invariance across both children's sex and age. This indicates that the factors of the C-TRF capture the same meaning regardless of the child is 1.5 or 5 years old, even though some behaviors are considered more normative than others at a given age. ECEC professionals may operate with different thresholds for what they perceive as normative behavior regarding children's age and sex (Stensen et al., 2021). Findings from the current study suggest that sex and age differences reported using the C-TRF (Marković et al., 2016; Rescorla et al., 2012) may reflect actual sex and age differences as perceived by ECEC professionals, as the latent means carried the same meaning across the groups tested. This supports the applicability of the C-TRF for assessing psychopathology in preschool children, as both reliability and validity are important for researchers and clinicians to interpret information accurately. However, the factor *Somatic complaints* and the three items with out-of-range parameters warrant awareness and should be excluded or used cautiously, because they may jeopardize the integrity of the instrument. If these 10 items are chosen to be retained for clinical or scientific purposes, they should be considered included in the category "*Other problems*" due to their factorial issues. This way the excluded items may still be a part of the C-TRF's total problem score.

## Strengths and Limitations

The current study included two large community samples that covered the full age range of children in ECEC centers. In particular, the inclusion of very young children was a major strength, as psychometric information on assessment

instruments involving this cohort are scarce or lacking. This study adds to the knowledge of the applicability of the C-TRF in the Norwegian ECEC context, and the findings can be used for cross-cultural comparison. One possible limitation was the reliance on one ECEC professional per child cluster. Even though the ECEC professional who knew a child best was informed to complete the C-TRF, we cannot know if this actually occurred. In general, children interact with several adults during the day in an ECEC center. Consequently, an adult's perception of a child's behaviors within the same cluster may vary. ECEC professionals who perceive their relationship with a child as conflictual are reported to rate the child as having more problem behaviors than they actually have (Berg-Nielsen et al., 2012). However, an inter-rater approach was not possible in the current study.

## Conclusion

This study adds to the knowledge of the MI of the C-TRF across children's sex and age, which has not been examined previously. The results support the applicability of the C-TRF to gather information of psychopathology in preschool children using ECEC professionals as informants. However, researchers and clinicians should be aware of the factor *Somatic complaints* and the three items with out-of-range parameters, as they may jeopardize the integrity of the instrument if included.

**Funding** No external funding was received as part of this work. Open access funding provided by NTNU Norwegian University of Science and Technology (incl St. Olavs Hospital - Trondheim University Hospital)

## Declarations

**Ethics approval and consent to participate** The study was approved by the Regional Committee for Medical and Health Research Ethics (REK) (reference number: 2011/2252; 2012/1030; 2013/1363; 2017/430).

**Informed Consent** Informed consent was obtained from all parents and ECEC professionals of included children.

**Conflict of Interest** None of the authors have declared any competing or potential conflicts of interest.

**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.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing,

adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## 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.
- 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.
- Blok, E., de Mol, C. L., van der Ende, J., Hillegers, M. H. J., Althoff, R. R., & White, T. (2022). Stability and Change of psychopathology symptoms throughout childhood and adolescence. *Child Psychiatry & Human Development*, *53*, 1330–1339.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.
- Bufferd, 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.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for Testing Measurement Invariance. *A Multidisciplinary Journal*, *9*, 233–255.
- DiStefano, C., & Morgan, G. B. (2014). A comparison of diagonal weighted least squares robust estimation techniques for Ordinal Data. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*, 425–438.
- 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.
- Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*, 399–412.
- 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.
- 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.
- 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>.
- Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a Multilevel Confirmatory factor analysis Framework. *Psychological Methods*, *19*, 72–91.
- Hirschfeld, G., & von Brachel, R. (2014). Improving multiple-group confirmatory factor analysis in R – a tutorial in measurement invariance with continuous and ordinal indicators. *Practical Assessment Research and Evaluation*, *19*, 1–12.
- 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.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.
- Ivanova, M. Y., Achenbach, T. M., Rescorla, L. A., Bilenberg, N., Bjarndottir, 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.
- 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.
- Wakschlag, L. S., Roberts, M. Y., Flynn, R. M., & Smith, J. D. Krogh-Jespersen, S., Kaat, A. J., Gray, L., Walkup, J., Marino, B. S., Norton, E. S., & Davis, M. M. (2019). Future directions for early Childhood Prevention of Mental Disorders: A Road Map to Mental Health, earlier. *Journal of Clinical Child & Adolescent Psychology*, *48*, 539–554.
- 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.
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*, 130–149.
- Marković, J., Rescorla, L., Okanović, P., Maraš, J. S., Bukorov, K. G., & Sekulić, S. (2016). Assessment of preschool psychopathology in Serbia. *Research in Developmental Disabilities*, *49*–50, 216–225.
- McNeish, D. (2018). Thanks Coefficient Alpha. We'll take it from Here. *Psychological Methods*, *23*, 412–433.
- Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*, *93*, 568–592.
- Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus User's Guide*. Eighth Edition. Los Angeles, CA: Muthén & Muthén.
- Peltonen, K., Kurki, M., Reedtz, C., Kaiser, S., Rasmussen, L. P., Merikukka, M., Rye, M., Laajasalo, T., Kyrrestad, H., Karjalainen, P., Pettersen, S. D., Eng, H., Breivik, K., & Martinussen, M. (2022). Psychological tests for expectant parents and young children in the nordic countries: A review of the evidence. *European Journal of Developmental Psychology*. <https://doi.org/10.1080/17405629.2022.2067141>.
- Peters, G. J. (2014). The alpha and the omega of scale reliability and validity: Why and how to abandon Cronbach's alpha and the route towards more comprehensive assessment of scale quality. *The European Health Psychologist*, *16*, 56–69.
- 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*. <https://doi.org/10.1111/jcpp.12381>.
- Pool, J. L., & Hourcade, J. J. (2011). Developmental screening: A review of contemporary practice. *Education and Training in Autism and Developmental Disabilities*, *46*, 267–275.
- Poulou, M. S. (2015). Emotional and behavioural difficulties in Preschool. *Journal of Child and Family Studies*, *24*, 225–236.
- Powell, D., Dunlap, G., & Fox, L. (2006). Prevention and intervention for the Challenging Behaviors of Toddlers and Preschoolers. *Infants & Young Children*, *19*, 25–35.
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90.
- 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.
- Sijtsma, K. (2009). On the use, misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*, 107–120.
- Stensen, K., Lydersen, S., Stenseng, F., Wallander, J. L., & Drugli, M. B. (2021). Teacher nominations of Preschool Children at Risk for Mental Health problems: How false is a false positive nomination and what make Teachers Concerned? *Journal of Psychopathology and Behavioral Assessment*, *43*, 646–656.
- Stensen, K., Lydersen, S., Stenseng, F., Wallander, J. L., Tveit, H., & Drugli, M. B. (2022a). 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*, *66*, 458–472.
- Stensen, K., Jozefiak, T., & Lydersen, S. (2022b). Psychometric properties of the caregiver-teacher report form in a sample of norwegian preschool children. *Nordic Journal of Psychiatry*, *76*, 616–622.
- Taber, K. S. (2018). The Use of Cronbach's alpha when developing and reporting Research Instruments in Science Education. *Research in Science Education*, *48*, 1273–1296.
- van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*, 486–492.
- Vasileva, M., Graf, R. K., Reinelt, T., Petermann, T., & Petermann, F. (2021). 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*, *62*, 372–381.
- 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.
- Wiley, J. F. (2022). Multilevel and Mixed Effects Model Diagnostics and Effect Sizes. Retrieved from: <https://cran.rproject.org/web/packages/multilevelTools/multilevelTools.pdf>.
- Yang, Y., & Green, S. B. (2011). Coefficient alpha: A reliability coefficient for the 21st Century? *Journal of Psychoeducational Assessment*, *29*, 377–392.
- Zaim, N., & Harrison, J. (2020). Pre-school mental health disorders: A review. *International Review of Psychiatry*, *32*, 189–201.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.