Convergent and discriminant validity of psychiatric symptoms reported in The Norwegian Mother and Child Cohort Study at age 3 years with independent clinical assessment in the Longitudinal ADHD Cohort Study

Guido Biele[1], Pål Zeiner[2] and Heidi Aase[1]

1) Division of Mental Health, Norwegian Institute of Public Health, Oslo, Norway
2) Oslo University Hospital, Child and Adolescent Mental Health Research Unit, Oslo, Norway
Correspondence: guido.biele@fhi.no

ABSTRACT

Epidemiological studies often use parent questionnaires to assess children's development and mental health. To date, few studies have investigated the validity of parent questionnaires with standardized clinical assessments as criterion. The current study examines discriminant and convergent validity of parent questionnaires for symptoms of Attention Deficit Hyperactivity Disorder (ADHD), Oppositional Defiance Disorder (ODD), and Conduct Disorder (CD) as well as symptoms of Separation Anxiety employed in the Norwegian Mother and Child Cohort Study by using structured clinical interviews performed 5 months later in the Longitudinal ADHD Cohort Study as a criterion. The comparison of confirmatory factor analysis models and examination of factor correlations indicate convergent and discriminant validity of MoBa parent questionnaires for preschool children, especially for the assessment of ADHD and ODD/CD. Future research should attempt to further improve parent questionnaires, examine their validity in representative samples, and explicitly test their utility for screening.

BACKGROUND

It is well documented that significant psychiatric symptoms and disorders may develop in early preschool age (1). Attention-Deficit/Hyperactivity Disorder (ADHD) is the most frequently diagnosed disorder in child psychiatric outpatient units (> 20%), and the large majority are diagnosed after school entry. Yet, while many of these children have significant symptoms with impairment in preschool years, very few are referred to specialist mental health services for help (2). As early psychosocial treatment of children at risk for developing ADHD may reduce negative consequences over time, early identification is warranted (3). However, the borders between age-appropriate and inappropriate behaviors and reactions in toddlers and young preschoolers are often fuzzy as normative behavior varies greatly in these age groups, making the identification of maladaptive behaviors difficult. It is an important clinical challenge to identify the children that will develop chronic behavioral problems while avoiding pathologizing normally developing active children (4). To meet this challenge, reliable and valid diagnostic tools that make this distinction need to be developed.

The Norwegian Mother and Child Cohort Study (MoBa) at the Norwegian Institute of Public Health (5) provided a unique opportunity to study early indicators and developmental trajectories of ADHD and co-occurring disorders. By identifying children with high scores on ADHD-related symptoms at a very young age and before symptoms develop to the full blown

clinical case we can follow a large group of children in MoBa over years in order to identify early manifestations of ADHD and factors that are associated with continuity of symptoms over time and co-occurring disorders. This was the aim of the MoBa substudy called The Norwegian Longitudinal ADHD Cohort Study (the ADHD Study). In the ADHD Study, children at risk to develop ADHD were identified based on the parent questionnaire in MoBa at 3 years and invited to a standardized clinical examination covering relevant psychiatric symptoms as well as cognitive and psychosocial functioning.

The MoBa Study is an exceptional data source for the analysis of a vast amount of health related questions, including psychiatric symptoms and disorders in children. An important assumption is that parent questionnaires can be used to reliably measure children's development and mental health (1). However, whereas reliability and construct validity – including convergent and discriminant validity – have been tested for a number of parent questionnaires (6-10), their convergent validity with an independent clinical assessment has rarely been examined.

The aim of this article is to utilize the clinical information collected in the ADHD Study in order to assess convergent and discriminant validity of parent questionnaires for assessing symptoms of Attention-Deficit/Hyperactivity Disorder (ADHD), Oppositional Defiant Disorder (ODD), Conduct Disorder (CD), and Separation Anxiety.

Based on previous research indicating co-occurring symptoms of ADHD, ODD, CD, and Separation Anxi-

ety (11), we expected the symptom load of different symptom groups to be correlated. More importantly, a strong correlation of symptom assessments for one symptom group through parent questionnaires and a structured clinical interview would indicate convergent validity of parent questionnaires, and higher correlations within than between symptom groups would indicate their discriminant validity.

METHODS

Sample. The ADHD Study sample consisted of participants recruited from the Norwegian Mother and Child Cohort Study (MoBa). Children who were eligible for inclusion were born between April 1, 2004 and January 31, 2008, and had mothers who had returned MoBa questionnaire Q6, which was sent to them at child age 3 years (n=33050).

The MoBa questionnaire Q6 included six questions from the Child Behavior Checklist (CBCL2-3) (12) and five items from the DSM-IV-TR criteria for ADHD (13) about hyperactivity, impulsivity, and in-attention. To identify children who might be at risk of developing ADHD, children in the MoBa population scoring at or above the 90th percentile on the sum score of these 11 questions were invited to the ADHD Study. Children with severe medical conditions and children with high scores on autistic symptoms were excluded. In total, 2798 children were invited from August 2007 to January 2011, and 1048 (37.5%) participated. In addition, a group consisting of 654 children randomly drawn from the MoBa sample of eligible children were invited, of which 149 agreed to participate, rendering a response rate of 22.5%. The overall response rate was 35%. Participating families attended a one-day clinical assessment one to eight months after responding to the MoBa questionnaire Q6.

Instruments. The Q6 contains a broad range of questions related to child health, development, and behavior. The questions used in the present study were selected items from the Child Behavior Checklist for toddlers (CBCL 2-3), the Infant and Toddler Social Emotional Assessment (14) (ITSEA), the Emotionality, Activity and Shyness Temperament Questionnaire (15) (EAS), and questions derived from the DSM-IV-TR clinical criteria for ADHD, ODD, and CD. Table 1 lists the questions, original questionnaires and scales, and the symptom group each item was assigned to for the current analysis.

In the ADHD Study, diagnostic assessment of the children was based on a Norwegian version of the Preschool Age Psychiatric Assessment (PAPA) interview with the caregiver (16). The interview was developed for preschool children from two to five years of age and includes questions about psychiatric symptoms and impairment in daily functioning. Interviewers probe until there is enough information to decide whether the symptom is present at pre-specified levels of severity. If so, its onset date is recorded along with its frequency of occurrence, when relevant. Interviews were conducted by graduate students in psychology trained in administration and scoring of the interview. A specialist in clinical psychology or child psychiatry supervised the scoring of the interview. The following PAPA sections were used to obtain symptom group scores: ADHD: all ADHD items covering the 18 diagnostic symptoms; ODD: all eight items; CD: eight of 15 items (age inappropriate items were left out); and six items for Separation Anxiety: avoidance of being alone, anticipatory stress and resistance to separation, withdrawal when attachment figure absent, physical symptoms (stomachaches) of separation, fear (anxiety) about daycare, fear/anxiety about calamitous separation. In the analyses, items from the criteria for ODD and CD are combined into the symptom group Behavior Problems.

For PAPA, interrater reliability was checked by a second rater, blind to any knowledge about the child and family, who rescored 79 randomly selected interviews from audio tape recordings. The average intra-class correlations (ICCs) were .98 for total number of ADHD symptoms, .99 for inattentive symptoms, and .97 for hyperactivity-impulsivity symptoms. ICC for anxiety and ODD symptoms were .86, and .98, respectively.

In addition to the structured diagnostic interview, the clinical assessment included questionnaires with standardized rating scales filled out by parents and preschool teachers, a short medical and neurological examination, as well as neuropsychological and neuromotor assessments. These additional assessments are not included in the present analyses.

Convergent and discriminant validity. We assessed the congruence of symptom groups by estimating multiple group confirmatory factor analyses (CFA) in Mplus (17). Because symptoms of ADHD and Behavior Problems more typically are expressed in boys, and we expected the convergent validity of parent questionnaires and structured clinical interview to depend on the length of the time period between the two assessments, we used a multiple groups model. In the model, 6 groups were defined by crossing 2 gender categories with 3 time period levels (participants were grouped into terciles based on the length of the period between both assessments). Factor loadings were constrained to be the same across all groups, and factor means and regression coefficients were allowed to vary between boys and girls (Figure 1 shows the model).

In the structural part of the model, each symptom group (ADHD, Behavior Problems, Separation Anxiety) was represented by one latent factor on which only the relevant MoBa items loaded, and by one latent factor on which only relevant items from the PAPA interview loaded. The following models were fit to test different hypotheses about the correlation of latent factors:

(**A**) Fixed factor correlations, zero between: In this model, latent factor correlations were the same for all

Table 1. *MoBa questions, source questionnaire, original scale and assigned symptom groups.* In this article the assigned symptom groups are used for clustering of items. "Beh & Man" is a group of Items assessing behavior and manner that was part of the 6th Moba questionnaire and not taken from another questionnaire. Response option were: ITSEA, CBCL 2-3, Beh & Man: 1=Not true, 2=Somewhat or sometimes true, 3=Very true or often true; EAS: 1=Very typical, 2=Quite typical, 3=Neither/nor, 4 = Not so typical, 5 = Not at all typical.

Source	Questions	Original scale	Symptom Group
EAS	Your child takes a long time to warm up to strangers	Shyness	Separation Anxiety
EAS	Your child is very friendly with strangers	Shyness	Separation Anxiety
CBCL 2-3	Clings to adults or too dependent	Anxious	Separation Anxiety
CBCL 2-3	Doesn't want to sleep alone	Sleep Problems	Separation Anxiety
CBCL 2-3	Gets too upset when separated from parents	Anxious	Separation Anxiety
ITSEA	Gets distressed when you go out and he/she is going to be looked after by family or a babysitter he/she knows	Separation Distress	Separation Anxiety
CBCL 2-3	Demands must be met immediately	Aggressive behavior	ADHD
Beh & Man	Becomes distracted or diverted by outside stimuli	Attention	ADHD
Beh & Man	Has problems keeping focused on tasks or activities	Attention	ADHD
CBCL 2-3	Can't concentrate, can't pay attention for long	Attention problems	ADHD
CBCL 2-3	Can't sit still, restless or hyperactive	Attention problems	ADHD
CBCL 2-3	Gets into everything	Attention problems	ADHD
CBCL 2-3	Quickly shifts from one activity to another	Attention problems	ADHD
Beh & Man	Doesn't seem to listen when he/she is being spoken to	Disobedience	ADHD
Beh & Man	Finds it difficult waiting his/her turn	Impulsivity	ADHD
Beh & Man	Is excessively talkative	Impulsivity	ADHD
CBCL 2-3	Can't stand waiting, wants everything now	Aggressive behavior	ADHD
ITSEA	Becomes aggressive when he/she is frustrated	Aggression/Defiance	Behavior Problems
ITSEA	Is disobedient or defiant	Aggression/Defiance	Behavior Problems
CBCL 2-3	Defiant	Aggressive behavior	Behavior Problems
CBCL 2-3	Doesn't seem to feel guilty after misbehaving	Aggressive behavior	Behavior Problems
CBCL 2-3	Gets in many fights	Aggressive behavior	Behavior Problems
CBCL 2-3	Hits others	Aggressive behavior	Behavior Problems
CBCL 2-3	Punishment doesn't change his/her behavior	Aggressive behavior	Behavior Problems
ITSEA	"Tests" other children to see whether they get angry	Peer Aggression	Behavior Problems
ITSEA	Hits, shoves, kicks and bites other children (not including siblings)	Peer Aggression	Behavior Problems

groups. In addition, the correlations of the assessment of the same symptom group through parent questionnaires and PAPA interviews was fixed to zero. This is a baseline model assuming no convergent validity. (**B**) Fixed factor correlations, fixed between: In this model the correlation between one factor from parent questionnaires and all factors from the PAPA interview were fixed to be equal. This is a baseline model assuming no discriminant validity. (**C**) Fixed factor correlations. This model assumes convergent and discriminant validity, but independence of factor correlations from gender and the period between the two assessments. (**D**) Gender depended factor correlations. (**E**) Period depended factor correlations. (**F**) Period & Gender depended factor correlations: This is the most flexible model, assuming that factor correlations depend on gender and the period between the two assessments.

We implemented the multiple group model as a mixture model with known classes because Mplus searches the maximum likelihood solution for such models, which in turn can be used to calculate penalized fit indices. Moreover, Mplus requires all entries in contingency tables to be nonzero for standard grouped CFA, which requires collapsing rare responses into more frequent categories. Modeling results were extracted with the R package MplusAutomation (18) and plotted with the R package ggplot2 (19).

Internal consistency. The internal consistency of each scale was calculated as its composite reliability (CR). The CR can be understood as the ratio of the factor variance divided by the sum of factor variance and the error variances of the factor items (20). Because error variances needed to calculate CRs can only be obtained by standard grouped CFAs, these were used here. For this analysis, rare items levels were collapsed to nearest next level in order to avoid nonzero entries in contingency tables.

RESULTS

A total of 1182 children, 560 girls and 622 boys, had sufficient data from the PAPA and from the relevant section of the MoBa Q6 to perform the analysis. At the time of participation in the ADHD Study, children were on average 3 years and 5 months old. Table 2 provides a more detailed sample description.

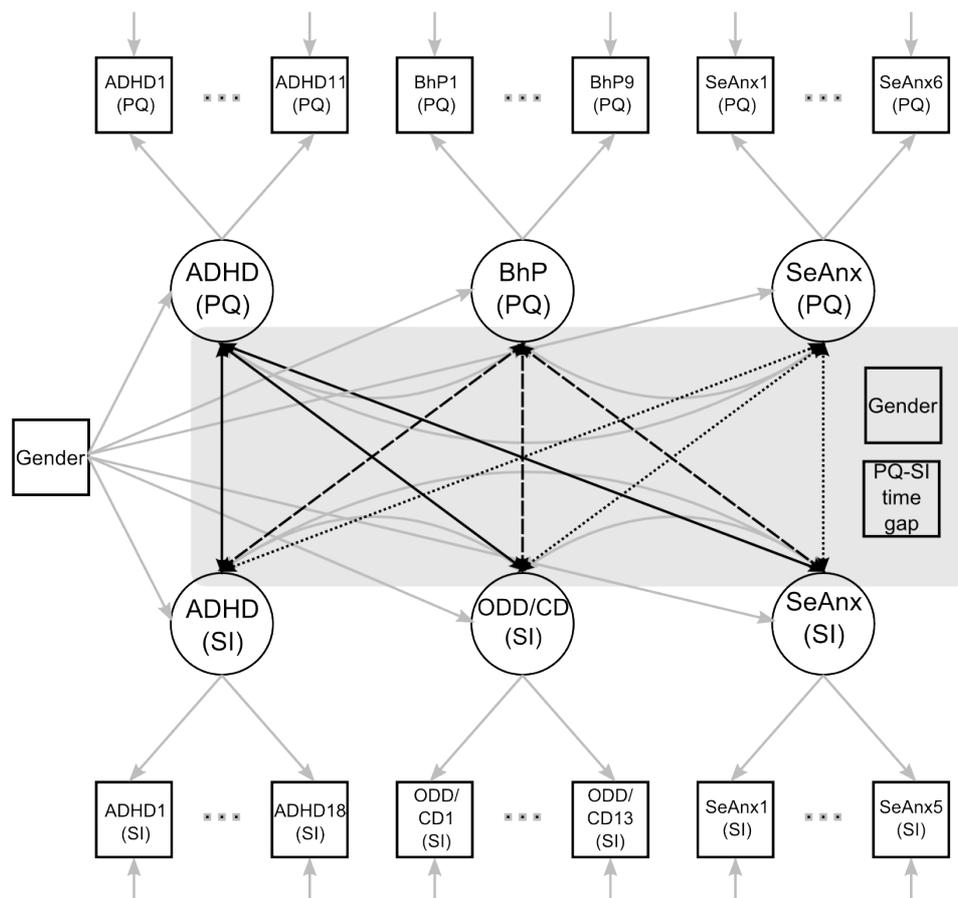*Internal consistency of symptom scores.* With the exception of the factor for Separation Anxiety symp-

Figure 1. *Model set up to test convergent and discriminant validity.* The upper part depicts the structural model for three factors, representing the three symptom groups ADHD, Behavior Problems (BhP), and Separation Anxiety (SeAnx) from the MoBa parent questionnaires (PQ). The lower part contains the model for the PAPA structured interview (SI). Gender is assumed to modify group means. Some models allow correlations within and between instruments (in the gray box) to be dependent on gender and/or the time period between PQ and SI. Convergent validity is estimated by the correlation of factors within symptom groups (vertical double headed arrows). For the baseline model assuming no convergent validity, correlations coefficients for vertical arrows are set to zero. For the baseline model assuming no discriminant validity, the coefficients for double headed arrows were fixed to be identical.

Table 2. Description of the study sample. Number of symptoms were calculated as the number of symptoms per scale for which a child scored either two or three in the PAPA interview (possible values are 0, 2, or 3).

	Child			Mother			Number sympt.			
Statistic	Age Q6 (months)	Age ADHDS (months)	delta (months)	Parity	Age (years)	Years edu	ADHD	ODD	CD	SepAnx
min	36	38	0.9	1	19	1	0	0	0	0
max	41	47	10.8	4	43	5	18	8	5	5
range	5	9	9.8	3	24	4	18	8	5	5
median	37	42	5.2	1	31	4	3	3	0	0
mean	37.3	42	5.2	1.5	31	4	4	3	0.6	0.4
std.dev	0.6	1	1.5	0.7	4	0.9	4	2	0.9	0

toms from the PAPA interview, all factors had good or very good internal consistency. Specifically, for the PAPA interview, the composite reliability (CR) was 0.936 (SE = 0.005) for ADHD, 0.807 (SE = 0.02) for Behavior Problems, and 0.684 (SE = 0.049) for Separation Anxiety. For the MoBa questionnaire the composite reliability was 0.783 (SE = 0.014) for ADHD symptoms, 0.83 (SE = 0.012) for Behavior Problems,

and 0.739 (SE = 0.018) for Separation Anxiety symptoms.

Validity of symptom evaluation. Figure 2 shows standardized factor means for boys as compared to girls. Consistent with the higher occurrence of significant ADHD symptoms in boys, the factor mean for ADHD symptoms assessed with the PAPA is higher for boys. The factor mean for ADHD from the MoBa

Table 3. *Model fits*. LL = Log Likelihood, AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion, aBIC = sample size adjusted BIC, AICC = AIC corrected for sample size. (A) and (B) are baseline models consistent with no convergent and no discriminant validity between instruments.

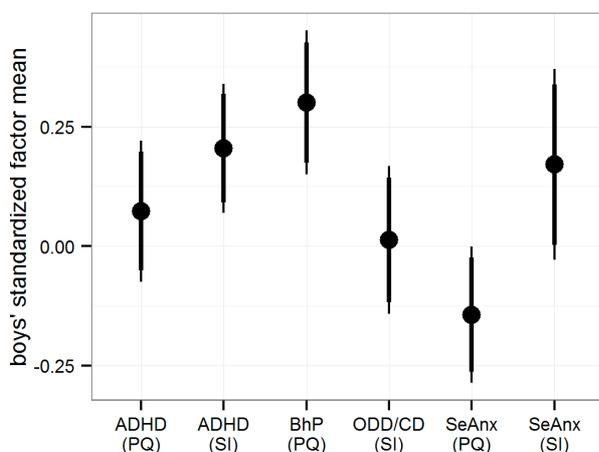| Model name	LL	AIC	aBIC	AICC	delta aBIC	delta AICC	delta AIC
(A) Fix factor correlations, zero between	-46246.86	92905.71	93286.11	92998.92	338.495	340.99056	344.034
(B) Fix factor correlations, fix between	-46171.34	92748.69	93123.55	92838.91	175.932	180.98322	187.011
(C) Fix factor correlations	-46071.84	92561.68	92947.61	92657.93	0	0	0
(D) Period depended factor correlations	-46049.50	92600.99	93064.49	92746.40	116.874	88.47390	39.317
(E) Gender depended factor correlations	-46056.18	92572.36	92997.08	92691.62	49.462	33.69226	10.683
(F) Period & Gender dependent factor correlations	-46005.55	92639.09	93218.92	92884.22	271.310	226.29711	77.417

Figure 2. *Factor means for boys*. Factor means for girls are fixed to 0. Thin (thick) lines show 95% (90%) confidence intervals. BhP: Behavior problems. ODD/CD: Oppositional Defiant Disorder (ODD) and Conduct Disorder (CD) symptoms. SeAnx: Separation anxiety symptoms. PQ: Parent Questionnaire (MoBa). SI: Standardized interview (ADHD Study). Factor means are obtained from the best model (F).

parent questionnaire does not differ meaningfully between boys and girls, presumably due to the sampling plan that oversampled children with high symptom scores independent of gender.

To assess congruence of symptom evaluations from parent questionnaires with structured clinical interviews, we estimated several grouped CFAs. Table 3 shows model fits for the estimated models. The first key result is that baseline models assuming either (A) zero correlations of latent factors between instruments or (B) equal correlations between each MoBa factor with the three ADHD Study factors, perform worse than models consistent with convergent and discriminant validity of questionnaires. Of these models, the best model had identical between factor correlations for all groups (only the least conservative model fit criterion, the AIC, indicated dependence of factor correlations on gender). This result suggests that convergent and discriminant validity were neither

influenced by gender, nor by the duration of the period between the two assessments.

Further examination of the latent factor correlations in the fixed factor correlations model supports the discriminant and convergent validity of the parent questionnaires. Convergent validity is supported by generally significant correlations between factors that measure the same symptom group, but are informed by either parent questionnaire or structured interview (see Table 4 and Figure 3). Discriminant validity is supported by the finding that, for between instruments correlations, the correlations of same-symptom group factors is generally higher than the correlation between different-symptom group factors. Only for Separation Anxiety did the confidence intervals of correlation coefficients overlap, which is mainly due to the large standard error for the correlation involving Separation Anxiety assessed in the structural interview.

As can be seen from the strong within-instrument correlation between symptoms of ADHD and Behavior Problems, our results also suggest an important co-occurrence of these symptoms. Indeed, these correlations have the same strength as the within symptom group correlations of ADHD and Behavior Problems across instruments, as indicated by their largely overlapping confidence intervals.

Discussion

We find that symptoms of ADHD, Behavior Problems, and Separation Anxiety, assessed with parent questionnaires in the MoBa Study, in 3 year old children are correlated with symptoms assessed in a structured clinical interview conducted on average 5 months later. That is, the scale items intended to measure symptoms of inattention/hyperactivity, Behavior Problems, and to a lesser degree Separation Anxiety, have both convergent and discriminant validity when compared with the results of a structured clinical interview.

Compared to other examinations of convergent and discriminant validity of parent questionnaires for ADHD symptoms (6,21,22), we observed lower con-
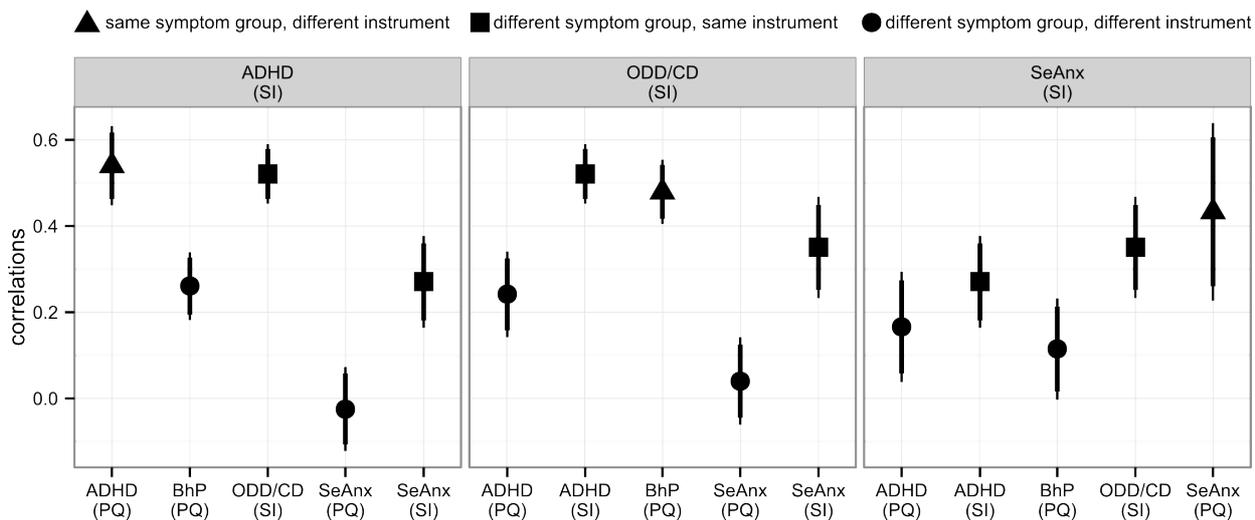
Figure 3. *Correlations between symptom scores.* The three panels show correlations between latent factors from PAPA items with all other latent factors. Triangles indicate correlations of latent factors for the same symptom group assessed at different time points, and show that this correlation is generally high compared to correlations with factors measuring different symptom groups at the same time or in Q6. The high correlation between symptoms of inattention/hyperactivity and Behavior Problems highlight the co-occurrence of ADHD and ODD/CD. The low frequency of anxiety symptoms leads to low variance and thus also to low correlations with anxiety symptoms. Thin (thick) lines show 95% (90%) confidence intervals. Symbols indicate if the correlation is between symptom groups assessed with same or different instruments. BhP: Behavior problems. ODD/CD: Oppositional Defiant Disorder (ODD) and Conduct Disorder (CD) symptoms. SeAnx: Separation anxiety symptoms. PQ: Parent Questionnaire (MoBa). SI: Standardized interview (ADHD Study).

Table 4. Standardized correlations (standard errors) between latent factors. Correlations of different symptom groups within one assessment instrument are italicized. Correlations between factors for the same symptom group from different instruments are in bold typeface. *: p< .0001, [a]: p=.623, [b]: p=.03, [c]: p=.001, [d]: p=.436, [e]: p=.011, [f]: p=.056. BhP: Behavior problems. ODD/CD: Oppositional Defiant Disorder (ODD) and/or Conduct Disorder (CD) symptoms. SeAnx: Separation anxiety symptoms. PQ: Parent Questionnaire (MoBa). SI: Standardized interview (ADHD Study).

	BhP (PQ)	SeAnx (PQ)	ADHD (SI)	ODD/CD (SI)	SeAnx (SI)
ADHD (PQ)	*0.551 (0.064)**	*0.152 (0.07)*[b]	**0.54 (0.047)**	0.242 (0.051)*	0.166 (0.065)[e]
BhP (PQ)		*0.173 (0.054)*[c]	0.261 (0.04)*	**0.479 (0.038)**	0.115 (0.06)[f]
SeAnx (PQ)			−0.025 (0.05)[a]	0.04 (0.052)[d]	**0.433 (0.105)***
ADHD (SI)				*0.521 (0.035)**	0.271 (0.054)*
ODD/CD (SI)					*0.351 (0.06)**

vergent validities (r = .54 for ADHD, r = .48 for Behavior Problems, and r = .43 for Separation Anxiety), which fall below the commonly referred threshold of r = .80 as a sign of good convergent validity (7). The relatively lower correlations we observed are at least partially due to both strengths and weaknesses of the current study. First, whereas other investigations of convergent validity typically use assessments employing the same questionnaire by different raters at approximately the same time (7,9,21), a strength of the current study is the use of a later, structured clinical interview as the comparison criterion. While the time lag and using a different instrument suppress the estimated convergent validity, our results are likely of higher practical relevance because they are primarily driven by stable symptoms and obtained through comparison with a validated clinical interview. Still, the low estimate of convergent validity likely also reflects limitations of the current study. In particular, while the

oversampling of children with a high ADHD symptom load is desirable for the main goal of the ADHD study, it also leads to a reduced variability of ADHD symptom load in the sample. This reduced variability, in turn, likely suppresses the estimated convergent validity. Lastly, compared to the population of children with symptoms of Separation Anxiety or Behavior Problems, children with such symptoms in the ADHD Study are also more likely to have ADHD symptoms, so that the current study potentially underestimates the discriminant validity of questionnaires. The overrepresentation of children with co-occurring symptoms of ADHD and especially Behavior Problems is also reflected in the high correlation of these symptom groups' factors. Complementary explanations could be that symptoms in preschool children are less differentiated, or that the population-wide co-occurrence of mental health symptoms is higher in young preschoolers. Still, others have reported a high co-occurrence of ADHD

and ODD (11,23,24), suggesting that our results are broadly in line with previous research.

One unexpected result of our analyses was that the effect of gender and the length of the period between two assessments on factor correlations was not strong enough to justify the inclusion of these variables into the grouped CFA model. While such a result could be explained by unreliable and unspecific assessment of all symptoms (i.e. when all assessments are unreliable, adding variables to explain their values is not effective), this is unlikely given that the correlation between factors for the same symptom is around 0.45. Instead, it is plausible that symptoms are already relatively stable in young preschool children, thus allowing for identifying convergent validity of instruments even when they are used on average 5 months apart.

An important methodological consideration is that we assessed validity not based on straight forward sum scores, but based on a measurement model in which items that are more representative of symptom groups have a larger influence on the symptom group score. Compared to the use of simple sum scores, the advantage of the inclusion of a measurement model is that symptom group scores are more reliable, which leads to higher estimates of convergent validity, if two instruments indeed measure the same symptom group.

In sum, our results show that relatively short parent questionnaires predict results of later interview assessments for symptoms of ADHD, ODD/CD, and Separation Anxiety in preschool children. These results suggest that at least some mental health problems are relatively stable over a time span of about five months in preschool children, and that parent questionnaires can potentially be used to identify children with clinically significant symptoms. Future research should attempt to further improve and test parent questionnaires' validity in representative samples, and explicitly test their utility for screening.

REFERENCES

1. Costello EJ, Egger H, Angold A. 10-year research update review: the epidemiology of child and adolescent psychiatric disorders: I. Methods and public health burden. *J Am Acad Child Adolesc Psychiatry* 2005;**44**(10):972-86.
2. Andersson HW, Ådnanes M, Hatling T. Nasjonal kartlegging av tilbud om diagnostisering og helhetlig behandling av barn og ungdom med hyperkinetiske forstyrrelser/ADHD. SINTEF, 2004.
3. Sonuga-Barke EJS, Koerting J, Smith E, McCann DC, Thompson M. Early detection and intervention for attention-deficit/hyperactivity disorder. *Expert Rev Neurother* 2011;**11**(4):557-63.
4. McClellan JM, Speltz ML. Psychiatric diagnosis in preschool children. *J Am Acad Child Adolesc Psychiatry* 2003;**42**(2):127-8; author reply 128-30.
5. Magnus P, Irgens LM, Haug K, Nystad W, Skjærven R, Stoltenberg C, et al. Cohort profile: the Norwegian Mother and Child Cohort Study (MoBa). *Int J Epidemiol* 2006;**35**(5):1146-50.
6. Johnson S, Hollis C, Marlow N, Simms V, Wolke D. Screening for childhood mental health disorders using the Strengths and Difficulties Questionnaire: the validity of multi-informant reports. *Dev Med Child Neurol* 2014;**56**(5):453-9.
7. Burns GL, Walsh JA, Servera M, Lorenzo-Seva U, Cardo E, Rodríguez-Fornells A. Construct validity of ADHD/ODD rating scales: recommendations for the evaluation of forthcoming DSM-V ADHD/ODD scales. *J Abnorm Child Psychol* 2013;**41**(1):15-26.
8. Kessler RC, Adler L, Ames M, Demler O, Faraone S, Hiripi E, et al. The World Health Organization Adult ADHD Self-Report Scale (ASRS): a short screening scale for use in the general population. *Psychol Med* 2005;**35**(2):245-56.
9. Bussing R, Fernandez M, Harwood M, Wei Hou, Garvan CW, Eyberg SM, et al. Parent and teacher SNAP-IV ratings of attention deficit hyperactivity disorder symptoms: psychometric properties and normative ratings from a school district sample. *Assessment* 2008;**15**(3):317-28.
10. Chen WJ, Faraone SV, Biederman J, Tsuang MT. Diagnostic accuracy of the Child Behavior Checklist scales for attention-deficit hyperactivity disorder: a receiver-operating characteristic analysis. *J Consult Clin Psychol* 1994;**62**(5):1017-25.
11. Wilens TE, Biederman J, Brown S, Tanguay S, Monuteaux MC, Blake C, et al. Psychiatric comorbidity and functioning in clinically referred preschool children and school-age youths with ADHD. *J Am Acad Child Adolesc Psychiatry* 2002;**41**(3):262-8.
12. Achenbach TM. Child Behavior Checklist for Ages 2-3. University of Vermont, Department of Psychatry, 1992.
13. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*: DSM-IV-TR®. American Psychiatric Association, 2000.
14. Carter AS, Briggs-Gowan MJ, Jones SM, Little TD. The Infant-Toddler Social and Emotional Assessment (ITSEA): factor structure, reliability, and validity. *J Abnorm Child Psychol* 2003;**31**(5):495-514.

15. Mathiesen KS, Tambs K. The EAS temperament questionnaire – factor structure, age trends, reliability, and stability in a Norwegian sample. *J Child Psychol Psychiatry* 1999;**40**(3):431-9.
16. Link Egger H, Angold A. The Preschool Age Psychiatric Assessment (PAPA): A structured parent interview for diagnosing psychiatric disorders in preschool children. In: DelCarmen-Wiggins R, Carter AS, editors. *Handbook of Infant and Toddler Mental Health Assessment*. Oxford University Press, 2004.
17. Mutén LK, Muthén BO. *Mplus User's Guide*. Los Angeles, CA: Mutén & Muthén, 2012.
18. Hallquist M, Wiley J. MplusAutomation: Automating Mplus Model Estimation and Interpretation. 2013.
19. Wickham H. *ggplot2: elegant graphics for data analysis*. Springer New York, 2009.
20. Raykov T. Behavioral scale reliability and measurement invariance evaluation using latent variable modeling. *Behav Ther* 2004;**35**(2):299-331.
21. Tehrani-Doost M, Shahrivar Z, Pakbaz B, Rezaie A, Ahmadi F. Normative data and psychometric properties of the child behavior checklist and teacher rating form in an Iranian community sample. *Iran J Pediatr* 2011;**21**(3):331-42.
22. Wichstrøm L, Berg-Nielsen TS. Psychiatric disorders in preschoolers: the structure of DSM-IV symptoms and profiles of comorbidity. *Eur Child Adolesc Psychiatry* 2014;**23**(7):551-62.
23. Greene RW, Biederman J, Zerwas S, Monuteaux MC, Goring JC, Faraone SV. Psychiatric comorbidity, family dysfunction, and social impairment in referred youth with oppositional defiant disorder. *Am J Psychiatry* 2002;**159**(7):1214-24.
24. Lavigne JV, Lebailly SA, Hopkins J, Gouze KR, Binns HJ. The prevalence of ADHD, ODD, depression, and anxiety in a community sample of 4-year-olds. *J Clin Child Adolesc Psychol* 2009;**38**(3):315-28.