Title

Factor Structure of the Children's Sleep Habits Questionnaire in Young Children with and Without Autism

Journal

Journal of Autism and Developmental Disorders, 51(9)

ISSN

0162-3257

Authors

Hatch, Burt
Nordahl, Christine Wu
Schwichtenberg, AJ
et al.

Publication Date

2021-09-01

DOI

10.1007/s10803-020-04752-5

Peer reviewed

HHS Public Access

Author manuscript

J Autism Dev Disord. Author manuscript; available in PMC 2022 September 01.

Published in final edited form as:

J Autism Dev Disord. 2021 September ; 51(9): 3126–3137. doi:10.1007/s10803-020-04752-5.

Factor structure of the Children's Sleep Habits Questionnaire in young children with and without autism

Burt Hatch[1], **Christine Wu Nordahl**[1], **AJ Schwichtenberg**[2], **Sally Ozonoff**[1], **Meghan Miller**[1]

[1]Department of Psychiatry & Behavioral Sciences and MIND Institute, University of California, Davis

[2]Department of Human Development and Family Studies, College of Health and Human Sciences, Purdue University, West Lafayette, IN

Abstract

The Children's Sleep Habits Questionnaire (CSHQ) is often used to assess sleep in children with autism spectrum disorder (ASD), but little is known about its factor structure in younger children with ASD. We evaluated alternative factor structures and measurement invariance for CSHQ items in 2- to 4-year-olds with ASD or typical development (TD). Bifactor models indicated subscales' variance was subsumed by a general factor predominantly reflecting sleep initiation and nighttime awakening items. A factor consisting of 7 of these items was measurement invariant across ASD and TD. Thus, comparisons between young children with ASD and TD is appropriate for a measure composed of 7 CSHQ items relating to sleep initiation and awakenings but not for other CSHQ item composites.

Relative to typically developing children, children with autism spectrum disorder (ASD) are more likely to experience sleep-related problems such as resisting going to bed, difficulty falling asleep, frequently waking during the night, and early morning waking (Díaz-Román et al., 2018). Indeed, sleep problems are amongst the most common difficulties that co-occur with ASD, affecting 50–80% of diagnosed individuals (Couturier et al., 2005; Richdale & Schreck, 2009; Souders et al., 2009). One approach to measuring sleep problems is caregiver rating scales. These provide a convenient method of assessing sleep-related problems in the home environment and measure caregiver perceptions including those relating to difficulty self-soothing, problems with sleep onset and maintenance, nighttime anxiety, and daytime sleepiness (Díaz-Román et al., 2018; Mancini et al., 2019; Owens et al., 2000). Alongside objective measures of sleep, caregiver rating scales can be important tools for understanding how to improve the delivery of parenting-based interventions, given that caregiver perception of sleep problems affects intervention engagement as well as parenting stress (Martin et al., 2019; Quach et al., 2011; Robinson & Richdale, 2004).

Although originally developed to assess sleep difficulties in typically developing children, the Children's Sleep Habits Questionnaire (CSHQ; Owens, Spirito, & McGuinn, 2000) has

Address correspondence to Burt Hatch, Ph.D., MIND Institute, University of California Davis Health System, 2825 50th Street, Sacramento, CA 95817. Phone: 916-734-4381, bhatch@ucdavis.edu.

The authors report no conflicts of interest.

often been used to study sleep in children with ASD (Hodge, Parnell, Hoffman, & Sweeney, 2012; Souders et al., 2009; Goodlin-Jones, Sitnick, Tang, Liu, & Anders, 2008; Reynolds et al., 2019). Developed by Owens et al. (2000), the CSHQ includes items that pertain to different sleep problems. Based upon a conceptual understanding of sleep problems and face validity of items, it was originally proposed that CSHQ items could be used to measure either specific sleep problems, by summing items grouped into eight different subscales, or more generalized sleep problems, by summing a broad set of items across the different subscales (Owens et al., 2000).

The CSHQ was initially intended to measure sleep problems in 4- to 10-year-olds (Owens et al., 2000). However, despite notable changes in sleep through infancy and childhood (Galland et al., 2012) the CSHQ has also been used with much younger children both with and without ASD (Levin & Scher, 2016; Reynolds et al., 2019; Richdale & Schreck, 2019; Wang et al., 2019). In particular, the CSHQ has been used to examine group differences between those with and without ASD in samples that include children as young as 2 years of age (Levin & Scher, 2016; Reynolds et al., 2019; Richdale & Schreck, 2019). For instance, one study of 2- to 4-year-olds found that, compared to those with typical development, those with ASD were higher on three CSHQ subscales purportedly measuring problems with sleep onset delay, parasomnias, and daytime sleepiness (Levin & Scher, 2016). Another recent investigation conducted with 2- to 5-year-olds found that the proportion of children surpassing a CSHQ total score cutoff value was significantly higher in children with ASD than those with other developmental delays, and those sampled from the general population (Reynolds et al., 2019). These studies highlight an interest in using the CSHQ to assess how young children with ASD differ from those with typical development in terms of sleep problems, either at a general level or at the level of more specific subscales.

Despite interest in using the CSHQ to investigate sleep problems in children with ASD, little is known about its factor structure in this context, particularly in children younger than the age for which the scale was originally designed. Investigating a scale's factor structure using an exploratory or confirmatory factor analysis framework provides a valuable component to the process of scale development and refinement by giving insight into the number and composition of subscale constructs with which the individual items best align (Timmerman et al., 2018). Moreover, investigating how individual items fit within bifactor models can evaluate the extent to which they reflect shared variance in a general overarching factor versus reflecting a subscale factor (Rodriguez et al., 2016a). Studies examining the factor structure of a measure can also include testing for measurement invariance to determine if items represent constructs equivalently across different groups (e.g., ASD versus typically developing) (Kline, 2016). This is an important step for scales used to examine group differences given that the presence of measurement invariance signifies that a measure is not biased and is measuring the same construct the same way across groups (Kline, 2016).

To date no study has sought to examine the factor structure of the CSHQ in a sample that includes both young children with typical development and young children with ASD, nor have any studies examined whether the scale exhibits measurement invariance across these groups. One previous study examined the CSHQ factor structure with a sample of typically developing 2- to 5-year-old children (Sneddon et al., 2013), and another with a sample of

children with ASD between 4–10 years of age (Katz et al., 2018). For both, the objective was to examine the number of different constructs reflected by CSHQ items. Notably, although results from each study indicated that CSHQ items related to four specific subscale factors, the precise item-to-subscale factor configuration differed. In the study of typically developing 2- to 5-year-olds, 24 items were subdivided into subscales of *Sleep Initiation*, *Sleep Distress*, *Sleep Transition*, and *Sleep Duration*. In the study of 4- to 10- year-olds with ASD, 23 items were subdivided into *Sleep Initiation and Duration, Night Waking/ Parasomnias, Daytime Alertness*, and *Sleep Anxiety/Co-sleeping* subscales. Thus, these two studies present two alternative factor structures for the configuration of CSHQ items into four different subscales. However, since one study focused on typically developing young children, and the other on older children with ASD, it remains unclear whether these factor structures might represent CSHQ subscales that are reliable for comparing young children who are typically developing to those with ASD.

Beyond the factor structure of the CSHQ subscales, two additional methodological considerations remain to be addressed. First, previous studies have not assessed bifactor models for CSHQ items. The importance of this is underscored by bifactor model analyses of other sleep rating scales similar to the CSHQ which demonstrate that a general sleep factor often accounts for a majority of the variance in subscale factors (Li et al., 2019; Mancini et al., 2019; Pushpanathan et al., 2018). Second, the measurement invariance of the CSHQ across typically developing children and those with ASD has remained untested. This raises concerns for interpreting group differences on CSHQ scales between children with ASD versus typical development, since the same set of items may not represent the same construct the same way for each group (Kline, 2016).

In summary, the factor structure of the CSHQ is not well understood in the context of comparisons of young children with ASD and typical development. The aim of the present study is to expand upon the two aforementioned studies that examined the factor structure of the CHSQ; one derived a factor structure for a set of 24 CSHQ items with typically developing 2- to 5-year-olds (henceforth referred to as CSHQ-TD item set; Sneddon et al., 2013) and the other derived a factor structure for a set of 23 CSHQ items with 4–10 year-olds with ASD (henceforth referred to as the CSHQ-ASD item set; Katz et al., 2018). First, we assess whether the subscale factor structure indicated by these two previous studies can be replicated in a sample of preschool age children that includes those with ASD and those who are typically developing. Second, we investigate whether the factor structures for each item set can be extended to a bifactor model and assess the extent to which reliable subscale variance is independent of variance in a general factor. Finally, we examine measurement invariance of factors across children with typical development and ASD in order to determine whether the CSHQ measures sleep difficulties similarly across these groups.

Method

Participants

Families who provided data for this study were participating in four longitudinal research projects across two sites. One site, the MIND Institute, University of California, Davis, houses three studies, the MIND Infant-Sibling Study, Girls with Autism - Imaging and

Neurodevelopment (GAIN) study, and Autism Phenome Project (APP); the other study, the Purdue Infant-Sibling study, is at the College of Health and Human Sciences, Purdue University. Each study obtained parent responses to the CSHQ and diagnostic information from structured research assessments conducted by, or under the supervision of, a licensed psychologist or pediatrician, that was used in the classification of children to an ASD group or TD group. All studies were conducted under the approval of the respective universities' Institutional Review Boards and informed consent was obtained from parents before assessments.

Two studies, the MIND Infant-Sibling study and the Purdue Infant-Sibling study, included younger (high-risk) siblings of children with ASD or younger (low-risk) siblings of children with typical development. Participants were enrolled by 9 (the MIND Infant-Sibling study) or 18 (The Purdue Infant-Sibling study) months of age. Exclusion criteria for high-risk siblings included birth before 32 weeks of gestation and a known genetic disorder (e.g., Fragile X syndrome) in the older affected sibling. Primary inclusion criterion for low-risk siblings was status as a younger sibling of a child/children deemed to be typically developing, confirmed by an intake screening questionnaire and scores below the ASD range on the Social Communication Questionnaire (SCQ; Rutter et al., 2003). Exclusion criteria for the low-risk groups were birth before 36 weeks of gestation, developmental, learning, or medical conditions in any older sibling, and ASD in first-, second-, or third-degree relatives.

The GAIN and APP studies recruited children at 24- to 48- months of age who were either typically developing or had confirmed diagnoses of ASD. These studies followed identical research protocols. At recruitment, all participants were required to be native English speakers; be ambulatory; have no suspected vision or hearing problems; and have no known genetic disorders or neurological conditions.

Children from all studies were classified into an ASD group (n = 269) or a typically developing (TD) group (n = 229). All children classified in the ASD group had scores at or above the ASD cutoff on the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2012; Lord et al., 2000) *and* met *DSM-5* criteria for ASD (for the Purdue Infant-Sibling study, GAIN study, and APP) or *DSM-IV-TR* criteria for Autistic Disorder or Pervasive Developmental Disorder-Not Otherwise Specified (for the MIND Infant-Sibling study, GAIN study, and APP). Children were classified into the TD group if they did not meet ASD classification criteria and had an ADOS score more than 3 points below the ASD cutoff (for the MIND and Purdue Infant-Sibling studies) or had scores below the Social Communication Questionnaire (Rutter et al., 2003) clinical cutoff (< 11; for the GAIN study and APP). Children were excluded from the TD group (n = 47 for The MIND Infant-Sibling study, n = 36 for The Purdue Infant-Sibling study, n = 9 for APP) if two or more Mullen Scales of Early Learning (Mullen, 1995) subscale scores fell below the normative mean by at least 1.5 standard deviations, or one or more Mullen subtest scores fell below the normative mean by at least 2 standard deviations. The number of participants with missing group classification information was n = 1 for The MIND Infant-Sibling study and n = 15 for The Purdue Infant-Sibling study.

Data were pooled across studies for children classified to the ASD or TD groups as follows: First, participants were included if parents completed any CSHQ items when children were 24- to 48- months of age. A single timepoint was selected for each participant. In instances when CSHQ data was available for multiple visits within the 24- to 48-month age range, only data for the most recent visit was included (n = 40 from The MIND Infant-Sibling study, n = 42 from The Purdue Infant-Sibling study), ensuring that the sample age diverged as little as possible from ages reported in previous studies. CSHQ data were excluded if collected at any other visit than the one at which group classification was determined (n = 2 from The Purdue Infant-Sibling study). The resulting sample consisted of n = 266 children with ASD (mean age = 37.43 months, 71% male) and n = 224 children classified as TD (mean age = 35.75 months, 57% male). Second, because a primary purpose of this study was to examine measurement invariance across children with ASD and TD children, participants were selected so that these groups were matched for age and sex using a nearest-neighbor matching algorithm from the *MatchIt* package (Ho et al., 2011) implemented in R. This resulted in a sample of 224 participants in each group, characteristics of which are presented in Table 1. Females were well represented in both the ASD and TD samples (34% and 43% respectively); the ASD and TD groups did not differ by age or sex (p > .05).

Measures

Children's Sleep Habits Questionnaire—The original version of the CSHQ (Owens et al., 2000) includes 33 items that, for the 'most recent typical week', are rated by caregivers as occurring 'Usually' (5–7 nights/week), 'Sometimes' (2–4 nights/week), or 'Rarely' (0–1 nights/week). Based on a conceptual understanding of common sleep problems in a community sample of 4- to 10-year-olds, the original version specified that the summation of all 33 items can provide an index of sleep problems at a general level, while eight subscales can be obtained by the summation of different subsets of items (*bedtime resistance*, *sleep onset delay*, *sleep duration*, *sleep anxiety*, *night-waking*, *parasomnias*, *disordered breathing*, and *daytime sleepiness*).

We sought to examine the factor structure for two alternative CSHQ item sets, CSHQ-TD (Sneddon et al., 2013) and CSHQ-ASD (Katz., 2018). The CSHQ-TD consists of 24 items divided into 4 subscales: 9 loading onto a *Sleep Initiation* subscale, 8 loading onto a *Sleep Distress* subscale, 4 loading onto a *Sleep Transition* subscale, and 3 loading onto a *Sleep Duration* subscale. Of the original 33 items 9 were excluded because either 90% or more of respondents endorsed "rarely" (4 items), the authors deemed content as being "less appropriate for preschool-age children" (3 items), or low factor loading (< 0.30, 2 items). The CSHQ-ASD consists of 23 items, also divided into 4 subscales: 6 loading onto a *Sleep Initiation and Duration* subscale, 5 onto a *Sleep Anxiety/Co-Sleeping* subscale, 6 onto a *Night Waking/Parasomnias* subscale, and 6 onto a *Daytime Alertness* subscale. Items were excluded because either 85% or more of endorsements were "rarely" or "usually" (4 items), or the item had low factor loadings (< 0.40, 7 items).

Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2012; Lord et al., 2000)—The ADOS is a structured observational assessment with modules designed for different levels of expressive language that measure the social and communication behaviors

indicative of autism. The assessment provides opportunities for interaction and play and "presses" for certain target behaviors within these interactions. An algorithm is used with cut-offs for autism and autism spectrum disorder. The ADOS has high test–retest reliability as well as good internal consistency (Lord et al., 2000). The ADOS was used to classify children into the ASD or TD groups.

Mullen Scales of Early Learning (MSEL; Mullen, 1995)—The Mullen is a standardized developmental assessment of verbal and non-verbal skills for children under 68 months of age. It provides an overall index score as well as verbal subscale scores (Receptive Language and Expressive Language) and non-verbal subscale scores (Visual Reception and Fine Motor). The Mullen has good test–retest reliability and high internal consistency (Mullen, 1995). The Mullen was used for sample characteristics and to exclude children from the TD group with delayed development.

Data analytic plan

Confirmatory factor analyses.—To compare alternative factor structure models, we used confirmatory factor analyses (CFA), implemented with a weighted least square mean and variance (WLSMV) adjustment estimator and delta-parameterization using the *lavaan* package (Russeel, 2012) for R. The CFA approach was employed because the focus was on testing already defined subscale factor structures for the two alternative CSHQ item sets based on the prior literature (Katz et al., 2018; Sneddon et al., 2013). For each item set, we first assessed whether a general factor model (all items loading on one common factor), correlated subscale factor model (items loading onto a specified subscale factor), or bifactor model (items loading onto both a specified subscale and a common factor) best represented the CSHQ item-level data spanning across the ASD and TD groups. Comparative Fit Index (CFI) and Root Mean Square Error of Approximation (RMSEA) were used to assess model fit with values of RMSEA < 0.05 and CFI value > 0.95 considered indicative of good model fit (West et al., 2012). Models nested within each item set were compared using Chi-square difference tests. Models best fitting the observed data for each item set were compared using Voung's test for non-nested models (Vuong, 1989), implemented in the *nonnest2* package (Merkle & Dongjun, 2018) for R, adapted for use with ordinal data using code shared by the package creator (Dr. Merkle, personal correspondence).

For bifactor models, we report psychometric indices that facilitate interpretation of the proportion of variance attributable to the general factor and subscale factors; namely Omega Total, Omega Total-Subscale, Omega Hierarchical, Omega Hierarchical-Subscale, Explained Common Variance (ECV) and Percentage of Uncontaminated Correlations (PUC) (Rodriguez et al., 2016a). Omega Total represents the proportion of all variance accounted for by *both* the general and specific factors together. A variant of this index, Omega Total-Subscale, represents the proportion of variance in the subscale attributable to all sources of common variance (i.e., variance of subscale items accounted for by both the given subscale factor and general factor). Omega Total is analogous to calculating Chronbach's alpha for all items and Omega Total-Subscale is analogous to calculating Chronbach's alpha for the subset of items of a given subscale; however, these Omega indices are more realistic representations of reliability than alpha (Bentler, 2009; Rodriguez et al., 2016a). Omega

Hierarchical for the general factor represents the proportion of variance that is accounted for by individual differences in the total score of all items *after* accounting for systematic variance associated with subscales. Likewise, Omega Hierarchical-Subscale represents the proportion of variance that is accounted for by individual differences in the subscale *after* accounting for differences due to the general factor. ECV and PUC provide an indication of the extent to which the general factor represents a multidimensional construct. ECV represents the proportion of all variance explained by all factors attributable to the general factor, and PUC is the proportion of all possible correlations between items that are influenced by the general factor. When both PUC and ECV are > 0.70, it can be considered that the general factor is strong, in that estimating a unidimensional model may offer a more parsimonious solution than the bifactor structure without introducing extensive bias in the factor loadings (Rodriguez et al., 2016b). At an item level, we also calculated Item Explained Common Variance (I-ECV), which represents the proportion of common variance attributable to the general factor for each individual item (Rodriguez et al., 2016a). Taken together, these indices can be used to identify which *factors* account for the majority of reliable variance and which *items* are most representative of these factors.

Measurement Invariance.—After identifying items that represented factors with reliable variance, we tested whether these items exhibited groupwise measurement invariance. Measurement invariance tests were conducted using a CFA framework (Meredith, 1993) with theta-parameterization and identification constraints for ordinal data (Wu & Estabrook, 2016). An initial *Configural* model tested whether constructs had the same pattern of factor loadings across groups. The second step tested whether the strength of item loadings differed significantly across groups by comparing the *Configural* model to a *Weak* model (a model with item loadings constrained to be equal across groups). The third step tested whether item intercepts differed significantly across groups by comparing the *Weak* model to a *Strong* model (a model that constrained item loadings and intercepts to be equal across groups). The fourth step tested whether item residual variance differed significantly across groups by comparing the *Strong* model to a *Strict* model (a model that constrained item loadings, item intercepts, and item residual variance to be equal across groups). A lack of measurement invariance for model comparisons was considered if Chi-square difference test was significant at $p < 0.05$, CFI values differed by more than 0.01, or RMSEA values differed more than 0.015 (Cheung & Rensvold, 2002).

Item descriptive statistics and missing data.—All items had skewness < 2.3, and kurtosis < 5 (Table S1), acceptable for CFA models implemented with WLSMV estimation (Hoyle, 2012, p.173). Missing data for individual CSHQ items ranged between 0 to 6.9% with most items having < 2% of missing data. The vast majority (87.7%, $n = 393$) of participants had no missing CSHQ data; 10.5% ($n = 58$) were missing 1–5 items, and 1.8% ($n = 8$) were missing 6–12 items. The two diagnostic groups did not significantly differ with respect to the number of missing items ($p = 0.27$). To account for missing data, all models were estimated on 200 datasets generated using multiple imputation by chained equations using the *mice* (Buuren & Groothuis-Oudshoorn, 2011) package for R, and estimates pooled using Rubin's rules (1987) as implemented in the *semTools* package (Jorgensen et al., 2020).

RESULTS

Alternative Factor structures

For each item set (i.e., CSHQ-ASD, CSHQ-TD), the fit of bifactor models was significantly better than correlated models, while the correlated model fit was significantly better than the general factor model (Table 2). Vuong's (1989) test for non-nested models indicated that the bifactor model for the CSHQ-ASD item set was a better representation of data than the bifactor model for CSHQ-TD item set ($z = -22.44$, $p < 0.001$). The CSHQ-ASD item set bifactor model fit the data well (RMSEA = 0.046, CFI = 0.983).

Given that the bifactor model for the CSHQ-ASD item set best fit the data, parameter estimates for this model are presented in Table 3 and described further (parameter estimates for additional models are presented in Tables S2–S6). Omega Total for all items was 0.93, meaning that the model accounted for 93% of the observed variance. Omega Hierarchical for the general factor was 0.73, meaning that 73% of the unit-weighted total score variance is attributable to the general factor, with 20% (i.e., difference between 93% and 73%) attributable to the four subscale factors. The ratio of Omega Hierarchical to Omega Total (i.e.,0.73 divided by 0.93) indicates that 78% of reliable variance can be attributable to the general factor. The ECV (i.e., proportion of the variance from all factors explained by the general factor) was 0.44 and the PUC (i.e., proportion of correlations between items influenced by the general factor) was 0.78. These values suggest that the general factor cannot necessarily be considered a unidimensional "broadband" construct that is equally common across subscales (Rodriguez et al., 2016a). The subscale with the highest average loadings on the general factor was *Initiation/Duration* (0.64), followed in descending order by *Anxiety/Co-Sleeping* (0.51), *Waking/Parasomnia* (0.45), and *Daytime Alertness* (0.28). Examining item factor loadings and I-ECV indicated that there were 12 items associated with the general factor more than their subscale factor (i.e., highest loading on the general factor and I-ECV > 0.50). Most of these items were from the *Initiation/Duration* (five items) and *Waking/Parasomnia* (four items) subscales, although some of these items were also from the *Anxiety/Co-Sleep* (two items) and *Daytime Alertness* (one item) subscales.

Omega Hierarchical-Subscale (the proportion of variance that is specifically accounted for by individual differences in the subscale after accounting for differences due to the general factor) was 0.18, 0.55, 0.30, and 0.71 for *Initiation/Duration*, *Anxiety/Co-Sleeping*, *Waking/Parasomnia*, and *Daytime Alertness,* respectively. The ratio of Omega Hierarchical-Subscale to Omega Total-Subscale indicated that reliable subscale variance independent of the general factor was 20% for *Initiation/Duration*, 60% for *Anxiety/Co-Sleeping*, 39% for *Waking/Parasomnia*, and 83% for *Daytime Alertness*. This pattern of results illustrates that, of all subscales, *Daytime Alertness* represents the greatest amount of reliable variance that is independent of the general factor. I-ECV indicated that, of the six items comprising the *Daytime Alertness* factor, five had more variance associated with this subscale factor than the general factor. Much of the reliable variance for the other three subscales is accounted for by the general factor: 80% for *Initiation/Duration*, 40% for *Anxiety/Co-Sleeping,* and 61% for *Waking/Parasomnia*. This suggests that these subscales are not well defined beyond the general factor. Moreover, many items loaded highest on these subscales (items with I-

ECV < 0.5 in Table 3), meaning that these items largely do not appear to relate to any reliable factor.

Although the bifactor model for the CSHQ-TD item set did not fit the data quite as well as that for the CSHQ-ASD item set, it is notable that it also indicated a similar pattern of results (Table S4); notably the general factor also accounted for the majority of reliable variance.

Measurement Invariance

Measurement invariance tests were conducted to examine if CSHQ items represented factors equivalently across the ASD and TD groups. Because the model that best fit the data spanning across the ASD and TD groups (i.e., the bifactor model for the CSHQ-ASD item set) demonstrated that a) 40–80% of reliable variance for three of four subscales was subsumed by the general factor, and b) items differed in the extent that they reflected the general factor versus their subscale factor, we took the approach of examining measurement invariance for two refined factors. In particular, we examined measurement invariance for the set of items that best represented variance in the two reliable factors; that is, we examined measurement invariance for 1) the subset of items that reflected the general factor more than their subscale factor, and 2) the subset of items that reflected the *Daytime Alertness* factor more than the general factor. This meant that we excluded items that loaded highest on unreliable factors so that they would not affect tests of measurement invariance.

The first refined factor tested for measurement was comprised of the items that each had a majority of variance associated with general factor (i.e., I-ECV > 0.5). For the general factor there were 12 such items. However, a model including all 12 of these items did not fit the data well when estimated across groups or separately for the ASD or TD group. A 10-item factor model, excluding the two items with lowest amount of variance associated with the general factor (*"Afraid to sleep alone"* and *"Awakens once during the night"*), did fit the data well (when estimated across groups CFI = 0.97 and RMSEA = 0.048 [90% *CI*: 0.032 – 0.064], when estimated for the ASD group CFI = 0.989 and RMSEA = 0.035 [90% *CI*: 0.001 – 0.063], when estimated for the typically developing group CFI = 0.970 and RMSEA = 0.032 [90% *CI*: 0.001 – 0.061]). RMSEA and CFI values were indicative of good fit for all models testing measurement invariance of the factor comprised of 10 items (Table 4). The fit of the *Weak* model (i.e., a model with item loadings constrained to be equal across groups) did not significantly differ from the *Configural* model, indicating that the strength of item loadings did not differ across groups. The *Strong* model (i.e., a model with item loadings and intercepts constrained to be equal across groups) significantly differed from the *Weak* model, meaning that mean group differences in the 10-item scale would be biased by group differences in scale properties of one or more items. Comparing CFA models with groupwise equality constraints released for single item intercepts to the fully constrained *Strong* model indicated that three items—*"Moves to someone else's bed"*, *"Struggles at bedtime"*, and *"Afraid of sleeping in dark"*—had intercepts that differed significantly between the ASD and TD groups. A partially invariant *Strong* model that did not constrain the intercepts of these three items to be equal across groups did not significantly differ from the *Weak* model. A partially invariant *Strict* model (i.e., a model with loadings equal across

groups for all items, and intercepts and residual variance equal across groups for all items except *Moves to someone else's bed"*, *"Struggles at bedtime"*, and *"Afraid of sleeping in dark"*) did not significantly differ from the partially invariant *Strong* model. A supplementary set of tests for the seven-item single factor model that excluded the items *"Moves to someone else's bed"*, *"Struggles at bedtime"*, and *"Afraid of sleeping in dark"* indicated no violation of *Configural*, *Weak*, *Strong*, or *Strict* measurement invariance (Table S7). Thus, of the 12 items that best reflected the general factor from the bifactor model for the CSHQ-ASD item set, a factor with 7 of these items was measurement invariant across the ASD and TD group.

The second refined factor tested for measurement was comprised of the items that each had a majority of variance associated with *Daytime Alertness* (i.e., I-ECV < 0.5). Models for the factor comprised of all five items with a majority of variance associated with the *Daytime Alertness* subscale did not fit the data well for data spanning across groups or for the ASD and TD groups separately. Models for a factor excluding two of these items (those with the least variance associated with *Daytime Alertness*) did fit well, for data spanning across groups (CFI = 1.00, RMSEA = 0.000 [90% *CI*: 0.000 – 0.000]), and for the ASD (CFI = 1.00, RMSEA = 0.000 [90% *CI*: 0.000 -).000]) and TD (CFI = 1.00, RMSEA = 0.000 [90% *CI*: 0.000 – 0.000]) groups separately. Measurement invariance models of the 3-item factor also fit the data well with no indication of violations in measurement invariance (Table S8). Indeed, for all models (including models reported above for data spanning across groups and the models specified for each group separately) indices were suggestive of a perfect fit. This is likely a result of a high correlation (0.89) between two of the items, "*Adults or siblings wake up child*" and *"Wakes up by him/herself"* (later being reverse scored), that seem to be simply alternative ways of wording the same concept.

In summary, of the tested models for data spanning across the TD and ASD groups, the bifactor model for the CSHQ-ASD item set represented the data best. There were two well defined factors represented in this model: the general factor (that subsumed much of the reliable variance for the *Initiation/Duration, Waking/Parasomnias,* and *Anxiety/Co-Sleeping* subscales), and the *Daytime Alertness* factor; however, items differed in the extent that they reflected these factors. There were 12 items that reflected the general factor more than their subscale factor (five from the *Initiation/Duration* subscale, four from the *Waking/Parasomnia* subscales, two from the *Anxiety/Co-Sleep* subscale, one from the *Daytime Alertness* subscale). A factor comprised of seven of these items fit the data well and demonstrated measurement invariance across the ASD and TD groups. For *Daytime Alertness*, there were five items that reflected this factor more than the general factor. A factor comprised of three of these items fit the data well and demonstrated measurement invariance across the ASD and TD groups; however, it is notable that all model fits were almost perfect and possibly due to a high correlation between two items that seem to be different ways of wording the same content rather than relating to slightly differing aspects of the same construct.

DISCUSSION

This is the first study to examine the factor structure of the CSHQ for young children that included both those with typical development alongside those with ASD, with females well

represented in the ASD group as well as the typically developing group. We examined factor structures for two different sets of CSHQ items: One of 24 items indicated from previous analysis of CSHQ data for typically developing 2- to 5-year-olds (CSHQ-TD) (Sneddon et al., 2013), and the other of 23 items for 4–10 year-olds with ASD (CSHQ-ASD) (Katz et al., 2018).

In this study, we found that, in children with ASD or typical development, bifactor models of the CSHQ were a better fit of the data than models that only included subscale factors, regardless of the item set. No previous study has examined the bifactor structure of the CSHQ specifically. However, the finding of bifactor models fitting better than other models is consistent with factor analytic investigations of other sleep questionnaires in other samples, including children and adolescents with ADHD (Mancini et al., 2019), adults (Li et al., 2019), and patients with Parkinson's disease (Pushpanathan et al., 2018). Moreover, we found that the general factor of the bifactor model accounted for the majority of reliable variance for both CSHQ item sets, which suggests that caregiver responses to the CSHQ items reflect impressions of children's sleep as a single factor that subsumes the reliable variance of subscale factors. This is consistent with results from efforts aimed at developing behavior rating scales, which have found that reliable measures of behavioral traits for young children are more difficult to obtain for more specific constructs than for broader ones (Gartstein & Rothbart, 2003; Rothbart & Bates, 2007).

Although the general factor accounted for the majority of reliable variance for both CSHQ item sets, it could not be considered as representing a single unidimensional construct. Instead of all items loading equally onto the general factor, items differed across subscales in the extent to which they reflected the general factor versus their subscale factor. For the best fitting bifactor model, there were 12 (out of 23) items that had more variance attributable to the general factor than to the subscale factor, many of which were from the *Initiation/ Duration* subscale, but some were from other subscales as well. Testing measurement invariance for a factor comprised only of the 12 items that reflected the general factor more than their subscale factors identified some items that could not be considered as indicating this new factor equivalently across the ASD and TD groups. However, a factor with seven of these items did exhibit measurement invariance across the TD and ASD groups, meaning that a scale comprised of these items (see Table 3) could be used to compare these groups without measurement bias.

Results did not provide substantial support for the reliability and measurement invariance of any other factor beyond the aforementioned factor comprised of seven items. For the best fitting bifactor model, only the *Daytime Alertness* subscale factor had a relatively large proportion of reliable variance (83%) that was independent of the general factor. Nevertheless, a measurement invariant factor reflecting this variance was comprised of 3 items covering a limited range of content. Two items, "*Adults or siblings wake up child*" and *"Wakes up by him/herself,"* which appear to simply be different ways of wording the same concept, were highly correlated (r = .89, after the later item was reverse scored). As such, it is not surprising that confirmatory factor analyses of this factor had RMSEA fit indices that were uncharacteristically low (0.00) and CFI fit indices uncharacteristically high (1.00). This suggests that the reliable variance independent of the general factor captured by the *Daytime*

Alertness factor is largely an artifact of similarity in wording content. This is in contrast to standard approaches to scale development that recommend items relate to slightly different aspects of the same domain (Irwing & Hughes, 2018).

In summary, results suggest that, using caregiver ratings of items from CSHQ, a scale comprised of the summation of seven items can be considered as forming an index that is reliable and also measurement invariant when measuring sleep problems in the context of 2- to 4-year-olds with ASD or typical development. The particular items are: *"Asleep within 20 minutes after bed", "Sleeps the same amount each day", "Sleeps too little", "Goes to bed the same time at night", "Awakes more than once during the night", "Restless and moved during sleep", and "Seems tired during the day"*. There was little evidence to suggest that any other scale could be considered to capture meaningful reliable variance. Taken together, these results have implications that are relevant to both research and the clinical understanding of sleep for young children with ASD.

In the context of research, one implication is that summation of the seven items from the new measurement invariant factor can be used to compute a sleep scale that can be used in research comparing those with ASD to TD children without concern for measurement bias. Critically this means that the scale can contribute to empirical studies seeking to understand the various underlying mechanisms (e.g., neurobiological and environmental) of sleep problems in children with ASD compared to those typically developing. Another implication is that results suggest that a level of caution should be taken when interpreting group differences between young children with ASD and those with typical development on other scales comprised of CSHQ items. In this study, a CSHQ general factor could not be considered unidimensional. This makes interpretation of group differences based on a "total score" with a scale derived by the summation of all CSHQ items (e.g., Reynolds et al., 2019) challenging; it is unclear if such group differences should be interpreted as reflecting sleep problems at a general level or sleep problems at the level of a more specific construct. Evidence that reliable variance from subscale factors is largely subsumed by a general factor means that many subscale composites of CSHQ items are also not necessarily suitable for examining specific sleep problems of young children with ASD or typical development. Notably, although there has been interest in using CSHQ subscales to examine whether young children with ASD or typical development differ with respect to specific sleep problems (Levin & Scher, 2016), the results of the present investigation suggest that interpreting these differences is problematic because subscales are not wholly independent of a common general factor. In addition, whether group differences are examined using a total score or subscale, results of the present study highlight the complications with interpretation of group differences on CSHQ subscales because scores cannot always be considered measurement invariant across groups (Kline, 2016).

Regarding clinical practice, since establishing evidence for a measure's reliability and measurement invariance is a necessary step before developing standardized scores (Irwing & Hughes, 2018), this study informs a new abbreviated scale that may be used in clinical settings, particularly when concern relates to understanding whether the sleep problems of young children with ASD are elevated beyond those experienced by typically developing peers. Future research that includes samples of TD and ASD children stratified for clinically

diagnosed sleep problems will allow development of standardized scores for this abbreviated scale that can be interpreted to indicate severity in the context of clinical practice.

The present study is not without limitations. First, the sample size precluded the opportunity to examine if CSHQ factors were measurement invariant across age ranges more specific than 2- to 4-years of age. Although the sample age range of 2- to 4-years is narrower than that of previous studies investigating the CSHQ factor structure (Katz et al., 2018; Sneddon et al., 2013), it still covers a period of development marked by significant changes in sleep-wake patterns (Galland et al., 2012). Considering these developmental changes, it will be important for future studies to examine whether the CSHQ items measure the same factor in the same way for children with ASD and those with typical development as children grow older, using data collected for narrower age ranges. Such investigations will be critical for determining whether CSHQ items are appropriate for examining developmental changes in sleep problems and could help address a critical need to understand differences in developmental trajectories of sleep problems among children with ASD compared to those with typical development (Sivertsen et al., 2012). A second limitation is that the study did not include CSHQ items pertaining to breathing difficulties, despite the possibility that breathing difficulties might affect the sleep of some children with ASD (Malow et al., 2012). The items relating to breathing difficulty were not included in the present study as we were extending on the factor structure for two CSHQ item sets derived in previous studies that also excluded these items. Breathing difficulty items were not included in these two previous studies because they did not load onto any specific factor, or respondents endorsed primarily one response option meaning items were uninformative (Katz et al., 2018; Sneddon et al., 2013). Consistent with this, in our data, responses for two of the three CSHQ items relating to breathing difficulty were also primarily limited to one option (for both items over 97% of respondents endorsed one option; *"Rarely"*). Although a recent exploratory factor analysis of CSHQ data including breathing difficulty items from a sample of 4–5 year-olds with ASD found that these items loaded onto a separate factor (Zaidman-Zait et al., 2020), it is notable that two of the items were similarly limited and therefore less informative. Overall, and consistent with findings from previous research (Certal et al., 2012; De Luca Canto et al., 2014), this suggests that, while the CSHQ sleep-related breathing problem items can be meaningful in clinical settings, in their current format they may not adequately capture breathing problems in the context of research with samples for whom this is not often a primary concern. A third limitation is that, while this study provided evidence for a factor comprised of CSHQ items that is a reliable measure in the context of young children with ASD or TD, further research is required to examine the validity of this measure as it relates to other sleep-related criterion. A final limitation of the present study arises in the context of empirical evidence of the heterogeneity in ASD across multiple domains including psychiatric comorbidities (e.g., anxiety, attention-deficit/hyperactivity disorder), cognitive and language abilities, and adaptive functioning (Cohen et al., 2014; Lord & Bishop, 2015). While the present study sought to examine the factor structure of the CSHQ across the whole sample of children with ASD, questions concerning heterogeneity within ASD as related to sleep problems remain open for further investigation (Cohen et al., 2014).

In conclusion, the present study makes several novel contributions toward improving the measurement and assessment of sleep problems experienced by young children with ASD. A

single general factor accounted for the majority of reliable variance in CSHQ items. A factor comprised of seven items that best reflected the general factor (the content of which pertains largely to difficulty settling to sleep at bedtime, nighttime awakenings, restlessness during sleep, and daytime tiredness) was demonstrated to be measurement invariant across the ASD and TD groups. Thus, this provides evidence that the sum of these items can be used in future research to reliably compare the sleep problems of children with ASD to those with typical development. In addition, normalized scores can now be developed for a scale composed of summing these items in order to use this measure in clinical practice. The results of this study suggest that specific sleep problems are not well addressed by CSHQ subscales and that there is an opportunity for further research to develop and utilize questionnaire items to assess specific sleep problems in young children.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding was provided by the National Institute of Mental Health (R01 MH068398 [SO] and K99 MH092431 [AJS] for the MIND Infant-Sibling study; R00 MH092431 [AJS] for the Purdue Infant-Sibling study; R01MH103284 [MS] and R01MH103371 [DGA] for the Autism Phenome Project [APP]; R01MH104438 [CWN] for the Girls with Autism - Imaging and Neurodevelopment [GAIN] study). The MIND Infant-Sibling study, APP and GAIN study were also supported by the MIND Institute Intellectual and Developmental Disabilities Research Center (U54HD079125) and the Autism Center of Excellence (P50 HD093079) awarded by the National Institute of Child Health and Development (NICHD). BH was supported by the MIND Institute Autism Research Training Program (T32MH073124)

We gratefully acknowledge the families who have participated.

REFERENCES

Bentler PM (2009). Alpha, Dimension-Free, and Model-Based Internal Consistency Reliability. Psychometrika, 74(1), 137–143. 10.1007/s11336-008-9100-1 [PubMed: 20161430]

Buuren, van S, & Groothuis-Oudshoorn K (2011). mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3). 10.18637/jss.v045.i03

Certal V, Catumbela E, Winck JC, Azevedo I, Teixeira-Pinto A, & Costa-Pereira A (2012). Clinical assessment of pediatric obstructive sleep apnea: A systematic review and meta-analysis. The Laryngoscope, 122(9), 2105–2114. 10.1002/lary.23465 [PubMed: 22886768]

Cheung GW, & Rensvold RB (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. Structural Equation Modeling: A Multidisciplinary Journal, 9(2), 233–255. 10.1207/S15328007SEM0902_5

Cho JR, Treweek JB, Robinson JE, Xiao C, Bremner LR, Greenbaum A, & Gradinaru V (2017). Dorsal raphe dopamine neurons modulate arousal and promote wakefulness by salient stimuli. Neuron, 94(6), 1205–1219.e8. 10.1016/j.neuron.2017.05.020 [PubMed: 28602690]

Cohen S, Conduit R, Lockley SW, Rajaratnam SM, & Cornish KM (2014). The relationship between sleep and behavior in autism spectrum disorder (ASD): A review. Journal of Neurodevelopmental Disorders, 6(1), 44. 10.1186/1866-1955-6-44 [PubMed: 25530819]

Couturier JL, Speechley KN, Steele M, Norman R, Stringer B, & Nicolson R (2005). Parental perception of sleep problems in children of normal intelligence with pervasive developmental disorders: Prevalence, severity, and pattern. Journal of the American Academy of Child and Adolescent Psychiatry, 44(8), 815–822. 10.1097/01.chi.0000166377.22651.87 [PubMed: 16034284]

De Luca Canto G, Singh V, Major MP, Witmans M, El-Hakim H, Major PW, & Flores-Mir C (2014). Diagnostic capability of questionnaires and clinical examinations to assess sleep-disordered

breathing in children. The Journal of the American Dental Association, 145(2), 165–178. 10.14219/jada.2013.26 [PubMed: 24487608]

Díaz-Román A, Zhang J, Delorme R, Beggiato A, & Cortese S (2018). Sleep in youth with autism spectrum disorders: Systematic review and meta-analysis of subjective and objective studies. Evidence Based Mental Health, 21(4), 146–154. 10.1136/ebmental-2018-300037 [PubMed: 30361331]

Galland BC, Taylor BJ, Elder DE, & Herbison P (2012). Normal sleep patterns in infants and children: A systematic review of observational studies. Sleep Medicine Reviews, 16(3), 213–222. 10.1016/j.smrv.2011.06.001 [PubMed: 21784676]

Gartstein MA, & Rothbart MK (2003). Studying infant temperament via the Revised Infant Behavior Questionnaire. Infant Behavior and Development, 26(1), 64–86. 10.1016/S0163-6383(02)00169-8

Goodlin-Jones BL, Sitnick SL, Tang K, Liu J, & Anders TF (2008). The Children's Sleep Habits Questionnaire in Toddlers and Preschool Children. Journal of Developmental & Behavioral Pediatrics, 29(2), 82–88. 10.1097/DBP.0b013e318163c39a [PubMed: 18478627]

Ho DE, Imai K, King G, & Stuart EA (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. Journal of Statistical Software, 42(8). 10.18637/jss.v042.i08

Hodge D, Parnell AMN, Hoffman CD, & Sweeney DP (2012). Methods for assessing sleep in children with autism spectrum disorders: A review. Research in Autism Spectrum Disorders, 6(4), 1337–1344. 10.1016/j.rasd.2012.05.009

Hoyle RH (Ed.). (2012). Handbook of structural equation modeling. Guilford Press.

Irwing P, & Hughes DJ (2018). Test Development. In Irwing P, Booth T, & Hughes DJ (Eds.), The Wiley Handbook of Psychometric Testing (pp. 1–47). John Wiley & Sons, Ltd. 10.1002/9781118489772.ch1

Jorgensen TD, Pornprasertmanit S, Schoemann AM, & Rosseel Y (2020). semTools: Useful tools for structural equation modeling. (R package version 0.5–2.918) [Computer software].

Katz T, Shui AM, Johnson CR, Richdale AL, Reynolds AM, Scahill L, & Malow BA (2018). Modification of the Children's Sleep Habits Questionnaire for Children with Autism Spectrum Disorder. Journal of Autism and Developmental Disorders. 10.1007/s10803-018-3520-2

Kline RB (2016). Principles and practice of structural equation modeling (Fourth edition). The Guilford Press.

Lazarus M, Chen J-F, Urade Y, & Huang Z-L (2013). Role of the basal ganglia in the control of sleep and wakefulness. Current Opinion in Neurobiology, 23(5), 780–785. 10.1016/j.conb.2013.02.001 [PubMed: 23465424]

Levin A, & Scher A (2016). Sleep Problems in Young Children with Autism Spectrum Disorders: A Study of Parenting Stress, Mothers' Sleep-Related Cognitions, and Bedtime Behaviors. CNS Neuroscience & Therapeutics, 22(11), 921–927. 10.1111/cns.12651 [PubMed: 27753226]

Li S, Fong DYT, Wong JYH, Wilkinson K, Shapiro C, Choi EPH, McPherson B, Lam CLK, & Ip MSM (2019). Nonrestorative sleep scale: Reliable and valid for the Chinese population. Quality of Life Research, 28(6), 1685–1692. 10.1007/s11136-019-02134-8 [PubMed: 30767089]

Lord C, Rutter M, DiLavore PC, Gotham K, & Bishop S (2012). Autism diagnostic observation schedule (2nd ed.). western Psychological Services.

Lord Catherine, & Bishop SL. (2015). Recent Advances in Autism Research as Reflected in DSM-5 Criteria for Autism Spectrum Disorder. Annual Review of Clinical Psychology, 11(1), 53–70. 10.1146/annurev-clinpsy-032814-112745

Lord Catherine, Risi S, Lambrecht L, Cook EH Jr., Leventhal BL, DiLavore PC, Pickles A, & Rutter M (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. Journal of Autism and Developmental Disorders, 30(3), 205–223. 10.1023/A:1005592401947 [PubMed: 11055457]

Malow BA, Byars K, Johnson K, Weiss S, Bernal P, Goldman SE, Panzer R, Coury DL, & Glaze DG (2012). A Practice Pathway for the Identification, Evaluation, and Management of Insomnia in Children and Adolescents With Autism Spectrum Disorders. Pediatrics, 130(Supplement 2), S106–S124. 10.1542/peds.2012-0900I [PubMed: 23118242]

Mancini VO, Rudaizky D, Pearcy BTD, Marriner A, Pestell CF, Gomez R, Bucks RS, & Chen W (2019). Factor Structure of the Sleep Disturbance Scale for Children (SDSC) in those with

Attention Deficit and Hyperactivity Disorder (ADHD). Sleep Medicine: X, 100006. 10.1016/j.sleepx.2019.100006 [PubMed: 33870165]

Martin CA, Papadopoulos N, Chellew T, Rinehart NJ, & Sciberras E (2019). Associations between parenting stress, parent mental health and child sleep problems for children with ADHD and ASD: Systematic review. Research in Developmental Disabilities, 93, 103463. 10.1016/j.ridd.2019.103463 [PubMed: 31446370]

Meredith W (1993). Measurement invariance, factor analysis and factorial invariance. Psychometrika, 58(4), 525–543. 10.1007/BF02294825

Merkle E, & Dongjun Y (2018). nonnest2: Tests of Non-Nested Models (R package version 0.5–2) [Computer software]. https://CRAN.R-project.org/package=nonnest2

Mitra A, Snyder AZ, Hacker CD, Pahwa M, Tagliazucchi E, Laufs H, Leuthardt EC, & Raichle ME (2016). Human cortical–hippocampal dialogue in wake and slow-wave sleep. Proceedings of the National Academy of Sciences, 113(44), E6868–E6876. 10.1073/pnas.1607289113

Mizrahi-Kliger AD, Kaplan A, Israel Z, & Bergman H (2018). Desynchronization of slow oscillations in the basal ganglia during natural sleep. Proceedings of the National Academy of Sciences of the United States of America, 115(18), E4274–E4283. 10.1073/pnas.1720795115 [PubMed: 29666271]

Mullen EM (1995). Mullen Scales of Early Learning, AGS Edition: Manual and Item Administrative Books. American Guidance Services, Inc.

Owens JA, Spirito A, & McGuinn M (2000). The children's sleep habits questionnaire (CSHQ): Psychometric properties of a survey instrument for school-aged children. Sleep, 23(8), 1043–1051. [PubMed: 11145319]

Pushpanathan ME, Loftus AM, Gasson N, Thomas MG, Timms CF, Olaithe M, & Bucks RS (2018). Beyond factor analysis: Multidimensionality and the Parkinson's Disease Sleep Scale-Revised. PLOS ONE, 13(2), e0192394. 10.1371/journal.pone.0192394 [PubMed: 29432463]

Quach J, Hiscock H, Ukoumunne OC, & Wake M (2011). A Brief Sleep Intervention Improves Outcomes in the School Entry Year: A Randomized Controlled Trial. PEDIATRICS, 128(4), 692–701. 10.1542/peds.2011-0409 [PubMed: 21890825]

Reynolds AM, Soke GN, Sabourin KR, Hepburn S, Katz T, Wiggins LD, Schieve LA, & Levy SE (2019). Sleep Problems in 2- to 5-Year-Olds With Autism Spectrum Disorder and Other Developmental Delays. Pediatrics, 143(3), e20180492. 10.1542/peds.2018-0492 [PubMed: 30745433]

Richdale AL, & Schreck KA (2009). Sleep problems in autism spectrum disorders: Prevalence, nature, & possible biopsychosocial aetiologies. Sleep Medicine Reviews, 13(6), 403–411. 10.1016/j.smrv.2009.02.003 [PubMed: 19398354]

Richdale AL, & Schreck KA (2019). Examining sleep hygiene factors and sleep in young children with and without autism spectrum disorder. Research in Autism Spectrum Disorders, 57, 154–162. 10.1016/j.rasd.2018.10.008

Robinson AM, & Richdale AL (2004). Sleep problems in children with an intellectual disability: Parental perceptions of sleep problems, and views of treatment effectiveness. Child: Care, Health and Development, 30(2), 139–150. 10.1111/j.1365-2214.2004.00395.x

Rodriguez A, Reise SP, & Haviland MG (2016a). Evaluating bifactor models: Calculating and interpreting statistical indices. Psychological Methods, 21(2), 137–150. 10.1037/met0000045 [PubMed: 26523435]

Rodriguez A, Reise SP, & Haviland MG (2016b). Applying Bifactor Statistical Indices in the Evaluation of Psychological Measures. Journal of Personality Assessment, 98(3), 223–237. 10.1080/00223891.2015.1089249 [PubMed: 26514921]

Rothbart MK, & Bates JE (2007). Temperament. In Handbook of child psychology (Vol. 3, pp. 99–166). John Wiley & Sons, Inc.

Rubin DB (Ed.). (1987). Multiple Imputation for Nonresponse in Surveys. John Wiley & Sons, Inc. 10.1002/9780470316696

Russeel Y (2012). lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software, 48(2), 1–36.

Rutter M, Bailey A, & Lord C (2003). The Social Communication Questionnaire. Western Psychological Services.

Saper CB, Fuller PM, Pedersen NP, Lu J, & Scammell TE (2010). Sleep state switching. Neuron, 68(6), 1023–1042. 10.1016/j.neuron.2010.11.032 [PubMed: 21172606]

Sivertsen B, Posserud M-B, Gillberg C, Lundervold AJ, & Hysing M (2012). Sleep problems in children with autism spectrum problems: A longitudinal population-based study. Autism, 16(2), 139–150. 10.1177/1362361311404255 [PubMed: 21478225]

Sneddon P, Peacock GG, & Crowley SL (2013). Assessment of Sleep Problems in Preschool Aged Children: An Adaptation of the Children's Sleep Habits Questionnaire. Behavioral Sleep Medicine, 11(4), 283–296. 10.1080/15402002.2012.707158 [PubMed: 23394069]

Souders MC, Mason TBA, Valladares O, Bucan M, Levy SE, Mandell DS, Weaver TE, & Pinto-Martin J (2009). Sleep behaviors and sleep quality in children with autism spectrum disorders. Sleep, 32(12), 1566–1578. 10.1093/sleep/32.12.1566 [PubMed: 20041592]

Timmerman ME, Lorenzo-Seva U, & Ceulemans E (2018). The Number of Factors Problem. In Irwing P, Booth T, & Hughes DJ (Eds.), The Wiley Handbook of Psychometric Testing (pp. 305–324). John Wiley & Sons, Ltd. 10.1002/9781118489772.ch11

Vuong QH (1989). Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica, 57(2), 307. 10.2307/1912557

Wang G, Li W-L, Han Y, Gao L, Dai W, Su Y, & Zhang X (2019). Sensory Processing Problems and Comorbidities in Chinese Preschool Children with Autism Spectrum Disorders. Journal of Autism and Developmental Disorders, 49(10), 4097–4108. 10.1007/s10803-019-04125-7 [PubMed: 31267289]

West SG, Taylor AB, & Wu W (2012). Model fit and model selection in structural equation modeling. In Handbook of structural equation modeling. (pp. 209–231). The Guilford Press.

Wu H, & Estabrook R (2016). Identification of Confirmatory Factor Analysis Models of Different Levels of Invariance for Ordered Categorical Outcomes. Psychometrika, 81(4), 1014–1045. 10.1007/s11336-016-9506-0 [PubMed: 27402166]

Zaidman-Zait A, Zwaigenbaum L, Duku E, Bennett T, Szatmari P, Mirenda P, Smith I, Vaillancourt T, Volden J, Waddell C, Kerns C, Elsabbagh M, Georgiades S, Ungar WJ, Fombonne E, & Roberts W (2020). Factor analysis of the children's sleep habits questionnaire among preschool children with autism spectrum disorder. Research in Developmental Disabilities, 97, 103548. 10.1016/j.ridd.2019.103548 [PubMed: 31901672]

Table 1.

Sample Characteristics

	ASD (*n* = 224)	Typically Developing (*n* = 224)	*p*-value
Study Sample			< 0.001
MIND Infant-Sibling study	11 (5%)	66 (29%)	
Purdue Infant-Sibling study	5 (2%)	40 (18%)	
GAIN study	165 (74%)	92 (41%)	
APP	43 (19%)	26 (12%)	
Sex (male)	148 (66%)	128 (57%)	*ns*
Age (months)	36.1 (5.9)	35.8 (5.7)	*ns*
Visual Reception: Mean (*SD*)[a]	31.4 (14.6)	60.1 (10.4)	< 0.001
Fine Motor: Mean (*SD*)[b]	28.8 (11.4)	51.7 (10.0)	< 0.001
Receptive Language: Mean (*SD*)[c]	30.3 (12.5)	53.1 (7.9)	< 0.001
Expressive Language: Mean (*SD*)[d]	31.1 (12.6)	55.8 (8.2)	< 0.001

p-values are for Fisher's exact test of overall independence for study sample, Chi-square test of independence for Sex, and one-way ANOVA for continuous variables.

APP Autism Phenome Project, *ASD* autism spectrum disorder, *GAIN* Girls with Autism - Imaging and Neurodevelopment, *SD* standard deviation.

ns = not significant at $p > 0.05$ level

[a]Missing $n = 17$ for ASD group

[b]Missing $n = 1$ for TD group & $n = 11$ for ASD group.

[c]Missing n=1 for TD group & n=43 for ASD group

[d]Missing n=2 for TD group & n=48 for ASD group.

Table 2

Model fit statistics for alternative CFA models of CSHQ

CSHQ Item Set	Factor Structure	χ^2 (*df*)	RMSEA (90% *CIs*)	CFI	$\Delta\chi^2$ (Δdf)
CSHQ-TD Sneddon et al. (2013)	Single General Factor	2175.96 (252)	0.131 (0.126 – 0.136)	0.836	
	Correlated Subscale Factors	1026.75 (246)	0.084 (0.079 – 0.090)	0.934	Correlated versus General: 1149.21 (6) ***
	Bifactor	499.94 (228)	0.052 (0.046 – 0.05)	0.977	Bifactor versus Correlated: 526.81 (16) ***
CSHQ-ASD Katz et al. (2018)	Single General Factor	2004.45 (230)	0.131 (0.126 – 0.137)	0.842	
	Correlated Subscale Factors	682.98 (224)	0.068 (0.062 – 0.074)	0.959	Correlated versus General: 1321.47 (6) ***
	Bifactor	400.52 (207)	0.046 (0.040 – 0.052)	0.983	Bifactor versus Correlated: 282.46 (17) ***

RMSEA root mean square error of approximation, CIs confidence intervals, CFI comparative fit index

* p<.05

** p<.01

*** p<.001

Table 3

Bifactor model standardized factor loadings for CSHQ-ASD item set

Abbreviated Item wording[a]	General Sleep	Initiation/ Duration	Anxiety/Co-sleep	Waking/ Parasomnia	Daytime Alertness	I-ECV
Asleep within 20 minutes after bed	0.54	0.14				**0.94**
Sleeps right amount	0.70	0.92				0.37
Sleeps same each day	0.51	0.37				**0.65**
Sleeps too little	0.71	0.40				**0.76**
Struggles at bedtime	0.71	0.04				0.99
Goes to bed same time at night	0.64	0.05				**0.99**
Needs parent in room to fall asleep	0.52		0.72			0.34
Afraid to sleep alone	0.56		0.55			0.51
Falls asleep alone in own bed	0.56		0.82			0.31
Falls asleep in parent/sibling bed	0.47		0.73			0.29
Moves to someone else's bed	0.44		0.29			0.70
Awakens alarmed from dream	0.38			0.52		0.34
Awakes more than once	0.61			0.53		**0.57**
Talks during sleep	0.30			0.44		0.32
Restless and moves during sleep	0.57			0.24		**0.85**
Awakes once during the night	0.52			0.49		0.53
Afraid of sleeping in dark	0.30			−0.06		0.96
Long time to become alert in morning	0.37				0.64	0.25
Difficulty getting out of bed	0.31				0.80	0.13
Wakes up in negative mood	0.43				0.45	0.48
Child seems tired	0.36				0.04	**0.99**
Adults or siblings wake up child	0.12				0.84	0.02
Wakes up by him/herself	0.09				0.90	0.01
Omega Total	0.93	0.89	0.91	0.76	0.86	
Omega H	0.73	0.18	0.55	0.30	0.71	
ECV	0.44	0.09	0.17	0.08	0.22	

ECV explained common variance, I-ECV item explained common variance

[a]Seven bolded items formed a single factor that was measurement invariant across the ASD and TD groups

Table 4.

Measurement invariance tests for 10-item factor

	χ^2 (*df*)	Δ χ^2 (Δ*df*)	RMSEA (90% *CI*s)	Δ RMSEA	CFI	Δ CFI
Baseline Model	87.90 (70)		0.034 (0.000 – 0.054)		0.985	
Metric Model	99.70 (79)	11.80 (9)	0.034 (0.000 – 0.053)	0.000	0.982	−0.003
Scaler Model	143.79 (88)	44.09 (9)***	0.053 (0.037 – 0.069)	0.019	0.952	−0.03
Scaler Model – Partial[a]	109.02 (85)	9.32 (6)	0.036 (0.008 – 0.054)	0.002	0.979	−0.001
Strict Model– Partial[b]	122.51 (92)	13.49 (7)	0.039 (0.017 – 0.056)	0.003	0.974	−0.005

Note. RMSEA = root mean square error of approximation; is = confidence intervals; CFI = comparative fit index.

[a]Groupwise intercept equality constraints relaxed for items: *"Moves to someone else's bed"*, *"Struggles at bedtime"*, and *"Afraid of sleeping in dark."*

[b]Groupwise intercept and residual variance equality constraints relaxed for items: *"Moves to someone else's bed"*, *"Struggles at bedtime"*, and *"Afraid of sleeping in dark."*