Assessing sleep problems and daytime functioning: a translation, adaption, and validation of the Athens Insomnia Scale for non-clinical application (AIS-NCA)

Sebastian Sattler, Daniel Seddig, Giulia Zerbini

Angaben zur Veröffentlichung / Publication details:

Sattler, Sebastian, Daniel Seddig, and Giulia Zerbini. 2023. "Assessing sleep problems and daytime functioning: a translation, adaption, and validation of the Athens Insomnia Scale for non-clinical application (AIS-NCA)." *Psychology & Health* 38 (8): 1006–31.
https://doi.org/10.1080/08870446.2021.1998498.

Assessing sleep problems and daytime functioning: a translation, adaption, and validation of the Athens Insomnia Scale for non-clinical application (AIS-NCA)

Sebastian Sattler[a,b,c], Daniel Seddig[a] and Giulia Zerbini[d]

[a]Institute of Sociology and Social Psychology, University of Cologne, Cologne, Germany; [b]Pragmatic Health Ethics Research Unit, Institut de recherches cliniques de Montréal, Montréal, Canada; [c]Faculty of Sociology, Bielefeld University, Bielefeld, Germany; [d]Department of Medical Psychology and Sociology, Medical Faculty, Augsburg University, Augsburg, Germany

ABSTRACT

Objective & Design: Sleep problems are common and have been linked to health problems, diminished well-being, and impaired performance. Many scales to diagnose clinically relevant sleep problems are time-consuming, complex, and difficult to administer in non-clinical and multi-thematic studies. Through a multi-stage translation (from English to German) and scale testing process, we developed a parsimonious measure of sleep problems and daytime functioning for non-clinical applications based on the Athens Insomnia Scale. **Results:** Exploratory ($N_{Study\ 1}$ = 25,140) and confirmatory ($N_{Study\ 2}$ = 14,797) factor analyses suggest a two-dimensional structure with the subscales "sleep problems" and "daytime functioning". Internal scale consistency was acceptable. Measurement invariance was found across time, gender, age, and diagnosed sleep disorders. The scale discriminates between people with and without sleep disorders and predicts emerging sleep disorders. Short-term retest reliability was acceptable ($N_{Study\ 3}$ = 78). Convergent validity with other sleep measures and discriminant validity with indicators of well-being were observed ($N_{Study\ 4}$ = 341). After a multi-stage translation to English, we confirmed the factor structure and found measurement invariance across languages ($N_{Study\ 5}$ = 623). **Conclusion:** Our short 7-item scale has good psychometric properties and is suitable for self-administration, making it useful in measuring sleep problems and daytime functioning efficiently and reliably, especially for large population studies.

Introduction

Sleep is essential for health, well-being, and performance (Grandner, 2017; Watson et al., 2015). Sleep also takes up about one third of our 24-hour day, making it one of the most important "activities" of our day. Obtaining a good and restorative night

CONTACT Sebastian Sattler sebastian.sattler@uni-bielefeld.de Bielefeld University, Cologne, Germany; Giulia Zerbini giulia.zerbini@med.uni-augsburg.de Department of Medical Psychology and Sociology, University of Augsburg, Augsburg, Germany.

Supplemental data for this article is available online at https://doi.org/10.1080/08870446.2021.1998498.

of sleep has, however, become challenging in modern society (Khubchandani & Price, 2020). Many factors contribute to sleep problems, such as stress (e.g., Kalmbach et al., 2018; Kim & Dimsdale, 2007), social demands (e.g., school and working schedules; Bannai & Tamakoshi, 2014; Crowley et al., 2018), illness (e.g., Besedovsky et al., 2019; Kaplow, 2016), and ageing (e.g., Gulia & Kumar, 2018; Li et al., 2018; Mander et al., 2017). The most common sleep disorder is insomnia, which is characterized by difficulties initiating or maintaining sleep and by impaired daytime functioning (including subjective complaints), despite appropriate circumstances for sleep. The prevalence of insomnia has dramatically increased in the past decades. For instance, between 1993 and 2015 the number of diagnosed insomniacs increased from 800,000 to 9.4 million in the United States (Moloney et al., 2019). In Europe, the prevalence of chronic insomnia is estimated to be around ten percent (Riemann et al., 2017).

Beside insomnia, there is a great variety and severity of sleep problems, ranging from non-clinical to clinical manifestations. Sleep problems have been linked to lower psychological well-being (Barros et al., 2019; Dai et al., 2020; Kuppermann et al., 1995) and to an increased risk of health complaints (e.g., obesity, cardiovascular disease; Cappuccio & Miller, 2017; Fatima et al., 2016; McDermott et al., 2018; Patel & Hu, 2008). It has also been linked to adverse health behaviors, such as substance misuse (e.g., Hasler et al., 2012; Logan et al., 2018). Finally, sleep problems have been found to be associated with lower cognitive and academic performance (e.g., Brownlow et al., 2020; Gomez Fonseca & Genzel, 2020), reduced work productivity (e.g., Barnes & Watson, 2019; Pilcher & Morris, 2020), and increased workplace accidents (Magnavita & Garbarino, 2017). Importantly, many of these aspects result in potential economic and social burdens for individuals and society. The relationship between sleep problems and these variables is often bidirectional, indicating that sleep problems can be both cause and effect.

So far, many sleep scales have been developed for clinical applications in order to screen the general population, and then diagnose and treat individuals with sleep disorders (e.g., Buysse et al., 1989; Crönlein et al., 2013; Jenkins et al., 1988; Morin et al., 2011; Pallesen et al., 2008; Soldatos et al., 2000). However, an important aspect of health promotion is prevention. In order to prevent the development of sleep disorders, it is essential to identify subclinical manifestations as well as risk and protective factors using longitudinal prospective large-population studies. There is therefore a need not only for diagnostic sleep scales (e.g., insomnia screening), but also for scales with a finer resolution capturing smaller variations in the severity of sleep problems and in the impairment of daytime functioning, ranging from subclinical to clinical manifestations. Short scales that provide data on sleep problems and daytime functioning in an economical and reliable way are especially important. They are the preferred choice for large-scale multi-topic longitudinal studies to assess potential changes of both constructs over time, and allow for cause-effect analyses between both constructs and, for example, stress, health, or substance misuse (Benham & Charak, 2019; Conroy, 2017; Lam et al., 2018). Succinct measures can have the advantage to reduce cognitive burdens for respondents and their fatigue. Moreover, short scales are also useful to screen participants for certain characteristics or when survey time is limited.

The aim of the present study was to develop and validate a scale to assess sleep problems and daytime functioning for non-clinical applications. For this purpose, we adapted the Athens Insomnia Scale (AIS; Soldatos et al., 2000; Soldatos et al., 2003), a cross-culturally validated and widely-used scale. To successfully apply the scale, we followed international guidelines for cross-cultural adaptation of health-related self-report measures (Beaton et al., 2000; Guillemin et al., 1993). The AIS items are based on the ICD-10 criteria for insomnia (e.g., difficulties in falling asleep, maintaining sleep, as well as distress induced by sleep problems and/or impaired daytime functioning). These are relevant in diagnosing insomnia and assessing general sleep problems and its immediate consequences, which we therefore elected to keep in our scale. Still, we rephrased some items and changed the response options to increase the versatility of the scale and to make it especially suitable for non-clinical applications. We chose the AIS because of its good psychometric properties, such as internal consistency, temporal validity, and criterion validity, and its wide use in many studies (e.g., Chung et al., 2011; Enomoto et al., 2018; Gómez-Benito et al., 2011; Hallit et al., 2019; Jeong et al., 2015; Lin et al., 2020; Okajima et al., 2013; Sirajudeen et al., 2020; Soldatos et al., 2000; Sun et al., 2011; Yen et al., 2010). Moreover, this questionnaire is short, simple to understand, and quick to administer, making it an interesting tool especially for large-scale studies.

In terms of sleep dimensions assessed by the AIS, Soldatos et al. (2000) identified only one factor, which was also found in further studies validating the AIS (Gómez-Benito et al., 2011; Jeong et al., 2015; Lin et al., 2020). However, more recent studies found that the items load on two or even three factors (Chung et al., 2011; Enomoto et al., 2018; Hallit et al., 2019; Okajima et al., 2013; Sirajudeen et al., 2020; Yen et al., 2010). Common factor structures identified in these studies were insomnia symptoms (items: 1-3), and subjective sleep and daytime functioning (items: 4-8) (Chung et al., 2011; Yen et al., 2010), or sleep problems (items: 1-5) and daytime functioning (items: 6-8) (Enomoto et al., 2018; Hallit et al., 2019; Okajima et al., 2013).

In our study, we found a two-factor structure of the developed Athens Insomnia Scale for Non-Clinical Application (AIS-NCA) with the subscales "sleep problems" and "daytime functioning" in an exploratory factor analysis (EFA) that was verified in a confirmatory factor analysis (CFA). Test-retest reliability, convergent, discriminant, and predictive validity, as well as measurement invariance across time, gender, age, diagnosed sleep disorders were satisfactory overall. We also translated the scale to English based on international guidelines for cross-cultural adaptations of scales (Beaton et al., 2000; Guillemin et al., 1993) and tested the dimensionality, internal consistency, as well as the measurement equivalence across languages (English and German). The final versions of the German and English scales can be found in the Supplementary Information (Figure S1).

Development of the Athens Insomnia Scale for non-clinical application (AIS-NCA)

We generally followed the guidelines for cross-cultural scale adaptation (Beaton et al., 2000; Guillemin et al., 1993) when translating the scale (Figure 1). In *stage 1*, two translators (native German speakers) naïve to the scale translated the eight original AIS items from English to German. In *stage 2*, four experts (three with methods expertise and one without, all fluent in English and German) judged and discussed

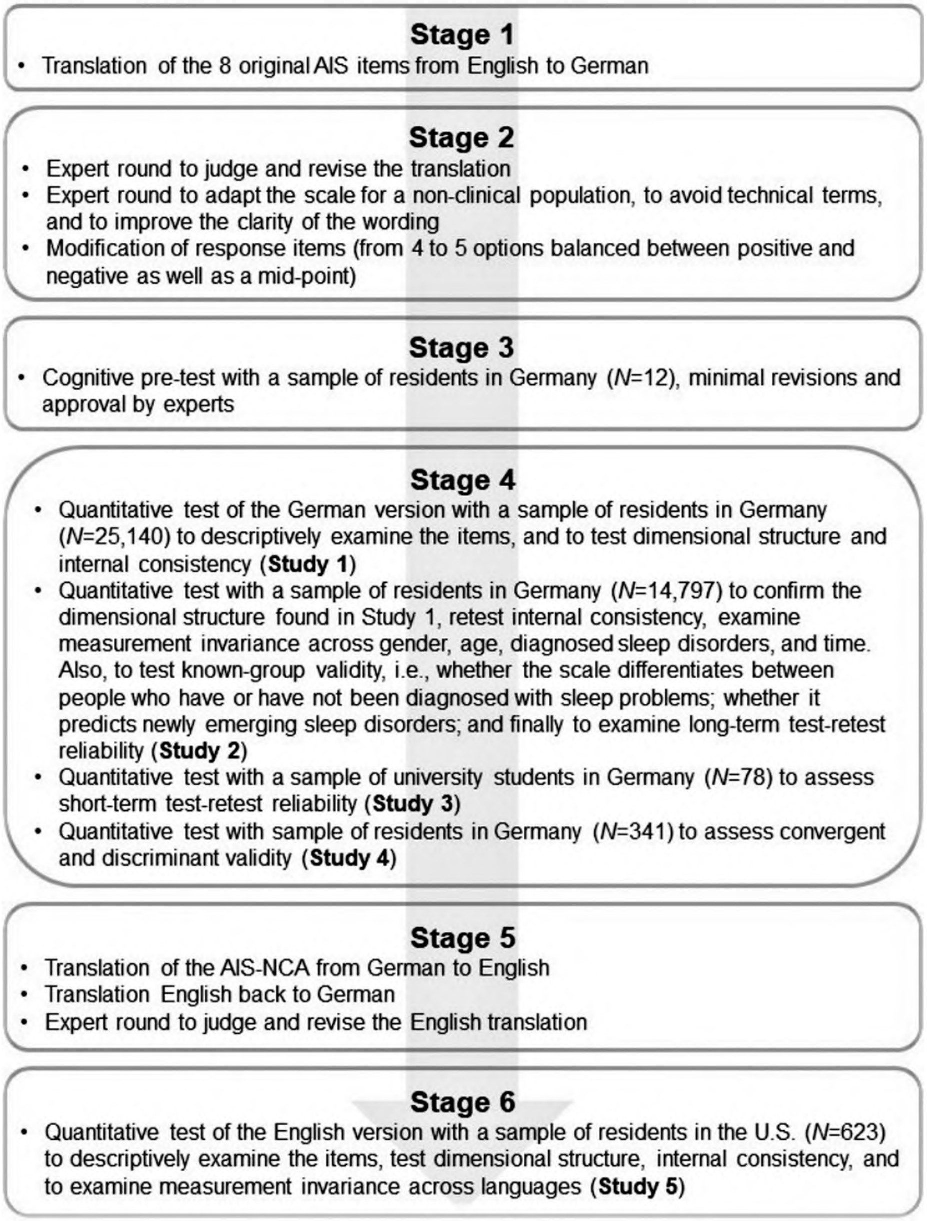

Figure 1. Description of the stages of the translation process and the validation of the scale.

the translation, which resulted in slight revisions of the translation. Then, five experts judged the translation (three with methods expertise, one with expertise in sleep research, one with non-specific expertise, all fluent in English and German) and reached a consensus with the aim of adapting the scale to a non-clinical population, avoiding technical terms, and improving the clarity of the wording. Thereby, items were also rephrased to avoid participants needing supervision while filling out the scale. For instance, the first item: “Sleep induction (time it takes you to fall asleep after turning off the lights" was changed to: "I could usually get to sleep (after turning off the lights)…". The most significant changes made were regarding the answer options, partly because the AIS was developed for clinical applications (i.e., screening and diagnosis of insomnia). We extended the response options from four to five categories to allow for a more detailed assessment and to provide a logical mid-point for most of the items. For instance, we changed the answer options for the item assessing sleep quality from: "satisfactory," "slightly unsatisfactory," "markedly unsatisfactory," "very unsatisfactory or did not sleep at all" to "very good," "good," "sometimes good/sometimes bad," "bad," and "very bad." Moreover, the time frame of the assessment was extended to twelve months (since the first application was planned for an annual panel study). We, however, also tested versions with shorter observation periods (i.e., four weeks; Studies 3 to 5).

In *stage 3*, cognitive pretests were conducted with a diverse sample of adult residents ($N = 12$) in Germany. We used the think-aloud technique and probing questions (Van Someren et al., 1994) to evaluate the comprehensibility and content validity of the instruction, items, and response categories (for example, by asking how respondents understood "finally woke up earlier than desired"). While no major difficulties and good comprehensibility were observed, we changed the phrase "after sleeping" to "during daytime" in the three daytime functioning items. Moreover, item eight and the response options were changed from "After sleeping, I was mostly…" ("very awake," "awake," "partly awake/partly tired," "tired," "very tired") to "My tiredness was during the day mostly…" ("not perceptible," "hardly perceptible," "moderately perceptible," "very perceptible," "very strongly perceptible"). These last two changes were made to ensure that participants would think about their daytime functioning across the entire day and not only in the morning after waking up (i.e., after sleeping). Based on these three stages, we continued with quantitative tests of this scale (Table 1 for the wording) and its properties, described in the following four studies.

Study 1

This first study aimed to test the extent to which the new items of the German version of the AIS-NCA scale reflect the underlying construct(s) (i.e., sleep problems and daytime functioning) that the scale is supposed to measure. In the original paper, Soldatos et al. (2000) found a one-dimensional structure, which was also found in other studies (Gómez-Benito et al., 2011; Jeong et al., 2015; Lin et al., 2020). Other studies reported two- or even three-factors structure of the AIS (Chung et al., 2011; Enomoto et al., 2018; Hallit et al., 2019; Okajima et al., 2013; Sirajudeen et al., 2020; Yen et al., 2010).

Methods

Participants: The initial sample for this web-based study was recruited offline in a multi-stage random process (based on the ADM telephone master sample[1] (ADM, 2012; Gabler & Häder, 1998)) via the forsa.omninet panel. It was nationally representative with regard to sex, age, education, and province for adult (18 or older) residents in Germany

Table 1. Items and descriptives of the AIS-NCA (Study 1, $N = 25{,}140$).[a]

#	Item and response options	Descriptives M	SD	Min	Max	Skew	Kurt
1	I could usually get to sleep (after turning off the lights)… *immediately; after a very short time; after a short time; after a long time; after a very long time* [Einschlafen konnte ich meist (nach dem Ausschalten des Lichts)… *sofort; nach sehr kurzer Zeit; nach kurzer Zeit; nach längerer Zeit; nach sehr langer Zeit*]	2.66	0.96	1.00	5.00	0.21	−0.45
2	Waking up during my sleep happened… *never; almost never; sometimes; quite often; very often* [Dass ich während des Schlafens aufwache, passiert… *nie; fast nie; manchmal; ziemlich oft; sehr oft*]	3.37	0.94	1.00	5.00	−0.01	−0.47
3	Waking up prematurely happened… *never; almost never; sometimes; quite often; very often* [Endgültig früher aufgewacht als gewünscht bin ich… *nie; fast nie; manchmal; ziemlich oft; sehr oft*]	2.91	0.96	1.00	5.00	0.24	−0.30
4	The overall duration of my sleep was usually… *more than sufficient; sufficient; partly sufficient/partly insufficient; insufficient; very insufficient* [Meine Schlafdauer war insgesamt meist… *mehr als ausreichend; ausreichend; teils ausreichend/teils unzureichend; unzureichend; stark unzureichend*]	2.68	0.84	1.00	5.00	0.59	−0.18
5	The overall quality of my sleep was usually… *very good; good; sometimes good/sometimes bad; bad; very bad* [Meine Schlafqualität war insgesamt meist… *sehr gut; gut; teils gut/teils schlecht; schlecht; sehr schlecht*]	2.57	0.86	1.00	5.00	0.49	0.14
6	Throughout the day, my level of well-being was usually… *very good; good; sometimes good/sometimes bad; bad; very bad* [Mein Wohlbefinden war tagsüber meist… *sehr gut; gut; teils gut/teils schlecht; schlecht; sehr schlecht*]	2.36	0.69	1.00	5.00	0.54	0.63
7	Throughout the day, my level of (physical and mental) performance was usually… *very good; good; sometimes good/sometimes bad; bad; very bad* [Meine Leistungsfähigkeit (physisch und psychisch) war tagsüber meist… *sehr gut; gut; teils gut/teils schlecht; schlecht; sehr schlecht]*	2.33	0.70	1.00	5.00	0.65	0.93
8	Throughout the day, my level of tiredness was usually… *not perceptible; hardly perceptible; moderately perceptible; very perceptible; very strongly perceptible* [Meine Müdigkeit war tagsüber meist… *nicht spürbar; kaum spürbar; mäßig spürbar; stark spürbar; sehr stark spürbar*]	2.57	0.82	1.00	5.00	0.13	0.04
	AIS-NCA: Sleep problems	2.88	0.68	1.00	5.00	0.28	−0.11
	AIS-NCA: Daytime functioning	2.42	0.63	1.00	5.00	0.51	0.62
	AIS-NCA: Total score	2.68	0.58	1.00	5.00	0.37	0.12

Notes: [a] This table shows the AIS-NCA in English and German (please note that Study 1 was conducted with a German-speaking sample and that the development of the English version was part of Study 5; see Figure S1 for the final versions of the German and English scales). The instruction was: "How do you rate the following aspects related to your sleep? Please base your answers on the past twelve months." [Wie bewerten Sie die folgenden Dinge, die mit Ihrem Schlaf zusammenhängen? Beziehen Sie sich bitte auf die letzten 12 Monate.]. M = Mean, SD = Standard deviation; Min = Minimum, Max = Maximum; Skew = Skewness, Kurt = Excess kurtosis.

(who have internet access, which applies to about 95% of all households; Statista, 2020). The gross sample consisted of 47,406 invited individuals. Of these individuals, 27,149 (57.27%) consented to participate, and 24,809 (91.38%) completed the study.[2] Our analysis comprises of answers of 25,140 individuals (45.69% women, mean age: 53.92 ± 15.00) with complete answers on all relevant variables. Upon completion, participants received bonus points as incentives (2€, approximately \$2.30, which was convertible to vouchers, a charity lottery ticket, or a donation to UNICEF). The ethics committee of the University of Erfurt approved the study (reference number: EV-20190917).

Instruments: The first five items (Table 1) of the self-administered AIS-NCA aim to measure subjectively perceived sleep problems (including difficulty with sleep induction, waking up during the night or the early morning, total sleep time, and overall sleep quality), while the remaining three items assess experienced consequences of sleep problems during the day (problems with well-being, physical and mental functioning, and tiredness). Respondents were asked to evaluate each item based on the last 12 months on a five-point scale, in which higher scores indicate more serious problems (Table 1). Respondents could also choose "does not apply" as a response option.

Statistical analysis: Descriptive information about the distribution of the answers are provided (i.e., mean values, standard deviations, minimum, maximum, skewness, and excess kurtosis[3]). To investigate the dimensionality of the scale, we used exploratory factor analysis (EFA) with oblimin rotation. We used full information maximum likelihood estimation and we treated our data as continuous normal. A factor was retained when the eigenvalue was larger than 1.00, factor determinacy coefficients were close to 1.00, and when the factor was substantively interpretable. Moreover, we retained an item when it had a factor loading larger than 0.30. We deleted an item when it loaded on more than one factor with a loading larger than 0.30 and when the difference of a cross loading to the main loading was smaller than 0.15 (see Brown, 2015; Worthington & Whittaker, 2006). Moreover, internal consistency was examined with *McDonald's* ω.

Results and discussion

Table 1 shows the mean, standard deviation, minimum, maximum, skewness, and excess kurtosis for all items. When testing for the normality of the data, we found that most scores were in acceptable limits (±2 as defined by George & Mallery, 2010) regarding $skewness_{range}$ = -0.01 to 0.65 and $excess\ kurtosis_{range}$ = -0.47 to 0.93, i.e., most items were fairly symmetrical (few had minimal more right-handed tails), while distributions of the first four items tended to be slightly platykurtic (i.e., thinner, heavier tails), and the latter three tended towards being minimally leptokurtic distributions (i.e., with heavier tails). However, for the EFA, we used a maximum likelihood estimator that is robust against non-normality.

The EFA showed two factors according to the eigenvalue criterion (Factor 1 = 3.808, Factor 2 = 1.138). Factor determinacy coefficients for both factors were close to 1.000, indicating that factor score estimates were closely related to the latent factor scores. The rotated standardized factor loadings are shown in Table 2. Items 1 to 3 and item 5 loaded on the first factor that we interpreted as "sleep problems" and items 6 to 8 loaded on another factor that we interpreted as "daytime functioning". Item 4 had a considerable cross loading on both factors, possibly due to being too unspecific with regard to both sleep problems and daytime functioning. We therefore omitted item 4 and performed a second EFA that revealed two factors ($Eigenvalue_{Factor\ 1}$ = 3.363, $Eigenvalue_{Factor\ 2}$ = 1.133), but this time without considerable cross loadings. Factor determinacy coefficients for both factors were again close to 1.000. Both factors correlated with $r = 0.573$. Moreover, we found an acceptable-to-good internal consistency of both subscales ($\omega_{Sleep\ problems}$ = 0.730; $\omega_{Daytime\ functioning}$ = 0.829) as well as the

Table 2. Results from exploratory factor analysis (with oblimin rotation) of the AIS-NCA (Study 1, $N = 25{,}140$).

#	Wording	Factor loadings with 8 items		Factor loadings with 7 items	
		Factor 1: Sleep problems	Factor 2: Impaired daytime functioning	Factor 1: Sleep problems	Factor 2: Impaired daytime functioning
1	I could usually get to sleep (after turning off the lights)…	**0.390**	0.092	**0.343**	0.151
2	Waking up during my sleep happened…	**0.708**	−0.112	**0.747**	−0.083
3	Waking up prematurely happened…	**0.621**	−0.105	**0.639**	−0.071
4	The overall duration of my sleep was usually	**0.410**	**0.319**	–	–
5	The overall quality of my sleep was usually…	**0.773**	0.146	**0.639**	0.272
6	Throughout the day, my level of well-being was usually…	0.052	**0.837**	0.033	**0.856**
7	Throughout the day, my level of (physical and mental) performance was usually…	−0.054	**0.842**	−0.041	**0.833**
8	Throughout the day, my level of tiredness was usually…	0.073	**0.627**	0.055	**0.631**
	Factor determinacy	0.914	0.932	0.886	0.932

Notes: Bold numbers refer to factor loadings of items on their respective target factor.

entire scale ($\omega_{AIS\text{-}NCA}$=0.815) after calculating composite scores. Altogether, the findings from Study 1 suggested two subscales composed of 4 items assessing sleep problems and 3 items assessing daytime functioning. The moderate correlation between the two subscales indicated that the two factors might be causally related (potentially with sleep problems negatively influencing daytime functioning), but were still separate constructs that are both relevant when assessing sleep characteristics.

Study 2

The aim of this study was to replicate the findings of Study 1 with a confirmatory factor analysis (CFA) of the dimensionality of both subscales and total scale, and to test internal consistency with *McDonald's ω*. We also exemplarily assessed the invariance of the measures across gender, age, people with and without a sleep disorder diagnosis, and time. The first two comparisons were conducted because of previously reported gender and age differences in sleep properties (Krishnan & Collop, 2006; Mander et al., 2017), while the last two comparisons, time and diagnosis of sleep disorders, might be interesting for longitudinal and clinical studies, respectively. Moreover, we wanted to explore known-group validity by examining whether the scale is able to differentiate participants with and without a sleep disorder diagnosis (e.g., insomnia). We expected the scale scores in the clinical group to be significantly higher compared to the non-clinical group. To examine predictive validity, we tested whether the scale predicts such a diagnosis one year later. We also tested the stability of the scale over one year.

Methods

Participants: For Study 2, 24,683 respondents who completed Study 1 and did not drop out of the panel were re-invited twelve months later. Of these participants, 17,818 (72.19%) individuals consented to participate, and 15,235 (85.50%) completed the study. The analytic sample comprises 14,797 individuals (50.42% women, mean age: 52.64 ± 14.30) with complete answers on all relevant variables.[4] Completers received an incentive worth 2.5€(approximately $3.00). The ethics committee of the University of Erfurt approved the study (reference number: EV-20200805).

Instruments: AIS-NCA was measured as in Study 1 (without item 4 due to the factor loading in Study 1, leading to a new numbering of items from 1 to 7) and mean-weighted total scores of the scale and both subscales were computed (for the descriptives see Table S1).

Diagnosis of sleep problems: Respondents were asked whether they were diagnosed with narcolepsy/cataplexy or other sleep disturbances (e.g., problems falling and/or staying asleep, or sleep apnea) and if yes, if they still received treatment or not. For our purposes, we did not separate those who still received treatment from those who no longer received treatment. Thus, the value 0 indicated "not diagnosed" and 1 "diagnosed." This information was assessed at t_1 and t_2 (i.e., Studies 1 and 2).

Statistical analysis: We used confirmatory factor analysis (CFA; Brown, 2015) to test the a priori formulated measurement model based on the EFA, and we performed multiple-group CFA (MGCFA Jöreskog, 1971) to test whether the measurement properties of the measurement model were invariant across gender, age groups, and waves of data collection. Measurement invariance (Davidov et al., 2014; Vandenberg & Lance, 2000) ensures that "under different conditions of observing and studying phenomena, measurement operations yield measures of the same attribute" (Horn & Mcardle, 1992, p. 117). When measurement invariance does not hold, comparisons of (co)variances, regression coefficients, and/or means across groups can be biased (Chen, 2007). We considered three levels of invariance: *configural* invariance (the pattern of factors, indicators, and factor loadings is invariant across groups), *metric* invariance (the factor loadings are held equal across groups), and *scalar* invariance (the item intercepts, in addition to the factor loadings, are held equal across groups). We used full information maximum likelihood estimation for CFA and MGCFA, and we treated our data as continuous normal, which can be deemed appropriate in the case of at least five response categories and when using robust corrections to standard errors and test statistics (Rhemtulla et al., 2012). We evaluated the fit of the models using the model chi-square test statistic and alternative fit indices such as the comparative fit index (CFI), root mean square error of approximation (RMSEA), and standardized root mean residual (SRMR). We considered model fit acceptable when CFI $\geq$ 0.90 and RMSEA/SRMR $<$ 0.08 (West et al., 2012). We used *T*-tests to examine whether the scale was able to differentiate between individuals reporting vs. not reporting a sleep disorder diagnosis at t_2. We used Logistic Regression Models (Long & Freese, 2014) to examine whether the AIS-NCA predicts newly diagnosed sleep disorders between t_1 and t_2.[5] We reported odds ratios (OR), whereby ORs above 1 indicate positive effects (i.e., AIS-NCA scores at $t_1$predict new diagnoses at t_2), ORs below 1 point to negative effects, and ORs equaling 1 indicate no effects. We also computed Intraclass Correlation Coefficients (ICC) to assess test-retest reliability over one year, whereby the individual was the cluster for grouping the repeated measurement (Qin et al., 2019).

Results and discussion

The CFA confirms the findings from the EFA in Study 1, suggesting that two factors best represent the data. While a one-factor confirmatory factor model did not fit the data well (χ^2 = 5529.659, df = 14, CFI = 0.823, RMSEA = 0.163, SRMR = 0.077), a two-factor model fit the data well (χ^2 = 1073.076, df = 13, CFI = 0.966, RMSEA = 0.074, SRMR = 0.034). Modification indices of the two-factor model pointed toward a residual covariance between item 2 and item 3, which might be due to the similarity in the content of the items (sleep disruption). After including this additional parameter, the model fit became excellent (χ^2 = 268.615, df = 12, CFI = 0.992, RMSEA = 0.038, SRMR = 0.016). The factor loadings, factor (co)variances, and the residual covariance estimates for this second model are presented in Panel A in Figure 2. All parameters were significant ($p < 0.001$). The standardized factor loadings ranged between 0.449 and 0.949, indicating moderate to (very) strong factor-indicator relationships. The factor variance was larger for the second factor, indicating that the self-assessments of daytime functioning had greater heterogeneity than sleep problems. Moreover, the correlation between the factors ($r = 0.692$)–while clearly showing an association–indicated that they are best treated as separate constructs (Brown, 2015).[6]

Omega coefficients from separate confirmatory factor models for each factor indicated that scale reliability for the second factor was sufficient ($\omega_{Daytime\ functioning}$ = 0.830). However, the omega coefficient was lower for the first factor ($\omega_{Sleep\ problems}$ = 0.722).

We additionally tested whether the two-factor model was invariant across gender (male: $N = 7{,}334$; female: $N = 7{,}463$), three age groups (younger adults, age 18-29: $N = 1{,}110$; middle-aged and older adults, age 30-63: $N = 10{,}439$; older and retired adults, age 64-97: $N = 3{,}248$), two time points (responses for study 1 and study 2: $N = 14{,}797$), and existing sleep disorder diagnosis (not diagnosed: $N = 13{,}213$; diagnosed: $N = 1{,}584$).

Therefore, we specified several multiple-group models for each comparison, beginning with the least constrained configural invariance model followed by metric and scalar invariance models. The model fit statistics for the invariance tests are shown in Table 3.

Scalar invariance was given across gender and time points because the differences in the fit criteria were extremely low when moving from less to more constrained models (see Chen, 2007). Thus, (co)variances, regression coefficients, and means may be compared with confidence across gender and time. However, full scalar invariance may not be given across age groups because the deterioration of fit between the metric and scalar models was considerable (although the global fit of the scalar model was still acceptable). Modification indices suggested that the intercepts of items 2, 4, and 6 were not invariant in the oldest group, which indicated that equal group means on the latent variables "sleep problems" and "daytime functioning" may not correspond with equal observed score means for the corresponding items and therefore that mean comparisons may be invalid. Thus, we estimated a model reflecting partial measurement invariance (Byrne et al., 1989) where we released the invariance constraints for the

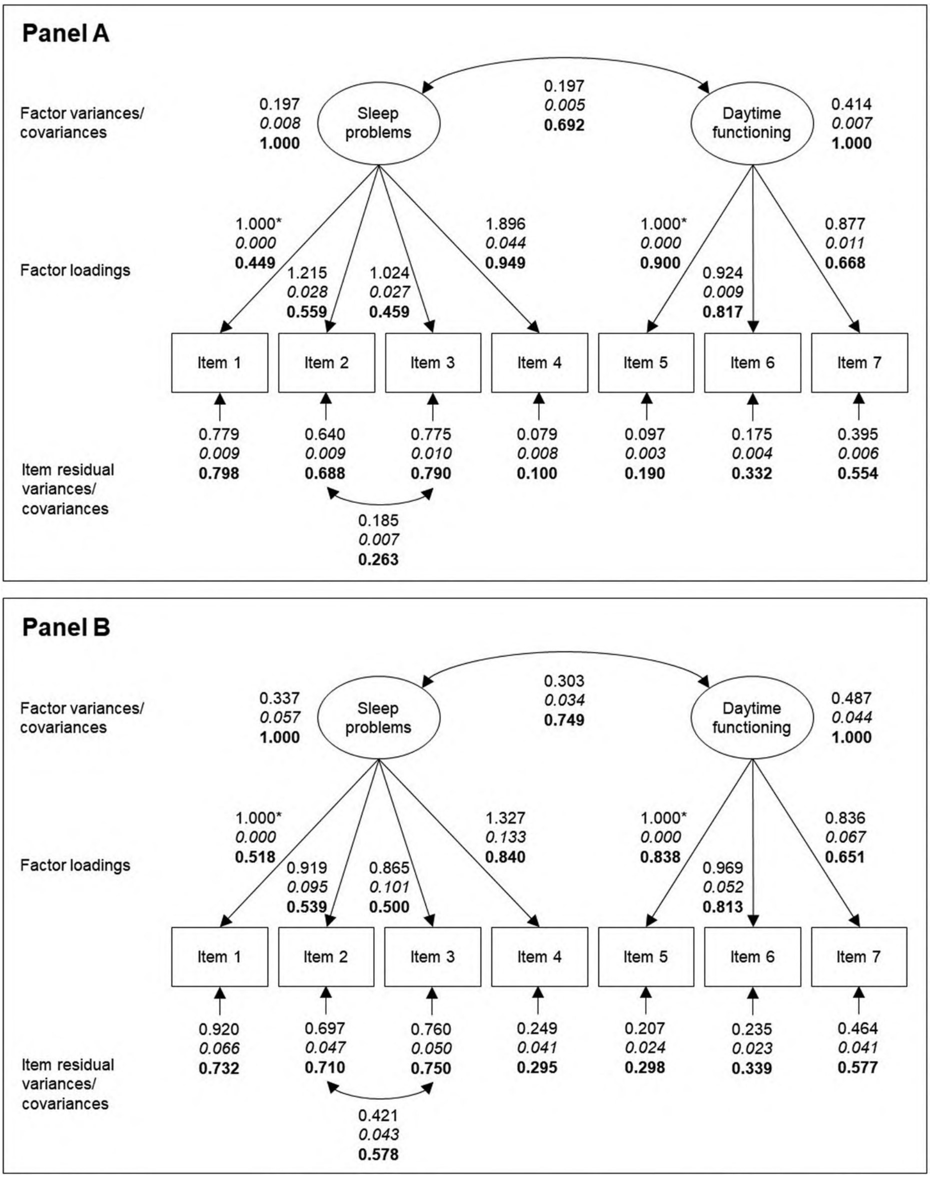

Figure 2. CFA parameter estimates for Study 2 (Panel A, $N = 14{,}797$) and Study 5 (Panel B, $N = 623$). *Notes:* unstandardized estimates are in normal font, standard errors are in italics, standardized estimates are in bold font, *parameter fixed for scale identification.

intercepts of items 2, 4, and 6. The deterioration of fit when moving from the metric model to the partial scalar model was less severe and we concluded that the latent means may still be compared across age groups, but one should be careful when comparing composite score means across age groups (see Steinmetz, 2013). Finally, scalar invariance was given across people with and without a sleep disorder diagnosis.

When weighted average scores for the two subscales and the total scale were used for comparing individuals not diagnosed with a sleep disorder to those with such a

Table 3. Measurement invariance tests (Study 2, $N = 14{,}797$).

Model	χ^2	df	CFI	RMSEA	SRMR
	Across gender (Study 2, $N = 14{,}797$)				
Configural	288.261	24	0.991	0.039	0.017
Metric	299.469	29	0.991	0.036	0.018
Scalar	465.847	34	0.986	0.041	0.026
Model	**Across time points (Study 2, $N = 14{,}797$)**				
	χ^2	df	CFI	RMSEA	SRMR
Configural	1698.488	62	0.982	0.042	0.032
Metric	1713.768	67	0.982	0.041	0.032
Scalar	1800.946	72	0.981	0.040	0.032
Model	**Across age groups (Study 2, $N = 14{,}797$)**				
	χ^2	df	CFI	RMSEA	SRMR
Configural	314.270	36	0.991	0.040	0.018
Metric	376.241	46	0.990	0.038	0.027
Scalar	1237.154	56	0.963	0.065	0.043
Partial scalar	668.185	50	0.981	0.050	0.033
Model	**Across sleep disorder diagnoses (Study 2, $N = 14{,}797$)**				
	χ^2	df	CFI	RMSEA	SRMR
Configural	290.257	24	0.991	0.039	0.017
Metric	306.760	29	0.991	0.036	0.020
Scalar	389.539	34	0.988	0.038	0.024
Model	**Across languages (Studies 4 and 5 combined, $N = 961$)**				
	χ^2	df	CFI	RMSEA	SRMR
Configural	32.493	24	0.996	0.027	0.024
Metric	35.878	29	0.997	0.022	0.028
Scalar	56.636	34	0.989	0.037	0.036

Notes: χ^2=model chi-square test statistic, df = model degrees of freedom, CFI = comparative fit index, RMSEA = root mean square error of approximation, SRMR = standardized root mean residual.

disorder, the latter group had higher values on sleep problems ($M_{not\ diagnosed}$ = 2.84 ± 0.68; $M_{diagnosed}$ = 3.29 ± 0.74; $t(14.795)$ = -24.744; $p < 0.001$; *Cohen's d* = -0.655), daytime functioning ($M_{not\ diagnosed}$ = 2.38 ± 0.63; $M_{diagnosed}$ = 2.82 ± 0.77; $t(14.795)$ = -25.672; $p < 0.001$; *Cohen's d* = -0.686), and the total score ($M_{not\ diagnosed}$ = 2.64 ± 0.57; $M_{diagnosed}$ = 3.09 ± 0.66; $t(14.795)$ = -25.672; $p < 0.001$; *Cohen's d* = -0.769) compared to the first group (see visualization Panel A in Figure S2). Thus, the AIS-NCA can discriminate between both groups with effect sizes ranging from moderate to large.

When examining whether the scales can predict newly diagnosed sleep disorders at t_2 ($N = 508$ out of 13,274 cases indicating no previous t_1), we found that an increase of the sleep problems subscale by one unit increased the likelihood of a new diagnosis by 156.7% (OR = 2.567, $p < 0.001$, see visualization Panel B in Figure S2). This likelihood also increased with higher scores on the daytime functioning subscale (OR = 2.299, $p < 0.001$), and the total score (OR = 3.304, $p < 0.001$). This lends credit to the predictive validity of the AIS-NCA.

The ICC for the overall scale was 0.762 (SE = 0.003), while it was 0.751 (SE = 0.003) for the sleep problems and 0.684 (SE = 0.004) for the daytime functioning subscale. The test-retest reliability over this relatively long interval of approximately one year shows that the measures are relatively stable over this period.

Study 3

This study examined short-term test-retest reliability of the German version of the scale and both subscales.

Methods

Participants: 171 medical students from a German university enrolled in their second and fourth semester were invited to fill out the AIS-NCA. Of these students, 96 entered the first page of the web-based survey and provided an informed consent, 89 (92.71%) completed t_1, of which 78 (81.25%; 70.51% women[7], with 96.15% in the age group 18-30) completed t_2[8] and comprises the analytic sample with complete answers on all relevant variables. The mean distance between participation was 7.00 days (range: 5 to 9 days). The Ethics Committee of the University of Munich approved the study (approval number: 20-1084).

Instruments: AIS-NCA was measured as in Study 2 (with the 7-item scale) but with a reference frame of four weeks and mean-weighted total scores of the scale and both subscales were computed (for the descriptives see Table S2).

Statistical analysis: We computed Intraclass Correlation Coefficient (ICC) to assess test-retest reliability, whereby the individual was the cluster for grouping the repeated measurement (Qin et al., 2019).

Results and discussion

The ICC for the overall scale was 0.883 (SE = 0.025), while it was 0.898 (SE = 0.022) for the sleep problems subscale and 0.756 (SE = 0.048) for the daytime functioning subscale. The test-retest reliability over a short-term interval of approximately 7 days showed that the measures (especially sleep problems) were relatively stable over this period and slightly more stable than after one year (Study 2). While these results suggested stability over time, we must acknowledge that the findings are based on a student sample and that future research may replicate the findings in other populations.

Study 4

The aim of this study was to explore the construct validity (convergent and discriminant) of the scale. Because we found that the items of the AIS-NCA loaded on two factors: “sleep problems” and “daytime functioning,” we decided to use three questionnaires to explore the convergent validity that also assess these and related constructs. We chose two instruments to assess sleep quality/sleep problems, the Pittsburgh Sleep Quality Index (PSQI; Buysse et al., 1989) and the Jenkins Sleep Problems Scale (JSPS; Jenkins et al., 1988), whereby the PSQI and the JSPS have one sub-dimension and one item, respectively, that assess daytime functioning. The third scale we used to measure convergent validity of our factor “daytime functioning” exclusively assessed sleepiness (Epworth Sleepiness Scale, ESS; Johns, 1991). All three questionnaires have been validated and are widely used to assess sleep quality and sleepiness. We expected strong positive correlations ($r > 0.7$; Munro, 2005) between the PSQI, JSPS, and the AIS-NCA subscale “sleep problems,” as well as between ESS, the subscale for daytime functioning of the PSQI, and the AIS-NCA subscale “daytime functioning.” Also, strong correlations were assumed between the three questionnaires and the AIS-NCA total score.

Discriminant validity was explored by comparing the AIS-NCA to a questionnaire assessing several aspects of well-being, namely happiness and life satisfaction, mental and physical health, meaning and purpose, character and virtue, close social relationships, and financial and material stability (VanderWeele, 2017). Compared to the measures used for convergent validity, we expected weaker (but possibly still significant) negative correlations between the AIS-NCA and the different well-being measures, since good sleep is considered an indicator for health and well-being (Tang et al., 2017).

Methods

Participants: We recruited adult participants (18 and older) in Germany via the respondi Online Panel (an actively managed panel used for market research with voluntary participation and a double opt-in registration process).[9] We used a quota sample representative for sex, age (18-74), and province for this web-based survey. The Ethics Committee of the Faculty of Management, Economics, and Social Sciences of the University of Cologne approved the study (approval number: 200028SeSa). The participants received a small incentive for completing the study (€0.45, approximately $0.55). A total of 345 respondents completed the survey. Four respondents did not provide either their sex or age and were therefore excluded from all analyses, leaving a sample of $N = 341$ (49.27% females, mean age 45.88 years ± SD 15.30).

Instruments: AIS-NCA was measured as in Study 2 and 3 (with 7 items and a reference frame of four weeks) and mean-weighted total scores of the scale and both subscales were computed (see Table 4 for descriptives of all questionnaires used and Table S3 for descriptives of the AIS-NCA single items).

Sleep problems and daytime functioning were assessed with three scales. The Pittsburgh Sleep Quality Index (PSQI) is one of the most used scales to assess sleep quality (Buysse et al., 1989). It comprises ten questions (some with several sub-questions) and provides a total score, as well as seven sub-scores. A cut-off for clinically relevant sleep problems is also provided. The Jenkins Sleep Problems Scale (JSPS) is a short 4-item questionnaire designed to assess clinically-relevant sleep problems (Jenkins et al., 1988). The Epworth Sleepiness Scale (ESS) is a short questionnaire asking about the probability of falling asleep in eight different situations (Johns, 1991).

Finally, we measured well-being using the German version of the Flourishing Index (FI), which consists of five dimensions with two questions and statements per dimension (happiness and life satisfaction, physical and mental health, meaning and purpose, character and virtue, and close social relationships) (Sattler et al., 2021; VanderWeele, 2017). The Secure Flourishing Index (SFI) adds the sub-domain financial and material stability with two items. Items were measured on 11-point scales (from 0 to 10) with higher scores indicating higher levels of well-being.

Statistical analysis: We used bivariate Pearson correlations to explore the convergent and discriminant validity of the AIS-NCA.

Results and discussion

Construct validity for the AIS-NCA subscale "sleep difficulties" was supported by strong positive correlations with both the PSQI total score ($r = 0.688$, $p < 0.001$, see Table 5

Table 4. Items and descriptives of validation measures (Study 4, $N = 273\text{-}338^{a}$).

Instrument	Descriptives M	SD	Min	Max	McDonald's ω	Number of observations
Non-Clinical Application (AIS-NCA)						
Sleep problems	2.93	0.72	1.25	5.00	0.755	338
Daytime functioning	2.53	0.72	1.00	5.00	0.848	338
Total score	2.76	0.65	1.14	4.71	0.854	338
Pittsburgh Sleep Quality Index (PSQI)						
Subjective sleep quality	1.26	0.65	0.00	3.00	–[b]	273
Sleep latency	1.14	0.95	0.00	3.00	–[b]	273
Sleep duration	0.91	0.96	0.00	3.00	–[b]	273
Sleep efficiency	0.56	0.89	0.00	3.00	–[b]	273
Sleep disturbance	1.19	0.54	0.00	3.00	–[b]	273
Use sleep medication	0.24	0.71	0.00	3.00	–[b]	273
Daytime functioning	1.04	0.75	0.00	3.00	–[b]	273
Total score	6.34	3.55	0.00	18.00	0.780	273
Jenkins Sleep Problems Scale (JSPS)						
Sum score of items 1, 2, 3	5.42	4.17	0.00	15.00	0.817	332
Item 4	1.61	1.59	0.00	5.00	–[b]	332
Total score	7.03	5.35	0.00	20.00	0.839	332
Epworth Sleepiness Scale (ESS)						
Total score	8.28	4.22	0.00	22.00	0.773	319
Flourishing Index (FI) and Secure Flourishing Index (SFI) with sub-dimensions						
Happiness and Life Satisfaction	12.65	4.02	0.00	20.00	0.854	317
Mental and Physical Health	12.98	4.05	0.00	20.00	0.730	317
Meaning and Purpose	12.96	4.81	0.00	20.00	0.827	317
Character and Virtue	14.24	3.20	5.00	20.00	0.434	317
Close Social Relationships	13.91	4.75	0.00	20.00	0.902	317
Financial and Material Stability	12.43	6.31	0.00	20.00	0.942	317
FI	66.74	16.57	5.00	100.00	0.896	317
SFI	79.16	19.43	5.00	120.00	0.888	317

Notes: [a]Differences in number of observations between scales derived from a different number of missing observations for each (sub-)scale (due to the complexity and length of the PSQI, more missing values were observed). [b]Not applicable. M = Mean, SD = Standard deviation; Min = Minimum, Max = Maximum, ω=McDonald's ω.

for all correlations) and the JSPS total score ($r = 0.747$, $p < 0.001$). The correlation between the ESS and the AIS-NCA subscale "daytime functioning" was weaker ($r = 0.346$, $p < 0.001$). However, the correlations between the AIS-NCA subscale "daytime functioning" and both the PSQI sub-score "daytime functioning" ($r = 0.672$, $p < 0.001$) and item 4 of the JSPS ("Wake up after your usual amount of sleep feeling tired or worn out?") were strong ($r = 0.711$, $p < 0.001$). The ESS had overall weak correlations with all other questionnaires, indicating that the ESS might address sleepiness specifically, rather than overall daytime functioning. Moreover, the AIS-NCA total score was strongly and positively correlated with the PSQI total score ($r = 0.749$, $p < 0.001$) and the JSPS total score ($r = 0.783$, $p < 0.001$).

The correlations between the different domains and scores of well-being and the AIS-NCA including its subscales indicated satisfactory discriminant validity. All correlations were negative and mainly weaker (r-values between –0.144 and –0.509, all p-values <0.05) compared to the correlations reported for convergent validity. Altogether, the AIS-NCA subscales and overall scale showed good convergent and discriminant validity.

Table 5. Inter-correlations of AIS-NCA, PSQI, SPS, ESS, and well-being (Study 4, $N = 272\text{-}331^{a}$).

	AIS-NCA: Sleep problems	AIS-NCA: Daytime functioning	AIS-NCA: Total score	Number of observations
Convergent validity				
Pittsburgh Sleep Quality Index (PSQI)				
Subjective sleep quality	0.676***	0.633***	0.736***	272
Sleep latency	0.575***	0.430***	0.573***	272
Sleep duration	0.369***	0.327***	0.392***	272
Sleep efficiency	0.327***	0.263***	0.334***	272
Sleep disturbance	0.479***	0.447***	0.520***	272
Use sleep medication	0.295***	0.244***	0.305***	272
Daytime functioning	0.453***	0.672***	0.614***	272
Total score	0.688***	0.645***	0.749***	272
Jenkins Sleep Problems Scale (JSPS)				
Sum score of items 1, 2, 3	0.751***	0.555***	0.742***	331
Item 4	0.544***	0.711***	0.686***	331
Total score	0.747***	0.644***	0.783***	331
Epworth Sleepiness Scale (ESS)				
ESS: Total score	0.161**	0.346***	0.268***	318
Discriminant validity				
Flourishing Index (FI) and Secure Flourishing Index (SFI) with sub-dimensions				
Happiness and Life Satisfaction	−0.321***	−0.437***	−0.411***	316
Mental and Physical Health	−0.441***	−0.484***	−0.509***	316
Meaning and Purpose	−0.218***	−0.395***	−0.326***	316
Character and Virtue	−0.155**	−0.292***	−0.238***	316
Close Social Relationships	−0.226***	−0.317***	−0.294***	316
Financial and Material Stability	−0.144*	−0.214***	−0.193***	316
FI	−0.344***	−0.486***	−0.449***	316
SFI	−0.339***	−0.483***	−0.445***	316

Notes: [a]Differences in number of observations between scales derived from a different number of missing observations for each (sub-)scale (due to the complexity and length of the PSQI, more missing values were observed);. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Study 5

To increase the developed scale's usability, we translated it to English (see stages 5 and 6 in Figure 1). The translation followed international guidelines for cross-cultural adaptation of health-related self-report measures (Beaton et al., 2000; Guillemin et al., 1993). First, the items were translated from German to English by two professional translators (native English speakers). Then, back-translations were completed by two professional bilingual native German speakers. Finally, discrepancies between translations and the original items were discussed by three team members (of whom two are native German speakers), two professional linguists in English (one was a native English speaker), one survey expert from Sociology (native German speaker), the original two translators, and the two back-translators. We informed the group that we wanted the translation to be fully comprehensible to the majority of people (e.g., short sentences, as simple as possible, active rather than passive voice, avoiding metaphors) and to consider cross-cultural source and conceptual equivalences (Guillemin et al., 1993). Table 6 shows the results of the translation process. The aim of Study 5 was to assess the factor structure and the internal consistency of the scale translated into English to offer a tool for a larger community and to test for measurement invariance across the two languages. Thus, the sample is used to validate and replicate the findings from Study 1 and Study 2.

Methods

Participants: We recruited adult participants (18 and older) in the United States via Clickworker, a commercial platform on which individuals can earn additional income for different tasks, such as survey participation. Of the more than two million accounts worldwide, approximately 46% are located in North America and 35% of account holders speak U.S. American English (Clickworker, 2021). For this study, 676 entered the first page of the web-based survey, 669 (98.96%) provided informed consent, of which 665 (99.40%) indicated commitment to participation in an attention check (Detailed Ethics Statement in Supplement Information). Due to missing values, the analytical sample comprises of 623 individuals (68.70% identified as female, 29.86% as male, and 1.44% as other, mean age: 34.08 ± 10.35 SD, age range: 18-69). The participants received a small incentive for completing the study (€1.35, approximately $1.60).

Instruments: We used the newly translated AIS-NCA (7 items) with a reference of four weeks and computed mean-weighted total scores of the scale and both subscales (for the descriptives see Table 6).

Statistical analysis: We provided descriptive information about the distribution of the answers (i.e., mean values, standard deviations, minimum, maximum, skewness, and excess kurtosis, see Study 1), computed a CFA to examine the dimensionality, ran a multiple-group CFA to test for measurement invariance across languages (with data from Study 4), and examined internal consistency with *McDonald's ω* (see Study 2).

Results and discussion

Table 6 shows the mean, standard deviation, minimum, maximum, skewness, and excess kurtosis for all items. When testing for the normality of the data, we found that scores for all items are within acceptable limits regarding $skewness_{range}$ = -0.09 to 0.32 and $excess\ kurtosis_{range}$ = -0.75 to 0.31, with most scores pointing towards fairly symmetrical distributions, with some slight tendencies towards being platykurtic.

The CFA confirms a two-dimensional structure, validating the findings from Study 1 and Study 2. However, standard errors of the CFA parameters were higher due to the smaller sample size in Study 5. Without a residual covariance between item 2 and item 3, the model did not fit the data (χ^2 = 200.975, df = 13, CFI = 0.848, RMSEA = 0.152, SRMR = 0.085). The model fit improved substantially when we included the residual covariance (χ^2 = 22.403, df = 12, CFI = 0.992, RMSEA = 0.037, SRMR = 0.025), which confirmed that the additional parameter is needed. All parameters were significant ($p < 0.001$, Panel B in Figure 2). The standardized factor loadings ranged between 0.500 and 0.840, indicating moderate to strong factor-indicator relationships. Similar to Study 2, we found a larger factor variance for the second factor, suggesting greater heterogeneity in respondents' self-assessments of daytime functioning than sleep problems. The correlation between the factors was 0.749. The residual correlation of item 2 and item 3 was strong, indicating that a part of the variance in both items (that is not explained by the sleep problems factor) may be rooted in a common unobserved cause or equal wording.

Table 6. Items and descriptives of the AIS-NCA (Study 5, $N = 623$).

		Descriptives					
#	Item and response options	M	SD	Min	Max	Skew	Kurt
1	I could usually get to sleep (after turning off the lights)… *immediately; after a very short time; after a short time; after a long time; after a very long time*	3.16	1.12	1.00	5.00	−0.07	−0.75
2	Waking up during my sleep happened… *never; almost never; sometimes; quite often; very often*	3.50	0.99	1.00	5.00	−0.09	−0.60
3	Waking up prematurely happened… *never; almost never; sometimes; quite often; very often*	3.35	1.01	1.00	5.00	−0.02	−0.56
4	The overall quality of my sleep was usually… *very good; good; sometimes good/ sometimes bad; bad; very bad*	2.78	0.92	1.00	5.00	0.25	0.10
5	Throughout the day, my level of well-being was usually… *very good; good; sometimes good/sometimes bad; bad; very bad*	2.55	0.83	1.00	5.00	0.32	0.31
6	Throughout the day, my level of (physical and mental) performance was usually… *very good; good; sometimes good/sometimes bad; bad; very bad*	2.53	0.83	1.00	5.00	0.12	−0.10
7	Throughout the day, my level of tiredness was usually… *not perceptible; hardly perceptible; moderately perceptible; very perceptible; very strongly perceptible*	3.00	0.90	1.00	5.00	0.20	−0.06
	AIS-NCA: Sleep problems	3.20	0.76	1.00	5.00	0.10	−0.24
	AIS-NCA: Impaired daytime functioning	2.70	0.72	1.00	5.00	0.17	0.04
	AIS-NCA: Total score	2.98	0.66	1.29	4.71	0.08	−0.30

Notes: The instruction was: "How do you rate the following aspects related to your sleep? Please base your answers on the past four weeks."; M = Mean, SD = Standard deviation; Min = Minimum, Max = Maximum; Skew = Skewness, Kurt = Excess kurtosis.

Omega coefficients from separate confirmatory factor models for each factor were $\omega_{Sleep\ problems} = 0.748$ and $\omega_{Daytime\ functioning} = 0.807$, pointing towards a sufficient degree of scale reliability for both factors when residual correlations are not considered. According to Table 3 full scalar measurement invariance is given across languages allowing comparisons of (co)variances, regression coefficients, and means.

Overall discussion

Insufficient and inadequate sleep is common in our society, and it has been linked to several health issues and performance deficits (Grandner, 2017; Watson et al., 2015). To better understand the implications of (poor) sleep for health and performance, longitudinal, prospective studies are needed. Objective measures of sleep (e.g., polysomnography) are unfortunately costly, cumbersome, and limited to relatively small sample sizes. Questionnaires, despite their limitations intrinsic to subjective measures, have the advantage of being time and cost-efficient, and give the possibility of collecting longitudinal within-subjects large datasets that can be used to explore cause-effect relationships and predictive effects of different variables (e.g., sleep quality and health-related outcomes).

The rationale behind developing the AIS-NCA was providing such a sleep scale, since many available scales are comparatively long and/or designed mainly for clinical applications, i.e., diagnosing sleep problems. Instead of developing an entirely new scale, we decided to adapt an existing scale, the AIS, because it is short, has very good psychometric properties, and has already been used in several studies (Chiu et al., 2016). By adapting the scale, we developed a widely-applicable tool, not only for screening and diagnosing insomnia, but also for exploring the great variety of sleep characteristics in the general population, ranging from very good sleep to subclinical and clinical sleep problems with different degrees of severity. To achieve this, items were rephrased and response options were changed from mainly negative to more balanced between negative and positive. Here, we presented four studies that aimed at describing the dimensional structure of the AIS-NCA, examined measurement invariance across different respondent groups and time, assessed its ability to discriminate between people with and without a diagnosis of sleep disorders and to predict emerging sleep disorders, as well as tested the scale's convergent/discriminant validity, internal consistency, and test-retest reliability. The validation studies were performed with German-speaking participants. Moreover, we provided data suggesting that the same dimensional structure of the scale can be found in an English-speaking sample and can show measurement invariance across the two languages. Additional validation studies with English-speaking participants could further confirm the good psychometric properties found here.

Dimensional structure: Our exploratory and confirmatory factor analyses identified a two-factor structure with 7 out of 8 items loading on the factors "sleep problems" (items 1, 2, 3, 4) and "daytime functioning" (items 5, 6, 7; for final item numbering, see Figure S1). One item assessing sleep duration was discarded because of a considerable cross loading on two factors. This is not surprising, since sleep duration partly depends on how much time one needs to fall asleep (item 1) and influences daytime functioning. Several experimental and observational studies have in fact shown the detrimental effects of sleep deprivation on daytime functioning (e.g., cognitive performance; Lim & Dinges, 2010; Lowe et al., 2017). A two-factor structure of the AIS has been shown in previous studies (Chung et al., 2011; Enomoto et al., 2018; Hallit et al., 2019; Okajima et al., 2013; Yen et al., 2010) and was therefore expected, although it was not found in the original validation study (Soldatos et al., 2000). Still, Soldatos and colleagues suggested that the first five items could be used on their own as a short AIS (Soldatos et al., 2000). Having a short scale that measures sleep quality and its consequences for daytime functioning is also an advantage of the AIS-NCA because it can be used to examine how these two important aspects of sleep relate to each other and other variables. For instance, one could explore whether sleep problems or daytime functioning have a greater impact on health-related outcomes and performance indexes.

Internal consistency: Internal consistency of both subscales and the overall scale were all above *McDonald's* $\omega > 0.700$. Thus, despite the brevity of the scale, its reliability was satisfactory.

Measurement invariance: Measurement invariance analysis shows that the measures of both sleep problems and impaired daytime functioning were scalar invariant across gender, over time (one year), and across existing sleep disorder diagnoses. Thus, (co) variances regression coefficients and mean scores may be compared without bias. However, for age the measures showed only partial scalar invariance, meaning that the latent means may be compared across age groups, but one should be cautious when comparing composite means of very distinct age groups.

Validity: Construct validity was assessed by looking at both convergent and discriminant validity. Because of the two-factor structure we wanted to test convergent validity using different questionnaires developed to specifically test sleep problems and daytime functioning. Convergent validity for the factor "sleep problems" was very good (all r-coefficients were close to or above 0.7). Convergent validity of "daytime functioning," assessed with the ESS for sleepiness, was less satisfactory ($r = 0.346$). This relatively low correlation may indicate that sleepiness and daytime functioning, despite being related, are two different constructs. This is supported by relatively low correlations between the ESS and some of the items from the PSQI and the JSPS that also assess daytime functioning. The correlation of these items with the factor "daytime functioning" of the AIS-NCA was relatively high (0.672 with PSQI and 0.711 with JSPS). Future validation studies may therefore use another questionnaire to more specifically assess daytime functioning. Discriminant validity was supported by lower correlations with various dimensions of well-being that assess a construct other than sleep, but still were expected to correlate with sleep characteristics (Gothe et al., 2020; Tang et al., 2017). Additionally, we found that the scale was able to discriminate between participants with and without a diagnosis of sleep problems (Study 2). We also found evidence for good predictive validity of the AIS-NCA since it was able to predict new sleep disorders within one year. Thus, the AIS-NCA may be also useful in clinical settings.

Test-retest reliability: Finally, test-retest reliability was affirmed both for short (7 days) and long (1 year) term intervals, suggesting very good stability of the scale over different time periods.

Applicability: The AIS was originally developed for research and clinical applications as an alternative to the traditionally lengthy and cumbersome sleep scales that are currently in use. For instance, the PSQI is considered the gold standard to assess sleep quality, but it is long, and the score calculations are complex. We developed the AIS-NCA with the same goal and argue that our new scale is more feasible in research settings, in which the focus is not necessarily only on clinically-relevant sleep problems (e.g., insomnia) but also on a broader assessment of sleep quality. Whether the AIS-NCA can be also used as a diagnostic tool for insomnia needs to be further elucidated. For now, the AIS provides clear cut-off scores and it is, therefore, better suited for clinical (diagnostic) applications (Soldatos et al., 2003). Future studies may also further evaluate the suggested advantages of our scale with qualitative and quantitative designs, by testing, for example, whether the avoidance of abstract technical terms (e.g., sleep induction) increases understanding and results in more valid responses, or whether the changes in the response options and the reduction of floor effects increase the predictive validity of the scale and allow for a more sensitive assessment of sleep problems and impaired daytime functioning (including sub-clinical manifestations).

Concluding remarks: With an extensive series of studies, we confirmed that the adapted AIS scale has good psychometric properties. Based on the exploratory factor analysis, one item was discarded, making the AIS-NCA even shorter, with a total of just 7 items. Given the two-factor structure, which was tested and supported through confirmatory factor analysis, researchers can choose to use just one of the two subscales according to their specific research questions. Altogether, the AIS-NCA is a short (on average, it takes less than 1 minute to fill out the scale), efficient, economical, valid, and reliable 7-item questionnaire that can be used in non-clinical research settings to measure sleep problems and daytime functioning. It might be especially suitable for large, multi-themed population studies to monitor changes and stabilities in these aspects of sleep, as well as to examine how they influence or are caused by other constructs, such as health and well-being.

Notes

1. The abbreviation ADM stands for "Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V." (Association of German Market and Social Research Institutes). Within this association, there is a group of market research agencies in Germany that is responsible for the sampling frame to member agencies. This is a sampling frame used for representatively selecting telephone samples of the population living in private households (Gäbler & Häder 1998; ADM, 2012).
2. The sample is based on an initial sample (37,003 invitees, 24,085 consenting individuals, and 22,024 completers) with an extended sample (10,403 invitees, 3,064 consenting individuals, and 2,785 completers) to counteract demographic imbalances (due to selective participation of more difficult-to-reach participants), and thus, to increase representation.
3. In order to ease interpretation and to enable a comparison to the standard normal distribution, we adjusted the kurtosis by subtracting the value 3, meaning that any value other than zero implies excess kurtosis.
4. We ran a logit regression model to test if sleep problems resulted in selective dropout and found no such effect (OR = 1.027, $p = 0.213$).
5. Respondents who indicated a diagnosed sleep disorder at t_1 were excluded from the analysis, since only new diagnoses were of interest.
6. In addition, we tested this model including the omitted item 4 (in this case the analytic sample decreased to $N = 14{,}792$). Modification indices pointed toward a cross-loading between item 4 and the second factor (impaired daytime functioning). When we included this cross-loading, standardized factor loadings of item 4 were 0.382 on the first factor and 0.311 on the second factor, which confirms our decision to omit item 4 from the scale.
7. Please note that more than half of the participants in Study 3 were females, which may limit the generalizability of the results. We suggest that future studies use samples that are more representative.
8. We ran a logit regression model to test if sleep problems resulted in selective dropout and found a tendency towards a lower probability of re-participation with increasing sleep problems (OR = 0.93). This effect failed to reach conventional levels of statistical significance ($p = 0.890$).
9. The provider takes care of data quality by an elaborate scoring and control process of the panel members.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the German Research Foundation [SA 2992/2-1; to Sebastian Sattler]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The views expressed do not necessarily reflect the policies of the funder. The authors declare that there is no conflict of interest.

ORCID

Sebastian Sattler http://orcid.org/0000-0002-6491-0754
Daniel Seddig http://orcid.org/0000-0003-1595-6948
Giulia Zerbini http://orcid.org/0000-0002-5348-9212

References

ADM. (2012). *ADM-Forschungsprojekt' Dual-Frame-Ansätze' 2011/2012*. Forschungsbericht. https://www.adm-ev.de/wp-content/uploads/2018/07/Forschungsbericht-Dual-Frame-Ans%C3%A4tze.pdf

Bannai, A., & Tamakoshi, A. (2014). The association between long working hours and health: A systematic review of epidemiological evidence. *Scandinavian Journal of Work, Environment & Health*, *40*(1), 5–18. https://doi.org/10.5271/sjweh.3388

Barnes, C. M., & Watson, N. F. (2019). Why healthy sleep is good for business. *Sleep Medicine Reviews*, *47*, 112–118. https://doi.org/10.1016/j.smrv.2019.07.005

Barros, M. B. D. A., Lima, M. G., Ceolim, M. F., Zancanella, E., & Cardoso, T. A. M. D. O. (2019). Quality of sleep, health and well-being in a population-based study. *Revista de Saúde Pública*, *53*, 82. https://doi.org/10.11606/s1518-8787.2019053001067

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, *25*(24), 3186–3191. https://doi.org/10.1097/00007632-200012150-00014

Benham, G., & Charak, R. (2019). Stress and sleep remain significant predictors of health after controlling for negative affect. *Stress and Health*, *35*(1), 59–68. https://doi.org/10.1002/smi.2840

Besedovsky, L., Lange, T., & Haack, M. (2019). The sleep-immune crosstalk in health and disease. *Physiological Reviews*, *99*(3), 1325–1380. https://doi.org/10.1152/physrev.00010.2018

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). The Guilford Press.

Brownlow, J. A., Miller, K. E., & Gehrman, P. R. (2020). Insomnia and cognitive performance. *Sleep Medicine Clinics*, *15*(1), 71–76. https://doi.org/10.1016/j.jsmc.2019.10.002

Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Research*, *28*(2), 193–213. https://doi.org/10.1016/0165-1781(89)90047-4

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*(3), 456–466. https://doi.org/10.1037/0033-2909.105.3.456

Cappuccio, F. P., & Miller, M. A. (2017). Sleep and cardio-metabolic disease. *Current Cardiology Reports*, *19*(11), 110. https://doi.org/10.1007/s11886-017-0916-0

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Chiu, H.-Y., Chang, L.-Y., Hsieh, Y.-J., & Tsai, P.-S. (2016). A meta-analysis of diagnostic accuracy of three screening tools for insomnia. *Journal of Psychosomatic Research*, *87*, 85–92. https://doi.org/10.1016/j.jpsychores.2016.06.010

Chung, K.-F., Kan, K. K.-K., & Yeung, W.-F. (2011). Assessing insomnia in adolescents: Comparison of insomnia severity index, Athens insomnia scale and sleep quality index. *Sleep Medicine*, *12*(5), 463–470. https://doi.org/10.1016/j.sleep.2010.09.019

Clickworker. (2021). Retrieved January 5, 2021, from Our Clickworker community website: https://www.clickworker.com/clickworker-crowd/

Conroy, D. A. (2017). The role of sleep on the pathway to substance abuse in teens. *Journal of Adolescent Health*, *60*(2), 129–130. https://doi.org/10.1016/j.jadohealth.2016.11.011

Crönlein, T., Langguth, B., Popp, R., Lukesch, H., Pieh, C., Hajak, G., & Geisler, P. (2013). Regensburg Insomnia Scale (RIS): A new short rating scale for the assessment of psychological symptoms and sleep in insomnia; Study design: Development and validation of a new short self-rating scale in a sample of 218 patients suffering from insomnia and 94 healthy controls. *Health and Quality of Life Outcomes*, *11*(1), 65. https://doi.org/10.1186/1477-7525-11-65

Crowley, S. J., Wolfson, A. R., Tarokh, L., & Carskadon, M. A. (2018). An update on adolescent sleep: New evidence informing the perfect storm model. *Journal of Adolescence*, *67*, 55–65. https://doi.org/10.1016/j.adolescence.2018.06.001

Dai, H., Mei, Z., An, A., Lu, Y., & Wu, J. (2020). Associations of sleep problems with health-risk behaviors and psychological well-being among Canadian adults. *Sleep Health*, *6*(5), 657–661. https://doi.org/10.1016/j.sleh.2020.02.003

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology*, *40*(1), 55–75. https://doi.org/10.1146/annurev-soc-071913-043137

Enomoto, K., Adachi, T., Yamada, K., Inoue, D., Nakanishi, M., Nishigami, T., & Shibata, M. (2018). Reliability and validity of the Athens Insomnia Scale in chronic pain patients. *Journal of Pain Research*, *Volume 11*, 793–801. https://doi.org/10.2147/JPR.S154852

Fatima, Y., Doi, S. A. R., & Mamun, A. A. (2016). Sleep quality and obesity in young subjects: A meta-analysis: Sleep quality and obesity. *Obesity Reviews : An Official Journal of the International Association for the Study of Obesity*, *17*(11), 1154–1166. https://doi.org/10.1111/obr.12444

Gabler, S., & Häder, S. (1998). Ein neues Stichprobendesign für telefonische Umfragen in Deutschland. In S. Gabler, S. Häder, & J. H. P. Hoffmeyer-Zlotnik (Eds.), *Telefonstichproben in Deutschland*. Springer.

George, D., & Mallery, M. (2010). *SPSS for windows step by step: A simple guide and reference, 17.0 update*. Allyn & Bacon.

Gomez Fonseca, A., & Genzel, L. (2020). Sleep and academic performance: Considering amount, quality and timing. *Current Opinion in Behavioral Sciences*, *33*, 65–71. https://doi.org/10.1016/j.cobeha.2019.12.008

Gómez-Benito, J., Ruiz, C., & Guilera, G. (2011). A Spanish version of the Athens Insomnia Scale. *Quality of Life Research : An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *20*(6), 931–937. https://doi.org/10.1007/s11136-010-9827-x

Gothe, N. P., Ehlers, D. K., Salerno, E. A., Fanning, J., Kramer, A. F., & McAuley, E. (2020). Physical activity, sleep and quality of life in older adults: Influence of physical, mental and social well-being. *Behavioral Sleep Medicine*, *18*(6), 797–808. https://doi.org/10.1080/15402002.2019.1690493

Grandner, M. A. (2017). Sleep, health, and society. *Sleep Medicine Clinics*, *12*(1), 1–22. https://doi.org/10.1016/j.jsmc.2016.10.012

Guillemin, F., Bombardier, C., & Beaton, D. (1993). Cross-cultural adaptation of health-related quality of life measures: Literature review and proposed guidelines. *Journal of Clinical Epidemiology*, *46*(12), 1417–1432. https://doi.org/10.1016/0895-4356(93)90142-n

Gulia, K. K., & Kumar, V. M. (2018). Sleep disorders in the elderly: A growing challenge: Sleep in elderly. *Psychogeriatrics : The Official Journal of the Japanese Psychogeriatric Society*, *18*(3), 155–165. https://doi.org/10.1111/psyg.12319

Hallit, S., Haddad, C., Hallit, R., Al Karaki, G., Malaeb, D., Sacre, H., Kheir, N., Hajj, A., & Salameh, P. (2019). Validation of selected sleeping disorders related scales in Arabic among the Lebanese population. *Sleep and Biological Rhythms*, *17*(2), 183–189. https://doi.org/10.1007/s41105-018-0196-0

Hasler, B. P., Smith, L. J., Cousins, J. C., & Bootzin, R. R. (2012). Circadian rhythms, sleep, and substance abuse. *Sleep Medicine Reviews*, *16*(1), 67–81. https://doi.org/10.1016/j.smrv.2011.03.004

Horn, J. L., & Mcardle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, *18*(3–4), 117–144. https://doi.org/10.1080/03610739208253916

Jenkins, C. D., Stanton, B.-A., Niemcryk, S. J., & Rose, R. M. (1988). A scale for the estimation of sleep problems in clinical research. *Journal of Clinical Epidemiology*, *41*(4), 313–321. https://doi.org/10.1016/0895-4356(88)90138-2

Jeong, H. S., Jeon, Y., Ma, J., Choi, Y., Ban, S., Lee, S., Lee, B., Im, J. J., Yoon, S., Kim, J. E., Lim, J.-H., & Lyoo, I. K. (2015). Validation of the Athens Insomnia Scale for screening insomnia in South Korean firefighters and rescue workers. *Quality of Life Research : An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *24*(10), 2391–2395. https://doi.org/10.1007/s11136-015-0986-7

Johns, M. W. (1991). A new method for measuring daytime sleepiness: The epworth sleepiness scale. *Sleep*, *14*(6), 540–545. https://doi.org/10.1093/sleep/14.6.540

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*(4), 409–426. https://doi.org/10.1007/BF02291366

Kalmbach, D. A., Anderson, J. R., & Drake, C. L. (2018). The impact of stress on sleep: Pathogenic sleep reactivity as a vulnerability to insomnia and circadian disorders. *Journal of Sleep Research*, *27*(6), e12710. https://doi.org/10.1111/jsr.12710

Kaplow, R. (2016). Sleep disturbances and critical illness. *Critical Care Nursing Clinics of North America*, *28*(2), 169–182. https://doi.org/10.1016/j.cnc.2016.02.005

Khubchandani, J., & Price, J. H. (2020). Short sleep duration in working American adults, 2010–2018. *Journal of Community Health*, *45*(2), 219–227. https://doi.org/10.1007/s10900-019-00731-9

Kim, E.-J., & Dimsdale, J. E. (2007). The effect of psychosocial stress on sleep: A review of polysomnographic evidence. *Behavioral Sleep Medicine*, *5*(4), 256–278. [Database] https://doi.org/10.1080/15402000701557383

Krishnan, V., & Collop, N. A. (2006). Gender differences in sleep disorders. *Current Opinion in Pulmonary Medicine*, *12*(6), 383–389. https://doi.org/10.1097/01.mcp.0000245705.69440.6a

Kuppermann, M., Lubeck, D. P., Mazonson, P. D., Patrick, D. L., Stewart, A. L., Buesching, D. P., & Filer, S. K. (1995). Sleep problems and their correlates in a working population. *Journal of General Internal Medicine*, *10*(1), 25–32. https://doi.org/10.1007/BF02599573

Lam, T., Ogeil, R. P., Allsop, S., Chikritzhs, T., Fischer, J., Midford, R., Gilmore, W., Lenton, S., Liang, W., Lloyd, B., Aiken, A., Mattick, R., Burns, L., & Lubman, D. I. (2018). Insomnia and regulation of sleep-wake cycle with drugs among adolescent risky drinkers. *Journal of Clinical Sleep Medicine*, *14*(09), 1529–1537. https://doi.org/10.5664/jcsm.7330

Li, J., Vitiello, M. V., & Gooneratne, N. S. (2018). Sleep in normal aging. *Sleep Medicine Clinics*, *13*(1), 1–11. https://doi.org/10.1016/j.jsmc.2017.09.001

Lim, J., & Dinges, D. F. (2010). A meta-analysis of the impact of short-term sleep deprivation on cognitive variables. *Psychological Bulletin*, *136*(3), 375–389. [Database] https://doi.org/10.1037/a0018883

Lin, C. -Y., Cheng, A. S. K., Nejati, B., Imani, V., Ulander, M., Browall, M., Griffiths, M. D., Broström, A., & Pakpour, A. H. (2020). A thorough psychometric comparison between Athens Insomnia Scale and Insomnia Severity Index among patients with advanced cancer. *Journal of Sleep Research*, *29*(1), e12891. https://doi.org/10.1111/jsr.12891

Logan, R. W., Hasler, B. P., Forbes, E. E., Franzen, P. L., Torregrossa, M. M., Huang, Y. H., Buysse, D. J., Clark, D. B., & McClung, C. A. (2018). Impact of sleep and circadian rhythms on addiction vulnerability in adolescents. *Biological Psychiatry*, *83*(12), 987–996. https://doi.org/10.1016/j.biopsych.2017.11.035

Long, J. S., & Freese, J. (2014). *Regression models for categorical dependent variables using Stata* (3rd ed., Vol. 7). Stata press.

Lowe, C. J., Safati, A., & Hall, P. A. (2017). The neurocognitive consequences of sleep restriction: A meta-analytic review. *Neuroscience & Biobehavioral Reviews*, *80*, 586–604. https://doi.org/10.1016/j.neubiorev.2017.07.010

Magnavita, N., & Garbarino, S. (2017). Sleep, health and wellness at work: A scoping review. *International Journal of Environmental Research and Public Health*, *14*(11), 1347. https://doi.org/10.3390/ijerph14111347

Mander, B. A., Winer, J. R., & Walker, M. P. (2017). Sleep and human aging. *Neuron*, *94*(1), 19–36. https://doi.org/10.1016/j.neuron.2017.02.004

McDermott, M., Brown, D. L., & Chervin, R. D. (2018). Sleep disorders and the risk of stroke. *Expert Review of Neurotherapeutics*, *18*(7), 523–531. https://doi.org/10.1080/14737175.2018.1489239

Moloney, M. E., Ciciurkaite, G., & Brown, R. L. (2019). The medicalization of sleeplessness: Results of U.S. office visit outcomes, 2008–2015. *SSM - Population Health*, *8*, 100388. https://doi.org/10.1016/j.ssmph.2019.100388

Morin, C. M., Belleville, G., Bélanger, L., & Ivers, H. (2011). The insomnia severity index: Psychometric indicators to detect insomnia cases and evaluate treatment response. *Sleep*, *34*(5), 601–608. https://doi.org/10.1093/sleep/34.5.601

Munro, B. H. (2005). *Statistical methods for health care research* (5th ed). Lippincott Williams & Wilkins.

Okajima, I., Nakajima, S., Kobayashi, M., & Inoue, Y. (2013). Development and validation of the Japanese version of the Athens Insomnia Scale: Development of Japanese version of AIS. *Psychiatry and Clinical Neurosciences*, *67*(6), 420–425. https://doi.org/10.1111/pcn.12073

Pallesen, S., Bjorvatn, B., Nordhus, I. H., Sivertsen, B., Hjørnevik, M., & Morin, C. M. (2008). A new scale for measuring insomnia: The Bergen Insomnia Scale. *Perceptual and Motor Skills*, *107*(3), 691–706. https://doi.org/10.2466/pms.107.3.691-706

Patel, S. R., & Hu, F. B. (2008). Short sleep duration and weight gain: A systematic review. *Obesity (Silver Spring, Md.)*, *16*(3), 643–653. https://doi.org/10.1038/oby.2007.118

Pilcher, J. J., & Morris, D. M. (2020). Sleep and organizational behavior: Implications for workplace productivity and safety. *Frontiers in Psychology*, *11*, 45. https://doi.org/10.3389/fpsyg.2020.00045

Qin, S., Nelson, L., McLeod, L., Eremenco, S., & Coons, S. J. (2019). Assessing test-retest reliability of patient-reported outcome measures using intraclass correlation coefficients: Recommendations for selecting and documenting the analytical formula . *Quality of Life Research : An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *28*(4), 1029–1033. https://doi.org/10.1007/s11136-018-2076-0

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*(3), 354–373. https://doi.org/10.1037/a0029315

Riemann, D., Baglioni, C., Bassetti, C., Bjorvatn, B., Dolenc Groselj, L., Ellis, J. G., Espie, C. A., Garcia-Borreguero, D., Gjerstad, M., Gonçalves, M., Hertenstein, E., Jansson-Fröjmark, M., Jennum, P. J., Leger, D., Nissen, C., Parrino, L., Paunio, T., Pevernagie, D., Verbraecken, J., … Spiegelhalder, K. (2017). European guideline for the diagnosis and treatment of insomnia. *Journal of Sleep Research*, *26*(6), 675–700. https://doi.org/10.1111/jsr.12594

Sattler, S., Lee, M. T., & Wilkinson, R. (2021). *A Brief Measure of Complete Subjective Well-Being in Germany: A Population-Based Validation of a German Version of the Flourish Index (FI) and the Secure Flourish Index (SFI)*. Unpublished manuscript.

Sirajudeen, M. S., Dilshad Manzar, M., Alqahtani, M., Alzhrani, M., Albougami, A., Somasekharan Pillai, P., Spence, D. W., & Pandi-Perumal, S. R. (2020). Psychometric properties of the Athens Insomnia Scale in occupational computer users. *Healthcare*, *8*(2), 89. https://doi.org/10.3390/healthcare8020089

Soldatos, C. R., Dikeos, D. G., & Paparrigopoulos, T. J. (2000). Athens Insomnia Scale: Validation of an instrument based on ICD-10 criteria. *Journal of Psychosomatic Research*, *48*(6), 555–560. https://doi.org/10.1016/s0022-3999(00)00095-7

Soldatos, C. R., Dikeos, D. G., & Paparrigopoulos, T. J. (2003). The diagnostic validity of the Athens Insomnia Scale. *Journal of Psychosomatic Research*, *55*(3), 263–267. https://doi.org/10.1016/s0022-3999(02)00604-9

Statista. (2020). *Haushalte in Deutschland—Internetzugang bis 2019*. Retrieved July 8, 2020, from https://de.statista.com/statistik/daten/studie/153257/umfrage/haushalte-mit-internetzugang-in-deutschland-seit-2002

Steinmetz, H. (2013). Analyzing observed composite differences across groups: Is partial measurement invariance enough? *Methodology*, *9*(1), 1–12. https://doi.org/10.1027/1614-2241/a000049

Sun, J.-L., Chiou, J.-F., & Lin, C.-C. (2011). Validation of the Taiwanese version of the Athens Insomnia Scale and Assessment of Insomnia in Taiwanese cancer patients. *Journal of Pain and Symptom Management*, *41*(5), 904–914. https://doi.org/10.1016/j.jpainsymman.2010.07.021

Tang, N. K. Y., Fiecas, M., Afolalu, E. F., & Wolke, D. (2017). Changes in sleep duration, quality, and medication use are prospectively associated with health and well-being: Analysis of the UK household longitudinal study. *Sleep*, *40*(3), zsw079. https://doi.org/10.1093/sleep/zsw079

Van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. (1994). *The think aloud method: A practical approach to modelling cognitive*. AcademicPress.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–70. https://doi.org/10.1177/109442810031002

VanderWeele, T. J. (2017). On the promotion of human flourishing. *Proceedings of the National Academy of Sciences*, *114*(31), 8148–8156. https://doi.org/10.1073/pnas.1702996114

Watson, N. F., Badr, M. S., Belenky, G., Bliwise, D. L., Buxton, O. M., Buysse, D., Dinges, D. F., Gangwisch, J., Grandner, M. A., Kushida, C., Malhotra, R. K., Martin, J. L., Patel, S. R., Quan, S. F., Tasali, E., Twery, M., Croft, J. B., Maher, E., Barrett, J. A., Thomas, S. M., & Heald, J. L. (2015). Recommended amount of sleep for a healthy adult: A joint consensus statement of the american academy of sleep medicine and sleep research society. *Journal of Clinical Sleep Medicine : JCSM : Official Publication of the American Academy of Sleep Medicine*, *11*(6), 591–592. https://doi.org/10.5664/jcsm.4758

West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In Rick H. Hoyle (ed.), *Handbook of structural equation modeling* (pp. 209–231). The Guilford Press.

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, *34*(6), 806–838. https://doi.org/10.1177/0011000006288127

Yen, C.-F., King, B. H., & Chang, Y.-P. (2010). Factor structure of the Athens Insomnia Scale and its associations with demographic characteristics and depression in adolescents. *Journal of Sleep Research*, *19*(1), 12–18. https://doi.org/10.1111/j.1365-2869.2009.00758.x