Title
Evaluating the Use of Commercially Available Wearable Wristbands to Capture Adolescents' Daily Sleep Duration

Journal
Journal of Research on Adolescence, 29(3)

ISSN
1050-8392

Authors
George, Madeleine J
Rivenbark, Joshua G
Russell, Michael A
et al.

Publication Date
2019-09-01

DOI
10.1111/jora.12467

Peer reviewed

Published in final edited form as:

J Res Adolesc. 2019 September ; 29(3): 613–626. doi:10.1111/jora.12467.

Evaluating the Use of Commercially Available Wearable Wristbands to Capture Adolescents' Daily Sleep Duration

Madeleine J. George,
Purdue University

Joshua G. Rivenbark,
Duke University

Michael A. Russell
Pennsylvania State University

Leonard Ng'eno, **Rick H. Hoyle**
Duke University

Candice L. Odgers
Duke University and University of California, Irvine

Abstract

Commercially available wearable devices are marketed as a means of objectively capturing daily sleep easily and inexpensively outside of the laboratory. Two ecological momentary assessment studies—with 120 older adolescents (aged 18–19) and 395 younger adolescents (aged 10–16)—captured nightly self-reported and wearable (Jawbone) recorded sleep duration. Self-reported and wearable recorded daily sleep duration were moderately correlated ($r \sim .50$), associations which were stronger on weekdays and among young adolescent boys. Older adolescents self-reported sleep duration closely corresponded with estimates from the wearable device, but younger adolescents reported having an hour more of sleep, on average, compared to device estimates. Self-reported, but not wearable-recorded, sleep duration and quality were consistently associated with daily well-being measures. Suggestions for the integration of commercially available wearable devices into future daily research with adolescents are provided.

Sleep is crucially important for adolescent health and development. The duration and quality of adolescents' sleep have been associated with their overall cognitive functioning (e.g., memory consolidation and inattention) and academic performance (Dewald, Meijer, Oort, Kerkhof, & Bogels, 2010; Wolfson & Carskadon, 1998). Inadequate sleep can have short- and long-term consequences for adolescents' health and well-being (for reviews see, e.g., Beebe, 2011; Owens, 2014; Shocat, Cohen-Zion, & Tzischinsky, 2014). Yet, most U.S. adolescents do not receive the recommended 8–10 hr of sleep per night (Carskadon, Mindell, & Drake, 2006; Keyes, Maslowsky, Hamilton, & Schulenberg, 2015), as national

Requests for reprints should be sent to Madeleine J. George, 701 S. 3rd St., West Lafayette, IN 47907-2059. georg107@purdue.edu.

Supporting Information
Additional supporting information may be found online in the Supporting Information section at the end of the article.
Appendix S1. Overview of study measures.

surveys have found that adolescents sleep, on average, 7.5 hr per night and more than half (58%) sleep less than 7 hr per night (Emsellem et al., 2014).

Parents and researchers alike are concerned about deficits in adolescents' sleep health and are eager to find ways to better study, understand, and improve adolescents' daily sleep habits. Studying daily sleep habits is important because both average sleep per night and variability in sleep patterns are associated with well-being (Fuligni, Arruda, Krull, & Gonzales, 2017; He et al., 2015; Lemola, Ledermann, & Friedman, 2013). A daily study with 750 adolescents (aged 14–15) found that on days when adolescents report less sleep during the night, they also report being more tired and experiencing more internalizing symptoms the following day, compared to themselves on days when they slept longer (Fuligni & Hardway, 2006).

The recent influx of commercially available wearable devices has dramatically increased the ease and accessibility with which adolescents' daily sleep may be monitored. Diary methods have long been used in psychological research to capture self-reports of daily behaviors and experiences (Iida, Shrout, Laurenceau, & Bolger, 2012; Nezlek, 2012; Shiffman, Stone, & Hufford, 2008). Mobile devices may also offer new ways of collecting real-time, objective data about adolescents' sleep patterns. Wearable devices hold promise for sleep researchers to capture adolescents' sleep habits in their everyday lives without the need for expensive or intrusive equipment or data collection methods. However, there is a need to understand the best practices for using commercial wearable devices in research. This paper reports findings from two daily studies, in which commercial wearable devices and daily self-report methodologies are used to measure daily sleep duration. Results provide practical guidance for researchers interested in studying adolescents' daily sleep patterns and behaviors.

Current Methodologies used in Daily Sleep Research

Given the importance of daily sleep health during adolescence, it is valuable to consider the possible ways that sleep can be measured in daily life and the relative merits and limitations of these methods for research. In this section, we detail the more frequently used methods for studying daily sleep health with adolescents. For each type of methodology, we explain (1) the components of sleep health the method can measure, (2) the merits of using the method in daily research, and (3) the limitations of using the method in daily research.

The traditional "gold standard" for studying sleep behaviors is polysomnography. Polysomnography measures eye and muscle movement, heart and respiratory rate, and electroencephalography (EEG) of brain waves to create a detailed picture of a participant's night of sleep (Marino et al., 2013). This method captures many of the components of sleep health such as sleep stages (e.g., slow wave, rapid eye movement [REM]), duration, and disturbances (Buysse, 2014). This method captures extensive features of sleep health measured objectively through physiological changes, some of which cannot be measured through self-reported items. However, the use of polysomnography in daily studies is expensive and does not capture the subjective components of sleep health and functioning, such as sleep quality. Polysomnography also requires specialized equipment and is typically conducted in a sleep laboratory, limiting its ecological validity. Although some studies have successfully used polysomnography in participants' homes (Stores, Fry, & Crawford, 1998; Zheng et al., 2012), the unusual or obtrusive nature of the equipment itself may disrupt typical sleep patterns. Thus, while polysomnography is the gold standard for studying sleep in the laboratory, this method is of limited value for research on the quality and quantity of adolescents' typical nightly sleep patterns in their own homes.

Self-report sleep scales are the most widely used measures of sleep in psychological research with adolescents (for a review see Ji & Liu, 2016). These measures often include single self-report items for different types of sleep patterns, including sleep duration, quality, disturbances or problems, and insomnia. Scales of sleep behaviors and health (e.g., Pittsburgh Sleep Quality Index; Epworth Sleepiness Scale) have been well validated in research with adolescents (e.g., Chung, Kan, & Yeung, 2011; Janssen, Phillipson, O'Connor, & Johns, 2017) and relate to their overall health and well-being (e.g., Dewald et al., 2010; Shocat et al., 2014). Self-report sleep scales can capture important subjective components of sleep health that are critical for the diagnosis and treatment of sleep and associated mental health problems (Mindell & Owens, 2015). Cross-sectional surveys that include self-reported sleep measures are highly feasible with large samples as they can be administered online and completed quickly. However, these measures rely on individuals to report accurately on their behaviors from days or months past, which may be subject to substantial recall bias. Self-reported measures also miss many of the physiological components of sleep health. Thus, although these methods measure a general assessment of adolescents' sleep behaviors or problems, they generally do not capture detailed, within-individual changes in sleep and associations with daily well-being.

Self-reports of sleep behaviors have been used in diary studies, including ecological momentary assessment (EMA; Shiffman et al., 2008) and experience sampling methodologies (ESM; Csikszentmihalyi, Larson, & Prescott, 1977). Diary methods can repeatedly and intensively assess sleep health, including sleep duration and quality, over a number of days. These "in the moment" measures of adolescents' behaviors or experiences reduce retrospective recall biases, resulting in more accurate self-reports (Conner & Barrett, 2012; Shiffman, 2009). Daily self-reported sleep measures have been validated with cross-sectional survey and polysomnography techniques (Rogers, Caruso, & Aldrich, 1993) and may provide more ecologically valid ways of assessing adolescents' daily sleep than laboratory or cross-sectional survey methods. However, associations between self-reported sleep measures and self-reported health outcomes are subject to shared method variance bias, in which some of the associations between self-reported sleep and subjective well-being can be caused by the common method of data collection rather than true individual or day-to-day differences (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). In daily sleep research with adolescents, self-reports of sleep duration and quality are the most commonly used and associated with well-being.

Wearable devices (e.g., wrist-worn actigraphy watches, chest-worn belts) measure momentary changes in heart rate and movement to record daily sleep duration and some types of sleep state and disturbances (e.g., waking from sleep, sleep efficiency and latency; Eatough & Shockley, 2016). "Research-grade" wearable devices (e.g., Actiwatch, Minimitter, MicroMini Motionlogger) may provide a less biased, objective measurement of sleep duration than daily self-reports. Daily studies using research-grade wearable devices, often referred to as daily actigraph studies, show moderate to high associations ($r \sim .30$ to .87) between device-recorded and participant-reported daily measures of sleep duration (Sadeh, 2011). For example, in a sample of over 300 adolescents (ages 13–20) followed over an 8-day period, there were moderate-to-strong associations between self-reported and actigraph (i.e., MicorMini Motionlogger) recorded sleep duration measures, especially on weekdays ($r = .61$) relative to weekend ($r = .38$) days (Wolfson et al., 2003). Research suggests that sleep duration recorded by actigraph devices is associated with, but may underestimate, sleep duration compared to laboratory polysomnography estimates of sleep (Johnson et al., 2007; Marino et al., 2013). However, more studies are needed that validate actigraph devices in normative adolescent samples (for a review see Meltzer, Montgomery-Downs, Insana, & Walsh, 2012). Furthermore, research-grade wearable devices may be uncomfortable or noticeable to wear (similar to laboratory polysomnography), can be expensive for large samples, and miss subjective sleep quality.

Using Commercial Devices in Daily Sleep Research

Commercial wearable devices (e.g., FitBit, Jawbone, Garmin) may afford researchers a cheaper alternative to research-grade wearable devices for the objective measurement of certain sleep behaviors. As with research-grade devices, these wrist-worn devices capture participants' movements and/or heart rate across the night to assess sleep duration and possible disturbances. Accompanying apps use algorithms to automatically translate these movements into daily estimates of sleep onset, efficiency, and duration. The increasing availability and affordability of commercial wearable devices capable of tracking daily sleep patterns beg the question of how these devices can be used in daily research with adolescents.

Studies are needed to assess the validity of using commercially available wearable devices to measure sleep, especially in children's and adolescents' daily lives. Studies comparing commercial wearable-recorded sleep (i.e., FitBit Flex, FitBit ChargeHR, FitBit Ultra) and sleep measures from polysomnography have found that many wearables have a high accuracy and sensitivity, but low specificity in detecting sleep states—that is, they are better at capturing sleep onset than interruptions, or wake from sleep, and they tend to overestimate total sleep and sleep efficiency (Cook, Prairie, & Plante, 2017; de Zambotti et al., 2016; Meltzer, Hiruma, Avis, Montgomery-Downs, & Valentin, 2015). A study of 63 children and adolescents aged 13–17 years found that the commercial devices overestimated (Fitbit Ultra Normal) or underestimated (Fitbit Ultra Sensitive) total sleep time compared to actigraphy and polysomnography (Meltzer et al., 2015). However, a study with 32 youth (12–21 years old) found comparable estimates of sleep with commercial devices (FitBit ChargeHR) and polysomnography (de Zambotti et al., 2016) and a study with 40 adults found similar sleep estimates when comparing actigraphy, commercial devices, and polysomnography (Mantua, Gravel, & Spencer, 2016). Similarly, a study of 21 adults with depression comparing sleep patterns collected via polysomnography, commercial (Fitbit Flex), and actigraphy (Actiwatch) devices found that they produced similar estimates of sleep duration, with the Actiwatch only slightly more accurate than the Fitbit Flex in its "normal setting" (Cook et al., 2017). Thus, it appears that, in small studies with adults, commercial devices can reliably assess sleep duration and these estimates are highly correlated with typical research-grade measures of sleep in the laboratory.

For researchers studying sleep health in adolescents, the limitations of the studies to date using commercial wearable devices are that they typically: (1) focus on adults and/or clinical samples, (2) have small samples (<100 participants), and (3) have non-daily study designs (i.e., do not have repeated objective sleep measures with corresponding actigraph or daily self-report measures). Thus, it remains unclear whether sleep duration assessed through commercial wearable devices is associated with self-reports of sleep duration and quality in adolescents, and how those measures in turn relate to other daily aspects of adolescents' well-being. Studies are needed that use multiple methods of capturing daily sleep with adolescent samples. Unfortunately, logistic hurdles and economic limitations present significant barriers to an idealized study that would compare and evaluate all possible sleep measurement methods at once; however, there are still scientific and methodological insights to be gained from smaller scope studies. For example, comparisons with actigraphs may show interchangeability in research-grade versus commercial devices, whereas comparisons with self-report methods may allow researchers to capture multiple components of sleep (i.e., duration and quality) and help to bridge findings with self-reported assessments gathered in cross-sectional studies.

Aims of the Present Studies

The goal of this article is to use evidence from daily studies as a basis for discussing the practical advantages and disadvantages to two types of ambulatory methodologies for studying adolescents' sleep behaviors: self-reported sleep (both duration and quality) and sleep duration as recorded by commercially available wearable wristbands. Specifically, we report findings from two relatively large EMA studies with adolescents. The first is a study with older adolescents (120 first year college students aged 18–19) and the second comprises young adolescents (395 adolescents aged 10–16). Four main research questions are addressed across the two daily studies: (1) How much daily sleep do adolescents receive as assessed by self-reported and wearable-recorded sleep measures? (2) What are the correlations between self-reported and wearable-recorded sleep duration? (3) Are there differences in the strength of the associations between these sleep measures by day of the week, age, or gender? and (4) Are self-reported and wearable-recorded sleep measures similarly associated with daily affect and well-being?

STUDY 1

Method

Participants.—One hundred and twenty-four older adolescents (aged 18–19) were recruited from a southeastern U.S. university in the 2016–2017 academic year and consented to participate in this 7–10-day EMA study. Four participants were excluded from the study because they were not in their first year of college or they did not complete any of the daily assessments, resulting in a final sample of 120 first year college students (70% women; 50% White, 24% Asian or Asian American, 13% Latino(a) or Hispanic, 11% Black or African

American, and 2% Other). Most students had highly educated parents (68% of had a professional degree).

Procedure and measures.—Students completed a brief baseline assessment and were given a wearable wristband (Jawbone UP3) with an accompanying application (UP) on their phone (Jawbone, 2014). The wristband passively monitored their steps and their sleep duration each day for the study period. Students also downloaded a survey application (MetricWire Inc, 2015) onto their personal mobile phones. Each day, students completed five short 2–3 min surveys every 4 hr—one in the morning (7 am), three during the daytime (11 am, 3 pm, and 7 pm), and one in the evening (11 pm). Students were given options for the settings of the notifications (i.e., vibrate, banner only) and had at least 4 hr to complete each survey so that the prompts would not disturb their normal activities (i.e., sleep, class). Each morning, students reported on their current affect, well-being, and the previous nights' sleep duration and quality. In each of the daytime and evening survey signals, students reported on their current affect and well-being since the previous survey. Each evening, students also reported on their emotional and attentional regulation across the entire day. Seventy-eight percent of survey prompts were answered, resulting in 4,681 total observations and giving an average of 7.8 study days per person. For the wearable devices, 104 of the 120 participants wore a functioning device on multiple nights, and of the 104, the completion rate across the 10 days was 75% (778 observations) with an average of 6.5 days per person.

Daily sleep experiences were assessed each day in three ways. Objective estimates of sleep duration were recorded with the Jawbone UP3 wearable wristband devices that students wore throughout the study period. As documented by Jawbone, the application creates an estimate of sleep duration with a proprietary algorithm that combines data from device's heart rate sensor and accelerometer (without bed or wake time input from the user). Because the devices report days without sleep data as "0 hours of sleep", the Jawbone data was cleaned by setting zero values to missing to prevent underestimating sleep duration. Although it is possible that some of these values represented "real" nights students did not sleep, it was more likely that the students did not wear the devices or that the devices ran out of battery (as discussed during exit interview with the participants). The data were also cleaned by setting values to missing if the recorded duration was between 2 hr and 50 min and 3 hr of sleep (four instances) due to a discovered error in processing of sleep duration when the devices were charged and active but not worn. Self-reported daily sleep duration was assessed each morning with a single self-reported estimate of the number of hours spent asleep during the night (i.e., How many hours did you sleep last night?). Subjective sleep quality was also assessed each morning with a single item (i.e., How well did you sleep last night? 0- poor to 9-well). The intraclass correlation coefficient (ICC) for the daily reports of sleep was calculated in the multilevel modeling framework to assess the within- versus between-person variance (Raudenbush & Bryk, 2002). The ICCs were calculated for wearable-recorded sleep duration (0.13), self-reported sleep duration (0.36), and subjective sleep quality (0.26); these show, for example, that 13% of the variance in daily recorded sleep duration was between participants, whereas the remaining 87% of the variance was within participants over time.

Daily affect and well-being were assessed in three ways each day. First, students reported on their affect at each survey assessment. They rated four adjectives for their current negative affect including, “anxious”, “stressed”, “angry”, and “sad”, and three adjectives for their positive affect including, “happy”, “excited”, and “calm” on the 5-point scale of their current state (1 = *definitely do not feel* to 5 = *definitely feel*) (adapted from PANAS and daily measures: Schimmack, 2003; Watson, Clark, & Tellegan, 1988). Composites were created by averaging the negative and positive items at each assessment time and across the entire day (alphas were calculated at each assessment and then averaged for the first full day of the study, negative affect: α = .74 and positive affect: α = .67). Second, students reported on their state self-esteem (i.e., How do you feel about yourself right now? 1 = *really bad* to 5 = *really good*) at each survey signal. A daily composite was created by averaging ratings across the entire day. Third, daily emotion dysregulation and inattention was measured only during the evening survey (adapted from the Affect Regulation Checklist; Moretti, 2003; and the Brief Self-Control Scale; Tangney, Baumeister, & Boone, 2004). Students reported on four binary (yes/no) questions about their attentional and emotional regulation throughout the day (i.e., Today, I had a hard time concentrating or focusing; Today, I have been doing or saying things without thinking; I am having a hard time controlling my emotions; even little things are getting on my nerves). A composite measure was created by counting the number of symptoms (0–4). Person-level averages for all sleep and well-being measures are presented in Table 1. The ICCs were then calculated for daily negative affect (0.55), positive affect (0.55), self-esteem (0.50), and dysregulation (0.30).

Analytical plan.—Basic descriptive statistics and person-level correlations were conducted in STATA15 (StataCorp, 2017) to find the average amount of sleep and associations between sleep measures. Multilevel models testing the within-individual associations between reported and recorded sleep measures and between sleep measures and daily well-being were conducted in Mplus8 (MPlus, 2017). Daily well-being composites were used to keep all measures on the same daily timescale. For the multilevel analyses, days were included for analyses only when there were available sleep measures (objective or self-report) and other missing values were estimated using a maximum likelihood with robust standard errors procedure. Equation 1 shows the multilevel model testing the same-day within-person association between daily wearable-recorded (WR) and self-reported (SR) sleep duration. Equation 2 shows an example multilevel model testing the same-day within-person association between daily self-reported sleep duration (SR) and negative affect (NA).

$$\left(\mathrm{WR_{ij}}\right) = \beta_0 + \beta_1\left(\mathrm{SR_{ij}}\right) + \beta_2\left(\mathrm{mSR_j}\right) + u_{0j} + \varepsilon_{ij} \quad (1)$$

$$\left(\mathrm{NA_{ij}}\right) = \beta_0 + \beta_1\left(\mathrm{SR_{ij}}\right) + \beta_2\left(\mathrm{mSR_j}\right) + u_{0j} + \varepsilon_{ij} \quad (2)$$

β_0 is the sample average intercept representing the average wearable sleep or negative affect on days without self-reported sleep. $\beta_1(\mathrm{SR_{ij}})$ is the within-person association in wearable-recorded sleep or negative affect by amount of self-reported sleep. $\beta_2(\mathrm{mSR_j})$ is the person-level (i.e., each person’s mean) association testing whether students who reported more sleep on average also experienced higher average wearable-recorded sleep or reported higher average negative affect, adjusting for their level of self-reported sleep on the current day. The random intercept [u_{0j}] captures the between-person variation in students' average wearable sleep duration or negative affect. The residual [ε_{ij}] captures the within-person variation in wearable sleep duration or negative affect that is unexplained by daily self-reported sleep.

Results

To address the first research question, Table 1 provides the person-level averages for the sleep and well-being measures. On average, the wearable devices recorded that students slept under 6 and ½ hr per night ($M = 6.41$, $SD = 1.19$) and students self-reported a similar average amount of sleep across the study period ($M = 6.32$, $SD = 1.25$). When comparing days when students had both self-reported and device-recorded sleep duration, students reported sleeping about 13 more minutes on average than was recorded on the devices ($M_{diff} = 0.22$, $SD = 1.71$).

To address the second research question, we tested the associations between wearable-recorded and self-reported sleep in three ways: (1) basic correlations in the sleep measures across all observation days, (2) the average person-level correlations across the study period, and (3) the daily within-individual associations. First, Table 1 shows the basic correlations between all the daily measures. There was a moderate correlation between wearable-recorded sleep duration and self-reported sleep duration ($r = .51$, $p < .001$) and a significant, but small, correlation between wearable-recorded sleep duration and subjective sleep quality ($r = .24$, $p < .001$). Second, there were weak, non significant person-level (i.e., each participant's average across the EMA period) associations between average wearable-recorded and self-reported sleep duration ($r = .15$, $p = .12$) and subjective sleep quality ($r = .11$, $p = .25$). Third, multilevel models were estimated to find the daily within-person associations between the sleep measures. Daily wearable-recorded sleep duration was significantly associated with same-day self-reported sleep duration ($\beta = .43$, 95% confidence interval (CI) [0.30, 0.55], $p < .001$) and subjective sleep quality ($\beta = .21$, 95% CI [0.12, 0.29], $p < .001$).

To address our third research question, we then tested whether there were differences in the associations between the sleep measures by daily context or individual characteristics. The length of the EMA period in Study 1 allowed for testing differences by day of the week such that a weekend marker was created to test the overlap of sleep measures on both week ($n \sim 600$) and weekend days ($n \sim 100$). There was a stronger association between recorded and reported sleep duration on weekdays ($r = .55$, $p < .001$) compared to weekend days ($r = .23$, $p = .06$), $z = -2.87$, $p = .004$. There was no statistical difference in the strength of the association between device-recorded sleep duration and subjective sleep quality on weekdays ($r = .26$, $p < .001$) versus weekend days ($r = .11$, $p = .39$), $z = -1.20$, $p = .23$. Individual differences in correlations are only reported for Study 2 (below) because the Study 1 sample was smaller (120) with little variation in age (all were first year college students, 18–20 years old) and majority (70%) female (see Appendix S1 in the online Supporting Information for analyses).

To address the fourth research question, multilevel models tested whether the daily sleep measures were associated with daily well-being and the results are shown in Table 2. These multilevel models estimated the same-day associations between measures of sleep and well-being outcomes, controlling for person-level means of the predictors to isolate associations at the daily, within-person level. Table 2 shows that wearable-recorded sleep duration was significantly and weakly associated with daily self-esteem ($\beta = .10$, $p = .01$), but the associations with same-day negative affect ($\beta = -.02$, $p = .52$), positive affect ($\beta = .04$, $p = .26$) and emotion dysregulation ($\beta = -.01$, $p = .76$) were not statistically significant. Conversely, self-reported sleep duration and subjective sleep quality were more consistently associated with well-being and were more likely to remain significant after false discovery rate (FDR) p-value correction (12 tests: $p < .004$). Specifically, reported sleep duration and sleep quality were weakly to moderately associated with same-day negative affect (duration: $\beta = -.14$, $p < .001$; quality: $\beta = -.17$, $p < .001$), positive affect (duration: $\beta = .08$, $p = .02$; quality: $\beta = .15$, $p < .001$) and self-esteem (duration: $\beta = .17$, $p < .001$; quality: $\beta = .21$, $p < .001$). This pattern of findings remained consistent when comparing associations by week- or weekend day (see Appendix S1 in the online Supporting Information).

STUDY 2

Method

Participants.—In Study 2, 395 adolescents (aged 10–16) were recruited to participate in a home visit and a 14-day EMA study between April 2016 and February 2017. Participants were drawn from a larger longitudinal study ($N = 2{,}104$), the Research on Adaptive Interests, Skills and Environments (RAISE) study, which follows a large representative sample of children and adolescents in North Carolina (see Rivenbark et al., in press, for more detailed information). The EMA sample was split evenly by gender (49.6% female), was racially and ethnically diverse (60.2% non-Hispanic White, 19.4% non-Hispanic Black, 13.8% Hispanic, 6.6% other race), and came from a mix of economic backgrounds (40.8% qualified as economically disadvantaged based on administrative records).

Procedure and measures.—During the home visit, adolescents completed a set of surveys, and interviewers installed the survey application (MetricWire Inc, 2015) onto the participant's own mobile phone or a study-administered phone (49.9% of adolescents used their own phone). Participants received three daily surveys (morning, afternoon, and evening) for the next 14 days that assessed their daily sleep behaviors, affect, and well-being. Eighty percent of survey prompts were answered, resulting in 13,017 total observations across an average of 13.6 days per participant. A subset of the sample ($n = 256$) also received the same type of wearable wristband (Jawbone, 2014) as Study 1 to passively monitor their nightly sleep duration until the battery on the device died, resulting in 856 person-days with wearable-recorded sleep data or an average of 3.3 days of recorded sleep per adolescent. Although we aimed to capture similar constructs, there were some differences in the procedures and measures used in Study 2 compared to Study 1, which we detail below (see also Appendix S1 in the online Supporting Information for a table of these measure differences).

Daily sleep patterns were assessed each day in three ways. Objective estimates of sleep duration were recorded and cleaned in the same manner as Study 1 (i.e., with the removal of 11 cases that recorded 2 hr 50 min to 3 hr of sleep from the Jawbone device). Self-reported daily sleep duration was assessed each morning with a two-item self-reported estimate of bed and wake times (i.e., What time did you go to sleep? What time did you wake up?), from which sleep duration was calculated. Bed and wake times were cleaned so that unreasonable or impossible values—perhaps from mixing up AM and PM—were corrected (e.g., more than 18 hr of sleep or negative values). Cases of uncertainty in self-reports of sleep were set to missing. Subjective sleep quality was assessed each morning with a single item about sleep problems (i.e., How well did you sleep last night? 0 = *very well*, 1 = *some trouble*, 2 = *very badly*). This item was then dichotomized (i.e., 0 = *no sleep problems*, 1 = *poor sleep*) for each day to create a marker for days when participants experienced poor sleep quality. Intraclass correlations coefficients (ICC) were calculated for wearable-recorded sleep duration (0.28), self-reported sleep duration (0.24), and subjective sleep problems (0.27).

Daily affect and well-being were assessed in three ways each day. Negative and positive affect daily composites were created from slightly different items from Study 1. At each survey assessment, adolescents rated four adjectives for their current negative affect including “nervous”, “stressed”, “mad”, and “sad”, and three adjectives for their positive affect including, “happy”, “excited”, and “calm” on the 100 point scale (1 = *not at all* to 100 = *very*) of their current state (adapted from PANAS; Schimmack, 2003; Watson et al., 1988). Composites were again created by averaging the across the entire day (average alphas during the first full day of the study, negative affect: α = .77 and positive affect: α = .65). Adolescents then reported on their state self-esteem and emotion dysregulation and inattention with the same questions as in Study 1, but emotion dysregulation and inattention items rated at every survey signal on a five-point scale (0 = *not at all* to 4 = *very*) averaged to create a composite daily measure of dysregulation (average alpha during the first full day of the study: α = .73). Person-level averages are presented in Table 3. The ICCs were then calculated for daily negative affect (0.64), positive affect (0.63), self-esteem (0.69) and dysregulation (0.73).

Results

Average amounts of sleep across the EMA period are presented in Table 3. On average, the wearable devices recorded that adolescents’ slept about 7 hr per night (M = 7.11, SD = 1.69), whereas adolescents reported sleeping almost eight and half hours per night (M = 8.48, SD = 1.06) on average across the EMA study period. When comparing days with both reported and recorded sleep duration, adolescents reported sleeping about 1 hr and 17 more minutes on average than was recorded on the devices (M_{diff} = 1.28, SD = 2.14).

To examine our second research question, we tested the basic, average person-level, and the daily within-individual associations between sleep measures. Table 3 provides the basic correlations between the sleep measures. First, with wearable-recorded sleep duration, there is a moderate and statistically significant correlation with reported sleep duration (r = .41, p < .001) and a small correlation with subjective sleep problems (r = .07, p = .04). Second, there was a small person-level association between average recorded and reported sleep duration ($r = .28$, $p < .001$), but not with average sleep problems ($r = -.02$, $p = .71$). Third, results from multilevel models demonstrate that daily wearable-recorded sleep duration was weakly to moderately associated with same-day self-reported sleep duration ($\beta = .36$, $p < .001$) and subjective sleep problems ($\beta = .15$, $p = .004$).

To address the third research question, we conducted a set of analyses to test the correlations across sleep measures separately by age and by gender, because the sample size was larger and more varied in age than Study 1. Recorded sleep duration was more strongly correlated with reported sleep duration in boys ($r = .52$, $p < .001$) than girls ($r = .34$, $p < .001$), $z = -2.97$, $p = .003$, and for younger adolescents 10–13 years old ($r = .48$, $p < .001$) compared to adolescents 14–16 years old ($r = .33$, $p < .001$), $z = -2.50$, $p = .01$. There were no statistical differences in the associations between recorded sleep duration and subjective sleep problems by gender (boys: $r = -.15$, $p = .003$; girls: $r = -.02$, $p = .77$), $z = -1.89$, $p = .06$, or by age (10–13 year olds: $r = -.07$, $p = .17$; 14–16 year olds: $r = -.09$, $p = .09$), $z = -0.27$, $p = .79$. Because the Study 2 had a shorter duration of recorded sleep, we do not report differences by day of the week (see Appendix S1).

As in Study 1, multilevel models estimated the same-day associations between measures of sleep and well-being outcomes and the results are shown in Table 4. Wearable-recorded sleep duration was not associated with same-day reports of affect and well-being (ps > .05). Self-reported sleep duration and subjective sleep problems were weakly associated with adolescents' same-day negative affect (duration: $\beta = .04$, $p = .03$; quality: $\beta = .10$, $p < .001$) and dysregulation (duration: $\beta = .04$, $p = .04$; quality: $\beta = .10$, $p < .001$), and sleep problems were weakly associated with positive affect ($\beta = -.09$, $p < .001$) and self-esteem ($\beta = -.11$, $p < .001$). After p-value correction (12 tests: $p < .004$), only associations between sleep quality problems and well-being remained significantly linked. Overall, results were consistent by gender and age (see Appendix S1 in the online Supporting Information).

GENERAL DISCUSSION

The two EMA studies using both wearable-recorded and self-report assessments of daily sleep produced four main sets of findings. First, older adolescents (18–19 year olds) slept an average of about 6 hr nightly as measured by both the wearable devices and the self-reported item as measured across the 7–10 day EMA period. This is in contrast to the notably larger mean difference of more than an hour in reported versus recorded sleep duration found in the younger adolescent sample (i.e., 10–16 year olds in Study 2) and in previous actigraph studies with younger adolescents (Arora, Broglia, Pushpakumar, Lodhi, & Taheri, 2013; Matthews, Hall, & Dahl, 2014). This difference may be due to a number of individual factors (e.g., population differences, greater accuracy in self-reported sleep in older adolescents) or features of the wearable data (e.g., greater accuracy in wearable-recorded sleep in older adolescents). College students live in a much different environment when compared to what young adolescents are experiencing with respect to sleep (e.g., living in a dorm or apartment away from parental home, bed-times monitored by parents), which may contribute to differences in the reported and recorded average sleep duration. For example, young adolescents' parents may encourage them to go to bed, even before they are ready to sleep, resulting in overestimation of self-reported sleep. In general, the total average sleep duration recorded by the devices in these studies corresponds with previous studies estimates of average sleep duration in younger and older adolescents (Emsellem et al., 2014; Lund, Reider, Whitin, & Prichard, 2010).

Second, daily sleep duration as recorded by commercially available devices (Jawbone UP3) and as reported daily by adolescents are moderately correlated ($r \sim .50$), especially when observing the same-day associations ($\beta \sim .40$). When comparing average sleep, there were weaker correlations between the person-level average amount of sleep reported and recorded. Not surprisingly, there were weaker associations between wearable-recorded sleep duration and subjective sleep quality, indicating more uniqueness in these measures. Although the percent of variance explained ($R^2 \sim 25\%$) may not be as high as expected from “interchangeable” daily measures, overall the data suggests that sleep duration captured via wearables corresponded well with same-day self-reports of hours slept.

Third, there are differences in the associations between sleep measures across days (i.e., weekend vs. week days) and across individuals (i.e., age, gender). Similar to previous studies (Wolfson et al., 2003), objective and subjective sleep measures were more highly associated on weekdays compared to weekends. This discrepancy may be due, in part, to some adolescents (especially older adolescents in college) more flexible, less regulated weekend schedules. If adolescents change or pay less attention to bed and wake times on weekends, they may be more likely to overestimate of the hours slept on weekends compared to weekday. Future studies may further examine weekend sleep patterns or at least account for them in analyses. The similarity in the associations across the studies suggests some reliability of the devices’ estimates. However, there also appeared to be some variation in the associations between sleep measures by individual characteristics in Study 2. Specifically, boys’ sleep measures were more highly correlated than girls and adolescents aged 10–13 years had a stronger association between recorded and reported sleep duration than adolescents aged 14–16. These differences were mostly driven by a much weaker association in recorded and reported sleep by a select group of older adolescent girls. Other studies comparing actigraph and self-report sleep measures have found mixed results regarding gender or age differences—some finding stronger correlations for boys (Tremaine, Dorrian, & Blunden, 2010) or for girls (Short, Gradisar, Lack, Wright, & Carskadon, 2012), some finding stronger correlations with middle-school age adolescents (Tremaine et al., 2010) or with elementary school children (Gaina, Sekine, Chen, Hamanishi, & Kagamimori, 2004). Future studies should explore the strength of the associations between sleep measures further with diverse groups of adolescents.

Fourth, objectively recorded sleep duration via a wearable device is not consistently associated with same-day affect and well-being. Across both studies, longer recorded sleep was only significantly associated with same-day self-esteem in older adolescents. In the younger sample, there were no associations between recorded sleep duration with daily outcomes. Self-reported sleep duration was associated with negative and positive affect and self-esteem in the older adolescent sample and negative affect and dysregulation in the younger sample. It should be noted that many of these associations were very small, especially in the younger adolescent sample (Betas ranged from −.04 to .02), and many failed to remain significant after correcting for false discovery rates. Only subjective sleep quality or problems were consistently, robustly associated with daily well-being.

There are a few possible explanations for the discrepancy in these findings given the correlated estimates of sleep duration. First, commercial wearable devices may not always measure sleep duration as accurately as individuals. A challenge for wearable devices is in defining when sleep starts and stops. It may be difficult for the devices to accurately detect whether a person is sleeping or simply resting in bed (e.g., reading a book before bed). Similarly, the devices often do not count some of the lighter, restless periods of sleeping (i.e., falling asleep, waking up, napping) in the total sleep duration. In addition, upon piloting the devices, we noticed that they would sometimes record about 2 hr and 50 min of sleep even when not worn (~1%–2% of days); hence, the impetus for our wearable data cleaning procedures. Second, the daily associations between self-reported sleep duration and quality and well-being outcomes may reflect shared method variance between the reported methods. Because both sleep and well-being were reported by the same participants on the same day, the correspondence in variance can be due to individuals' and/or daily systematic biases in reporting (e.g., a "good day effect" when participants feel good they report similarly positively across items). Because we did not compare both daily measures to polysomnography standards we cannot conclude which measures capture the "real" associations between sleep duration and well-being based on our results alone. It is tempting to attribute the daily associations between self-reported measures of sleep and well-being only to biases in self-report, but daily self-reported sleep measures have been validated with survey and polysomnography techniques (Rogers et al., 1993). In addition, clinical research has shown subjective sleep quality may continue to be informative, even with a lack of objective corroboration, for diagnosis and treatment of sleep disturbances and mental health symptoms (e.g., Alfano, Patriquin, & De Los Reyes, 2015; Bertocci et al., 2005). Assessing sleep behaviors by commercially available devices and self-reported measures each has advantages and disadvantages and inclusion of both can allow for more comprehensive understanding of the facets of daily sleep health and their associations with well-being.

Suggestions for Future Daily Studies

On the basis of our summed experience from two daily studies with adolescents, we offer practical suggestions for researchers interested in using commercially available wearable devices (for our full list of considerations see https://github.com/kodvinci/wearable_data_collection/blob/master/README.md). First, when possible, researchers should use wearable devices in conjunction with self-reported data. There is utility in capturing self-reported sleep measures for validating wearable-recorded sleep duration and collecting more complete measures of daily sleep health. Commercial devices may not always be able to detect sleep patterns (e.g., bed/wake times) and cannot obtain subjective aspects of sleep (e.g., quality). Researchers should carefully consider the strengths and limitations of each method for assessing sleep and use prior research to identify which components of sleep (and the appropriate methods of gathering that information) are likely to be associated with adolescents' well-being. When possible, daily studies with adolescents should include multiple measures of sleep duration and quality to offset the limitations of all measures.

Second, researchers should fully investigate commercial wearable devices before using them in the field. This ensures that devices were developed by reputable companies and have been tested for reliability and safety. Prior to selecting the Jawbone UP3 for our studies, we piloted a wide range of commercially available and research-grade devices. Common limitations with the research-grade devices included discomfort in wearing the device, low battery efficiency, and difficulty syncing with non-commercially available apps on the participants' devices. Major limitations with the commercial devices included a longer, but still limited, battery life and the need for participants to recharge the device over the course of the study, a lack of transparency in terms of how and for how long data are saved on the devices, and specialized programming that is required to prevent the participant from viewing feedback on sleep or activity from the device. Extensive pilot testing of devices is critical for the identification of issues with the devices or their applications before they are used in the field. Anecdotally, our research team accumulated hundreds of person-days of pilot testing devices, and it was only through this testing that the previously mentioned phenomenon of the wearable recording about 3 hr of sleep when not worn was discovered. However, even after extensive piloting of devices there are risks to investigators launching studies with wearable devices. In the case of Jawbone, the company went out of business following the completion of our study. If the timing of Jawbone's exit from the market would have came earlier, it would have negatively impacted our ability to retrieve study member's sleep data.

Third, researching or fostering communications with companies can aid in understanding the devices and in manipulating the data. Note that most companies do not have open APIs for the algorithms used to calculate daily sleep measures and the amount of information they share with researchers may be limited. In addition, the algorithms used to calculate sleep measures can (and in our experience, often do) change without notice from the manufactures. Therefore, keeping in contact with both participants and companies throughout the data collection process can help identify and resolve problems as they occur.

CONCLUSION

In sum, commercially available wearable devices are an emergent technology that allows researchers to unobtrusively estimate daily sleep patterns in adolescents' everyday lives. These devices have a number of strengths as methodological tools—including ecological validity, objective estimates of sleep duration, reliability across studies, and unobtrusiveness. Researchers must also be aware of their limitations, should be cautious not to over-state findings associated with these devices, and should not be surprised if objectively gathered sleep data do not closely correspond with subjective indicators of adolescents' daily well-being. As these new technologies improve, the ability to more reliably and accurately capture objective features of sleep should lead to important new discoveries about the daily (and nightly!) experiences of adolescents.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported with funding from the National Institute on Drug Abuse Center for the Study on Adolescent Risk and Resilience (C-StARR) (P30DA023026). We thank the C-StARR study team, study participants and their families. JGR received funding from T32 GM007171 as part of the Medical Scientist Training Program. CLO is a Jacobs Foundation Advanced Fellow and a Fellow of the Canadian Institute of Advanced Research.

REFERENCES

Alfano CA, Patriquin MA, & De Los Reyes A (2015). Subjective-objective sleep comparisons and discrepancies among clinically-anxious and healthy children. Journal of Abnormal Child Psychology, 43, 1343–1353. 10.1007/s10802-015-0018-7 [PubMed: 25896729]

Arora T, Broglia E, Pushpakumar D, Lodhi T, & Taheri S (2013). An investigation into the strength of the association and agreement levels between subjective and objective sleep duration in adolescents. PLoS ONE, 8, e72406–e72412. 10.1371/journal.pone.0072406 [PubMed: 23951321]

Beebe D (2011). Cognitive, behavioral, functional consequences of inadequate sleep in children and adolescents. Pediatric Clinics of North America, 58, 649–665. 10.1016/j.pcl.2011.03.002 [PubMed: 21600347]

Bertocci M, Dahl R, Williamson D, Iosif A, Birmaher B, Axelson D, & Ryan N (2005). Subjective sleep complaints in pediatric depression: A controlled study and comparison with EEG measures of sleep and waking. Journal of the American Academy of Child and Adolescent Psychiatry, 44, 1158–1166. 10.1097/01.chi.0000179057.54419.17 [PubMed: 16239865]

Buysse D (2014). Sleep heath: Can we define it? Does it matter? Sleep Health, 37, 9–17. 10.5665/sleep.3298

Carskadon M, Mindell J, & Drake C (2006). 2006 Sleep in America poll: Teens and SLEEP. Washington, DC: National Sleep Foundation.

Chung KF, Kan KKK, & Yeung WF (2011). Assessing insomnia in adolescents: Comparison of Insomnia Severity Index, Athens Insomnia Scale, and Sleep Quality Index. Sleep Medicine, 12, 463–470. 10.1016/j.sleep.2010.09.019 [PubMed: 21493134]

Conner T, & Barrett L (2012). Trends in ambulatory self-report: The role of momentary experience in psychosomatic medicine. Psychosomatic Medicine, 74, 327–337. 10.1097/PSY.0b013e3182546f18 [PubMed: 22582330]

Cook J, Prairie M, & Plante D (2017). Utility of the Fitbit Flex to evaluate sleep in major depressive disorder: A comparison against polysomnography and wrist-worn actigraphy. Journal of Affective Disorders, 217, 299–305. 10.1016/j.jad.2017.04.030 [PubMed: 28448949]

Csikszentmihalyi M, Larson R, & Prescott S (1977). The ecology of adolescent activity and experience. Journal of Youth and Adolescence, 6, 281–294. 10.1007/bf02138940 [PubMed: 24408457]

Dewald JF, Meijer AM, Oort FJ, Kerkhof GA, & Bogels SM (2010). The influence of sleep quality, sleep duration, and sleepiness on school performance in children and adolescents: A meta-analytic review. Sleep Medicine Reviews, 14, 179–189. 10.1016/j.smrv.2009.10.004 [PubMed: 20093054]

de Zambotti M, Baker F, Willoughby A, Godino J, Wing D, Patrick K, & Colrain I (2016). Measures of sleep and cardiac functioning during sleep using a multi-sensory commercially-available wristband in adolescents. Physiology and Behavior, 158, 143–149. 10.1016/j.physbeh.2016.03.006 [PubMed: 26969518]

Eatough E, & Shockley K (2016). A review of ambulatory health data collection methods for employee experience sampling research. Applied Psychology, 65, 322–354. 10.1111/apps.12068

Emsellem H, Knutson K, Hillygus D, Buxton O, Montgomery-Downs H, LeBourgeois M, & Spilsbury J (2014). 2014 Sleep in America Poll: Sleep in the modern family. Arlington, VA: National Sleep Foundation.

Fuligni A, Arruda EH, Krull JL, & Gonzales NA (2017). Adolescent sleep duration, variability, and peak levels of achievement and mental health. Child Development, 89, e18–e28. [PubMed: 28129442]

Fuligni A, & Hardway C (2006). Daily variation in adolescents' sleep, activities, and psychological well-being. Journal of Research on Adolescence, 16, 353–378. 10.1111/j.1532-7795.2006.00498.x

Gaina A, Sekine M, Chen X, Hamanishi S, & Kagamimori S (2004). Sleep parameters recorded by Actiwatch in elementary school children and junior high school adolescents: Schooldays vs weekends. Sleep and Hypnosis, 6, 55–66.

He F, Bixler EO, Liao J, Berg A, Kawasawa YI, Fernandez-Mendoza J, … Liao D (2015). Habitual sleep variability, mediated by nutrition intake, is associated with abdominal obesity in adolescents. Sleep Medicine, 16, 1489–1494. 10.1016/j.sleep.2015.07.028 [PubMed: 26611945]

Iida M, Shrout P, Laurenceau J, & Bolger N (2012). Using diary methods in psychological research. In Cooper H, Camic PM, Long DL, Panter AT, Rindskopf D, & Sher KJ (Eds.), APA Handbook of research methods in psychology, Vol 1: Foundations, planning, measures, and psychometrics (vol. 1, pp. 277–305). Washington, DC: American Psychological Association. 10.1037/13619-000

Janssen KC, Phillipson S, O'Connor J, & Johns MW (2017). Validation of the Epworth Sleepiness Scale for children and adolescents using Rasch analysis. Sleep Medicine, 33, 30–35. 10.1016/j.sleep.2017.01.014 [PubMed: 28449902]

Jawbone (2014). Retrieved from www.jawbone.com Accessed: June 8th, 2018.

Ji X, & Liu J (2016). Subjective sleep measures for adolescents: A systematic review. Child: Care, Health, and Development, 42, 825–839. 10.1111/cch.12376

Johnson NL, Kirchner L, Rosen CL, Storfer-Isser A, Cartar LN, Ancoli-Israel S, … Redline S (2007). Sleep estimation using wrist actigraphy in adolescents with and without slepe disordered breathing: A comparison of three data modes. Sleep, 30, 899–905. 10.1093/sleep/30.7.899 [PubMed: 17682661]

Keyes KM, Maslowsky J, Hamilton A, & Schulenberg J (2015). The great sleep recession: Changes in sleep duration among US adolescents 1991–2012. Pediatrics, 135, 460–468. 10.1542/peds.2014-2707 [PubMed: 25687142]

Lemola S, Ledermann T, & Friedman EM (2013). Variability of sleep duration is related to subjective sleep quality and subjective well-being: An actigraphy study. PLoS ONE, 8, e71292–e71292. 10.1371/journal.pone.0071292 [PubMed: 23967186]

Lund HG, Reider BD, Whitin AB, & Prichard JR (2010). Sleep patterns and predictors of disturbed sleep in a large population of college students. Journal of Adolescent Health, 46, 124–132. 10.1016/j.jadohealth.2009.06.016

Mantua J, Gravel N, & Spencer RMC (2016). Reliability of sleep measures from four personal health monitoring devices compared to research-based actigraphy and polysomnography. Sensors, 16, 646–657. 10.3390/s16050646

Marino M, Li Y, Rueschman M, Winkelman J, Ellen-bogen J, Solet J, … Buxton O (2013). Measuring sleep: Accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. Sleep, 36, 1747–1755. 10.5665/sleep.3142 [PubMed: 24179309]

Matthews K, Hall M, & Dahl R (2014). Sleep in healthy black and white adolescents. Pediatrics, 133, e1189–e1196. 10.1542/peds.2013-2399 [PubMed: 24753532]

Meltzer L, Hiruma L, Avis K, Montgomery-Downs H, & Valentin J (2015). Comparison of a commercial accelerometer with polysomnography and actigraphy in children and adolescents. Sleep, 38, 1323–1330. 10.5665/sleep.4918 [PubMed: 26118555]

Meltzer L, Montgomery-Downs H, Insana SP, & Walsh CM (2012). Use of actigraphy for assessment in pediatric sleep research. Sleep Medicine Reviews, 16, 463–475. 10.1016/j.smrv.2011.10.002 [PubMed: 22424706]

MetricWire Inc (2015). Retrieved from https://www.metricwire.com/ Accessed: June 8th, 2018.

Mindell JA, & Owens JA (2015). A clinical guide to pediatric sleep: Diagnosis and management of sleep problems (3rd ed.). Philadelphia, PA: Wolters Kluwer.

MPlus (2017). [Computer Software] (Version 8). Los Angeles, CA: Muthén & Muthén.

Moretti MM (2003). The Affect Regulation Checklist. Unpublished measure and data. British Columbia, Canada: Simon Fraser University.

Nezlek J (2012). Diary methods. Thousand Oaks, CA: Sage.

Owens J, Adolescent Sleep Working Group, & Committee on Adolescence (2014). Insufficient sleep in adolescents and young adults: An update on causes and consequences. Pediatrics, 134, e921–e932. 10.1542/peds.2014-1696 [PubMed: 25157012]

Podsakoff P, MacKenzie S, Lee J, & Podsakoff N (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. Journal of Applied Psychology, 88, 879–903. 10.1037/0021-9010.88.5.879

Raudenbush S, & Bryk A (2002). Hierachical linear models: Applications and data analysis methods (Vol. 1). Thousand Oaks, CA: Sage.

Rivenbark JG, Copeland WE, Davisson EK, Gassmnan-Pines A, Hoyle RH, Piontak JR, Russell MA, Skinner AT, & Odgers CL (in press). Perceived social status and mental health among young adolescents: Evidence from census data to cell-phones. Developmental Psychology.

Rogers A, Caruso C, & Aldrich M (1993). Reliability of sleep diaries for assessment of sleep/wake patterns. Nursing Research, 42, 368–372. [PubMed: 8247821]

Sadeh A (2011). The role validity of actigraphy in sleep medicine: An update. Sleep Medicine Reviews, 15, 259–267. 10.1016/j.smrv.2010.10.001 [PubMed: 21237680]

Schimmack U (2003). Affect measurement in experience sampling research. Journal of Happiness Studies, 4, 79–106. 10.1023/A:1023661322862

Shiffman S (2009). How many cigarettes did you smoke? Assessing cigarette consumption by global report, time-line follow-back, and ecological momentary assessment Health Psychology, 28, 519–526. 10.1037/a0015197 [PubMed: 19751076]

Shiffman S, Stone A, & Hufford M (2008). Ecological momentary assessment. Annual Review of Clinical Psychology, 4, 1–32. 10.1146/annurev.clinpsy.3.022806.091415

Shocat T, Cohen-Zion M, & Tzischinsky O (2014). Functional consequences of inadequate sleep in adolescents: A systematic review. Sleep Medicine Reviews, 18, 75–87. 10.1016/j.smrv.2013.03.005 [PubMed: 23806891]

Short MA, Gradisar M, Lack LC, Wright H, & Carskadon MA (2012). The discrepancy between actigraphic and sleep diary measures of sleep in adolescents. Sleep Medicine, 13, 378–384. 10.1016/j.sleep.2011.11.005 [PubMed: 22437142]

StataCorp. (2017). Stata statistical software (Version 15). College Station, TX: StataCorp LLC.

Stores G, Fry A, & Crawford C (1998). Sleep abnormalities demonstrated by home polysomnography in teenages with chronic fatigue syndrome. Journal of Psychosomatic Research, 45, 85–91. 10.1016/S0022-3999(98)00024-5 [PubMed: 9720858]

Tangney J, Baumeister R, & Boone A (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. Journal of Personality, 72, 271–322. 10.1111/j.0022-3506.2004.00263.x [PubMed: 15016066]

Tremaine RB, Dorrian J, & Blunden S (2010). Subjective and objective sleep in children and adolescents: Measurement, age, and gender differences. Sleep and Biological Rhythms, 8, 229–238. 10.1111/j.1479-8425.2010.00452.x

Watson D, Clark LA, & Tellegan A (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. Journal of Personality and Social Psychology, 54, 1063–1070. 10.1037/0022-3514.54.6.1063 [PubMed: 3397865]

Wolfson A, & Carskadon M (1998). Sleep schedules and daytime functioning in adolescents. Child Development, 69, 875–887. 10.1111/j.1467-8624.1998.tb06149.x [PubMed: 9768476]

Wolfson A, Carskadon M, Acebo C, Seifer R, Fal-lone G, Labyak S, & Martin J (2003). Evidence for the validity of a sleep habits survey for adolescents. Sleep, 26, 213–216. 10.1093/sleep/26.2.213 [PubMed: 12683482]

Zheng H, Sowers M, Buysse D, Consens F, Kravitz H, Matthews K, …Hall M (2012). Sources of variability in epidemiological studies of sleep using repeated nights of in-home polysomnography: SWAN Sleep Study. Journal of Clinical Sleep Medicine, 8, 87–96. [PubMed: 22334814]

TABLE 1

Older Adolescents' Means (Person-Level) and Standard Deviations and Basic Correlations (Across All Observations in the EMA Period Between Sleep and Well-Being Measures in Study 1.

	Mean **(SD)**	***1***	***2***	***3***	***4***	***5***	***6***	***7***
1. Wearable-recorded sleep duration (hours)	6.41 (1.19)	–						
2. Self-reported sleep duration (hours)	6.32 (1.25)	**0.51****	–					
3. Subjective sleep quality (0–9)	5.63 (1.47)	**0.24****	**0.45****	–				
4. Negative affect (1–5)	2.16 (0.52)	−0.05	−0.12**	−0.26**	–			
5. Positive affect (1–5)	3.30 (0.54)	0.05	0.15**	0.27**	−0.52**	–		
6. Self-esteem (1–5)	3.70 (0.46)	0.08*	0.17**	0.29**	−0.57**	0.64**	–	
7. Dysregulation (0–4)	0.79 (0.60)	0.009	0.009	−0.15**	0.43**	−0.26**	−0.25**	–

$^{\dagger}p < .10$;

$^{*}p < .05$;

$^{**}p < .01$.

TABLE 2

Multilevel Standardized Coefficients (*Beta*) and 95% Confidence Intervals (CI) Showing the Daily Associations Between Sleep and Same-Day Well-Being in Study 1.

	Same-Day Well-being			
	Negative Affect β [95% CI]	**Positive Affect β [95% CI]**	**Self-Esteem β [95% CI]**	**Dysregulation β [95% CI]**
Wearable-recorded sleep duration	−.02 [−0.09, 0.05]	.04 [−0.03, 0.11]	.10* [0.02, 0.17]	−.01 [−0.09, 0.07]
Self-reported sleep duration	**−.14** [−0.21, −0.07]**	.08 [0.01, 0.15]*	**.17** [0.09, 0.24]**	.003 [−0.08, 0.08]
Self-reported sleep quality	**−.17** [−0.25, −0.09]**	**.15** [0.08, 0.22]**	**.21** [0.15, 0.28]**	.06 [−0.15, 0.03]

Notes. Coefficients are bolded if they remain significant after false discovery rate *p*-value correction ($p < .004$). All models control for the person-level associations (i.e., person means) of each predictor.

† $p < .10$;

* $p < .05$;

** $p < .01$.

TABLE 3

Younger Adolescents' Means (Person-Level), Standard Deviations and Basic Correlations (Across All Observations in the EMA Period) Between Sleep and Well-Being Measures in Study 2.

	***Mean* (SD)**	***1***	***2***	***3***	***4***	***5***	***6***	***7***
1. Wearable-recorded sleep duration (hours)	7.11 (1.69)	–						
2. Self-reported sleep duration (hours)	8.48 (1.06)	**.41****	–					
3. Subjective sleep problems (0–1)	0.33 (0.28)	−.07*	**-.07****	–				
4. Negative affect (1–100)	12.87 (12.61)	−.005	−.10**	.23**	–			
5. Positive affect (1–100)	52.65 (17.85)	−.02	.05**	−.18**	−.27**	–		
6. Self-esteem (1–5)	2.93 (0.79)	−.006	.07**	−.27**	−.51**	.49**	–	
7. Dysregulation (0–4)	0.39 (0.58)	−.01	−.08**	.22**	.64**	−.12**	−.33**	–

† $p < .10$;

* $p < .05$;

** $p < .01$.

TABLE 4

Multilevel Standardized Coefficients (*Beta*) and 95% Confidence Intervals (*CI*) Showing the Daily Associations Between Sleep and Same-Day Well-Being in Study 2. All Models Control for the Person-Level Associations (i.e., Person Means) of Each Predictor

	Same-Day Well-Being			
	Negative Affect β [95% CI]	**Positive Affect β [95% CI]**	**Self-Esteem β [95% CI]**	**Dysregulation β [95% CI]**
Recorded sleep duration	−.008 [−0.08, 0.06]	−.02 [−0.09, 0.04]	.003 [−0.07, 0.08]	.003 [−0.07, 0.08]
Self-reported sleep duration	−.04 [−0.07, −0.004]*	.02 [−0.02, 0.05]	.02 [−0.01, 0.06]	−.04 [−0.08, −0.001]*
Subjective sleep problems	**.10** [0.06, 0.14]**	**−.09** [−0.12, −0.06]**	**−.11** [−0.15, −0.07]**	**.10** [0.05, 0.14]**

Notes. Coefficients are bolded if they remain significant after false discovery rate *p*-value correction ($p < .004$). All models control for the person-level associations (i.e., person means) of each predictor.

† $p < .10$;

* $p < .05$;

** $p < .01$.