Validation of a non-wearable device in healthy adults with normal and short sleep durations

David A. Hsiou, B.S.,[1,2] Chenlu Gao, PhD,[3,4] Robert C. Matlock, MD,[5] & Michael K. Scullin, PhD[1]

[1]Baylor University, Department of Psychology and Neuroscience

[2]Baylor College of Medicine, School of Medicine

[3]Division of Sleep and Circadian Disorders, Departments of Medicine and Neurology, Brigham and Women's Hospital

[4]Division of Sleep Medicine, Department of Medicine, Harvard Medical School

[5]Ascension Medical Group Providence, Department of Sleep Medicine

*Corresponding author: Michael K. Scullin, Ph.D., Department of Psychology and Neuroscience, Baylor University, One Bear Place 97334, Waco, TX 76798, Phone: 254-710-2241

Email: Michael_Scullin@Baylor.edu

All authors have read and approved this manuscript for submission.

This study is supported by the National Science Foundation (1920730 and 1943323).

The authors have no conflict of interest to disclose.

Number of tables: 1

Number of figures: 2

Number of supplemental figures: 4

Abstract word count: 228

Brief summary word count: 114

Manuscript word count: 2,534

Abstract

Study Objectives: To determine the accuracy of early and newer versions of a non-wearable sleep tracking device relative to polysomnography and actigraphy, under conditions of normal and restricted sleep duration.

Methods: Participants were 35 healthy adults (M_{age}=18.97, SD=0.95 years; 77.14% female; 42.86% Caucasian). In a controlled sleep laboratory environment, we randomly assigned participants to go to bed at 10:30pm (normal sleep) or 1:30am (restricted sleep), setting lights-on at 7:00am. Sleep was measured using polysomnography, wristband actigraphy (Actiwatch Spectrum Plus), self-report, and an early or newer version of a non-wearable device that uses a sensor strip to measure movement, heart rate, and breathing (Beddit, Apple Inc.). We tested accuracy against polysomnography for total sleep time (TST), sleep efficiency (SE%), sleep onset latency (SOL), and wake after sleep onset (WASO).

Results: The early version of the non-wearable device (Beddit 3.0) displayed poor reliability (ICCs<0.30). However, the newer non-wearable device (Beddit 3.5) yielded excellent reliability with polysomnography for TST (ICC=0.998) and SE% (ICC=0.98) across normal and restricted sleep conditions. Agreement was also excellent for the notoriously difficult metrics of SOL (ICC=0.92) and WASO (ICC=0.92). This non-wearable device significantly outperformed clinical grade actigraphy (ICCs between 0.44 and 0.96) and self-reported sleep measures (ICCs below 0.75).

Conclusions: A non-wearable device demonstrated better agreement with polysomnography than actigraphy. Future work is needed to test the validity of this device in clinical populations.

Keywords: consumer sleep technology, non-wearable device, sleep tracking

Brief Summary

Current Knowledge/Study Rationale: Polysomnography is the gold standard for clinical sleep assessment and diagnosis, but this method is too cumbersome and costly for long term monitoring of natural sleep patterns. This study compares the accuracy of sleep metrics collected from a non-wearable device, Apple, Inc.'s Beddit Sleep Monitor (versions 3.0 and 3.5), against industry standards such as polysomnography, actigraphy, and self-report.

Study Impact: While many commercial sleep devices have displayed poor accuracy, the current findings show that at least one recent non-wearable device has excellent accuracy in healthy adults. Reducing the burden on participants in sleep studies (via non-wearable devices) has broad implications for improving accessibility, monitoring duration, study costs, and representativeness of findings.

Introduction

Polysomnography (PSG) is the gold standard measurement of sleep. In research settings, PSG allows for precise measurement of sleep stages and detection of sleep microevents (utilizing EEG monitoring). In clinical settings, PSG enables diagnoses and therapeutic testing (e.g., positive airway pressure titration). Despite its strengths, PSG has some disadvantages such as being expensive and burdensome on patients and participants. In addition, sleeping in a laboratory is known to detrimentally affect sleep quality[1] and key sleep measures can fluctuate from night to night, limiting the value of PSG if only a single night can be obtained.[2] One alternative (or complementary) approach is to use wristband actigraphy, in which sleep/wake state is estimated via accelerometry in home settings.[3] Actigraphy has acceptable levels of accuracy for total sleep time,[4] but the cost of clinical-grade actigraphy devices remains high and some individuals do not like to sleep while wearing devices (some special populations, such as children with autism, do not tolerate actigraphy or PSG).[5] Therefore, there is a need for devices that can inexpensively and accurately measure sleep in the home environment with minimal or zero contact to the person ("non-wearables").

As public interest in sleep health has surged,[6-7] there has been a proliferation of commercial devices to track sleep.[8] However, when these commercial devices are tested relative to PSG and actigraphy, many of them show poor congruence. For example, Meltzer et al.[9] found that the Fitbit Ultra either overestimated or underestimated total sleep time by greater than 1 hour, depending on whether the normal or sensitive analysis mode was used. Similar inaccuracies in device measurement have been reported in other independent tests of commercial sleep tracking devices.[10-12] Nevertheless, the sleep tracking consumer market has continued a compound annual growth rate of 18.5%,[13] and now has attracted leading technology companies such as Apple, Samsung, Bose, and others.

In the current study, we evaluated the accuracy of a non-wearable device, the Beddit Sleep Monitor (hereafter, Beddit). Beddit is a flat strip sensor that is placed under the bedsheet. Apple, Inc. acquired this start-up company in 2017 (version 3.0) and released an updated version of Beddit in

December, 2018 (version 3.5). At the time of acquisition, Beddit 3.0 demonstrated poor accuracy relative to PSG for total sleep time (TST), wake after sleep onset (WASO), and sleep efficiency (SE%).[12] The current study sought to replicate these findings for Beddit 3.0 and also test the accuracy of the subsequent Beddit version 3.5 against PSG (gold standard), as well as actigraphy and self-report. Because length and quality of sleep have been known to affect the accuracy of some sleep monitoring devices,[14] we randomly assigned participants to 8.5-hour or 5.5-hour time-in-bed opportunities .[15] Thus, the current work provides a comprehensive comparison of research-grade, commercial, and subjective measures of sleep across normal and restricted sleep schedules.

Methods

Participants

Thirty-five adults (M_{age} = 18.97, *SD* = 0.95 years; 77.14% female; 42.86% Caucasian, 14.29% Hispanic, 14.29% Asian, 5.71% African American, 20% other/multiple races) were recruited through advertisements at a university in the United States. All participants provided informed consent and completed the study between October 2019 and March 2020. Data collection ceased in response to the coronavirus pandemic. This study was approved by the Baylor University Institutional Review Board.

Sleep Measurements

Polysomnography (PSG). We used the Grass Comet XL Plus system to record sleep from electroencephalography (Fp1, Fp2, F3, F4, F7, F8, FZ, C3, C4, CZ, P3, P4, PZ, T3, T4, T5, T6, O1, and O2 sites; referenced to contralateral mastoids), bilateral electrooculography, and mentalis electromyography. Sleep stages were scored in 30-second epochs according to AASM guidelines by a registered polysomnography technician who was masked to participants' Beddit, actigraphy, and self-report data.[16]

Actigraphy. For actigraphy monitoring, we used Philips Respironics Actiwatch Spectrum Plus devices (Bend, Oregon, USA). Relative to PSG, actigraphy tends to show modest specificity,[14] but strong sensitivity to sleep states.[17-19] Participants wore the actigraphy device on their non-dominant wrist throughout the night. We set the devices to medium-threshold sensitivity (40 activity counts per epoch) and recorded in 30-second epoch lengths. Trained personnel then scored the data based on the scoring technique developed by Rijsketic and colleagues.[20]

Beddit. The Beddit Sleep Monitor device is a thin, flat sensor strip placed immediately under the bedsheet at chest level. Beddit records movement, heart rate, and breathing rate and provides automated estimates of time in bed, total sleep time, sleep onset latency, wake after sleep onset, time away from bed, sleep efficiency, average breathing rate, and average heart rate (as well as maximum and minimum). Automated estimates are provided through Beddit's iOS-supported companion applications. Two successive Beddit versions were tested – Beddit 3.0 (released October 2016) and Beddit 3.5 (released December 2018). The Beddit device was connected to an Apple iPhone or iPad via Bluetooth technology. To avoid potential cross-device interference, only one Beddit version was used for each participant (*ns* = 21 and 14, for versions 3.0 and 3.5, respectively).

Self-report. Upon awakening in the morning, participants estimated their sleep latency, total sleep time, and wake after sleep onset. Previous validation studies demonstrated moderate correlations (r=0.45) between self-report and objective sleep measures.[21] Additionally, participants completed the Pittsburgh Sleep Quality Index (PSQI)[22] to assess subjective sleep quality during the past month.

Procedures

Participants arrived at a light- and sound-controlled sleep lab around 8:00pm. They completed a series of questionnaires, including the PSQI, while research personnel set up the Beddit devices, applied PSG electrodes, and configured the actigraphy device. Participants were randomly assigned to go to bed at 10:30pm (*n*=19) or 1:30am (*n*=16). Following random assignment, participants were allowed to sit at a desk and complete homework, use their phones, and watch television until 15 minutes prior to their scheduled bedtime. Biocalibration testing was conducted at the scheduled bedtime and was immediately followed by lights-out. Lights-on occurred between 6:50am and 7:10am, depending on natural awakening or a switch to N1 sleep. Upon awakening, participants reported their total sleep time, sleep latency, and wake after sleep onset, prior to completing learning tasks that addressed a separate research question.

Statistical analyses

First, we used two-way repeated measures ANOVA (device × sleep condition) to assess whether Beddit-measured bedtime, SOL, WASO, TST, and SE% differed from PSG-measured values across normal sleep and restricted sleep schedules. We focused on the main effect of device and interaction effects. We also conducted Pearson's correlation coefficients to quantify the relationships between sleep measures. Importantly, to quantify the level of consistency between measures, we used intraclass correlation coefficient (ICC), which is a standard measure of inter-rater reliability that is suitable for ratio variables and for studies with two or more raters (or, in the current study, two or more devices).[23] ICC values were computed through two-way mixed effect models that examined the consistency between measures and estimated the reliability of a single measure. We interpreted ICC values according to Koo and Li's guidelines,[23] in which 0.50 indicates moderate agreement, 0.75 indicates good agreement, and 0.90 indicates excellent agreement. We then computed and examined the 95% confidence intervals of ICCs to determine whether reliability differed significantly across devices. We plotted the consistency between measures on Bland-Altman plots.[24] On Bland-Altman plots, the mean difference represents systematic/bias error and the 95% confidence intervals represent random/precision error. Because correlation measures (e.g., ICC, Pearson's correlation) can be influenced by restriction of range, we combined the normal and restricted sleep conditions in ICC and correlational analyses when the initial ANOVA showed no significant device by condition interactions.[25] Note that neither Beddit 3.0 nor Beddit 3.5 provided epoch-by-epoch sleep/wake data, and therefore we were unable to examine sensitivity and specificity outcome measures. All analyses were performed using SPSS software (version 27). All statistical tests were two-tailed, with alpha set to .05. Study data are publicly accessible at https://osf.io/vg345/.

Results

Overview. Four actigraphy reports and one self-report sleep estimate were missing or incomplete. Therefore, the final dataset includes PSG for 35 participants, Beddit data for 35 participants, actigraphy data for 31 participants, and self-report data for 34 participants.

Bedtime. We first investigated whether the actigraphy and Beddit devices would accurately capture participant bedtimes, which were experimentally assigned. For this initial measure, both actigraphy and Beddit 3.5 were successful (Table 1). Figure 1A demonstrates that Beddit 3.5 had excellent reliability (ICC > 0.99), and the Bland-Altman plot in Supplemental Figure 1 demonstrates minimal bias error (<5 min mean deviation; 95%CI -11.79, 2.00), relative to PSG.

Total Sleep Time (TST). Table 1 shows that TST was more difficult (than bedtime) to estimate for all non-PSG methods. Beddit 3.0 correlated weakly (r=.38, p=.089), and self-report correlated moderately (r=.58, p<.001) with PSG; however, the TST correlation with PSG was nearly perfect for actigraphy (r=.96, p<.001) and Beddit 3.5 (r=.998, p<.001; Supplemental Table 1). The ICCs showed a similar pattern. Beddit 3.5 and actigraphy demonstrated excellent reliability relative to PSG in measuring TST (ICCs>.95), significantly outperforming Beddit 3.0 and self-reported TST (ps <.05). Figure 2 demonstrates that, relative to PSG, the raw bias error for Beddit 3.5 was less than 2 minutes (95%CI: -12.93, 16.08), which was superior to the Beddit 3.0 deviation of 45 minutes. Similar outcomes for both Beddit versions were observed across the normal sleep and restricted sleep conditions (Table 1).

Sleep Efficiency (SE%). The SE% data mirrored the TST findings. Beddit 3.0 SE% correlated weakly with PSG SE% (r=.34, p=.133; Supplemental Table 2); by contrast, Beddit 3.5 correlated nearly perfectly with PSG for SE% (r=.99, p<.001) and showed only a 1.1% bias error (95%CI: -5.19, 3.00; Supplemental Figure 2). Self-report (r=.45, p=.008) and actigraphy (r=.81, p<.001) estimates of SE% also correlated with PSG estimates, albeit not as highly as Beddit 3.5 (r=.99, p<.001). Relative to PSG, SE% ICC values for the Beddit 3.5 (ICC=0.98; Figure 1B) were significantly better than the levels of agreement with PSG for Beddit 3.0 or actigraphy (ps<.05; Table 1). Self-reported SE% was more than 5% lower (SD = 15.35%) than PSG-measured SE% and showed poor levels of agreement (ICC=0.38).

Sleep Onset Latency (SOL). The early version of the non-wearable device (Beddit 3.0) underestimated SOL by 13 minutes (95%CI: -58.99, 32.13), and showed essentially zero reliability (ICC=0.02), relative to PSG (Table 1; Supplemental Figure 3). By contrast, Figure 1c shows that the newer version of the non-wearable device (Beddit 3.5) demonstrated excellent reliability with PSG

(ICC=0.92) with only 6 minutes of bias error (95%CI: -5.49, 18.28; Supplemental Figure 3). Actigraphy and self-report performed less well, demonstrating moderate-to-good reliability (ICC=0.74) and poor reliability (ICC=0.37), respectively, relative to PSG.

Wake After Sleep Onset (WASO). WASO often shows the weakest levels of agreement between sleep tracking devices and PSG. Indeed, Beddit 3.0 displayed very poor reliability for WASO (ICC=-0.02), overestimating WASO duration by 25 minutes relative to PSG (95%CI: -89.90, 141.23; Table 1; Supplemental Figure 4). Interestingly, Figure 1D illustrates that the newer Beddit 3.5 device provided excellent reliability for WASO relative to PSG (ICC=0.92), though the bias error remained high at 22 minutes (95%CI: -7.47, 50.90; Supplemental Figure 4). In other words, Beddit 3.5 overestimated WASO for all participants, but its overestimations were very consistent in relation to PSG measured WASO variability. Actigraphy and self-report showed a different pattern: they were just as likely to overestimate or underestimate WASO for a given individual, which led to relatively low average bias error (actigraphy: *M*=-3.19; 95%CI: -60.52, 54.14; self-report: *M*=5.47; 95%CI: -52.57, 63.51), but the ICC values revealed these measures' weak to modest reliability (actigraphy: ICC=0.44; self-report: ICC=0.72).

Discussion

Previous work found that an early non-wearable device, Beddit 3.0, yielded poor accuracy across measures of sleep (SOL, WASO, TST, and SE%).[12] We replicated these findings. The newer version of Beddit (3.5), however, distinguished itself from its earlier version with strong agreement to PSG. Compared to PSG, the mean bias (and 95% CIs) of Beddit 3.5 was within the field's standards for TST (threshold: 40 minutes), SE% (threshold: 5%), SOL (threshold: 30 minutes), and WASO (threshold: 30 minutes).[26] Importantly, Beddit 3.5 even demonstrated greater agreement with PSG than actigraphy (and self-report) on measures of SE%, SOL, and WASO, and agreement levels were consistent across longer and shorter time-in-bed opportunities (i.e., the two sleep conditions). Therefore, at least in healthy adults, Beddit 3.5 is a valid measure of TST, SE%, SOL, and WASO.

The current findings with Beddit 3.5 add to the developing literature on measuring sleep using minimal or no-contact sensors.[27] Such non-wearable devices estimate sleep by detecting movement, heart rate, and/or respiration using radiofrequency waves (e.g., ResMed S+,[28] impulse radio ultra-wideband technology (Somnofy),[29] cardioballistic sensors (Beddit,[12] and piezoelectric sensors placed under the mattress (EarlySense.[30] In general, these non-wearables have shown high sensitivity (detecting sleep state), modest specificity (detecting wake state), and poor sleep staging relative to PSG.[27] In some cases, however, commercial sleep tracking devices have demonstrated equal or better accuracy than wristband actigraphy for TST.[31] The current findings add to this literature that the non-wearable Beddit 3.5 outperforms wristband actigraphy under multiple sleep schedules (8.5 hour and 5.5 hour time-in-bed opportunities). Like other wearable and non-wearable devices, Beddit 3.5 has difficulty precisely measuring WASO (overestimates by 22 minutes), but at least for Beddit 3.5, its levels of overestimation were consistent on across participants. Therefore, the Beddit 3.5 WASO estimate would be appropriate for research studies focused on inter-individual variability in WASO, but inappropriate for studies in which absolute threshold values must be identified.

This study has limitations. First, the single-center design and limited study population restricts data generalization. All participants were healthy, young adults that presented with normal sleep patterns and behaviors. Thus, these findings might fail to be replicated in other populations, such as those with increased variability in sleep or frequent disturbances, such as the elderly or critically-ill patients. Second, the Beddit 3.5 device does not provide epoch-by-epoch data, which is necessary to calculate sensitivity and specificity. Third, because Apple, Inc. has not shared what was changed across Beddit devices, it remains unclear whether changes in firmware, device sensitivity, algorithms, or some other modification was responsible for the improved accuracy. Fourth, this validation study was conducted in a controlled sleep laboratory, and it remains to be tested whether Beddit 3.5 is equally accurate when there are bedpartners or pets who sleep in bed, as commonly occurs in home settings.

In conclusion, some commercial technology, including the non-wearable Beddit 3.5, has reached acceptable levels of accuracy for sleep research in healthy adults. This demonstrates a clear advancement over the state of the commercial sleep tracking field just 6-8 years ago.[9-12] Future studies on Beddit 3.5 and other devices will need to determine if the high levels of accuracy are retained in diverse groups and clinical populations. Furthermore, future work will need to determine if these devices can be used accurately outside of a sleep laboratory (e.g., home settings, in-patient settings). If the validity of commercial sleep tracking devices are upheld in such contexts that will have broad implications for sleep research, as well as clinical applications. For example, in clinical settings, non-wearable devices could be utilized in conjunction with traditional sleep testing to inform circadian rhythm and insomnia disorders that are dominant in home settings. Such long-term, low cost-monitoring would provide depth to the clinical picture, while avoiding first-night effects and other disturbances to sleep quality that are commonly observed with PSG and wearable sleep monitors.

Abbreviations:

AASM: American Academy of Sleep Medicine

CAGR: compound annual growth rate

CI: confidence interval

EEG: electroencephalography

ICC: intraclass correlation coefficient

PSG: polysomnography

PSQI: Pittsburgh Sleep Quality Index

SE: sleep efficiency

SOL: sleep onset latency

SPSS: Statistical Package for the Social Sciences (software)

TST: total sleep time

WASO: wake after sleep onset

References

1. Agnew HW, Webb WB, Williams RL. The first night effect: An EEG study of sleep. *Psychophysiology*. 1966;2(3):263-266. doi:https://doi.org/10.1111/j.1469-8986.1966.tb02650.x

2. Alshaer H, Ryan C, Fernie GR, Bradley TD. Reproducibility and predictors of the apnea hypopnea index across multiple nights. *Sleep Sci*. 2018;11(1):28-33. doi:10.5935/1984-0063.20180007

3. Stone KL, Ancoli-Israel S. Actigraphy. In: Principles and practice of sleep medicine. 6th ed. Elsevier; 2016:1671-1678.

4. Conley S, Knies A, Batten J, et al. Agreement between actigraphic and polysomnographic measures of sleep in adults with and without chronic conditions: A systematic review and meta-analysis. *Sleep Med Rev.* 2019;46:151-160.

5. Adkins KW, Goldman SE, Fawkes D, Surdyka K, Wang L, Song Y, Malow BA. A pilot study of shoulder placement for actigraphy in children. *Behav Sleep Med*. 2012;10(2):138-147.

6. Perry GS, Patil SP, Presley-Cantrell LR. Raising awareness of sleep as a healthy behavior. *Prev Chronic Dis*. 2013;10. doi:10.5888/pcd10.130081

7. Buysse DJ. Sleep health: Can we define it? Does it matter? *Sleep*. 2014;37(1):9-17. doi:10.5665/sleep.3298

8. Shelgikar AV, Anderson PF, Stephens MR. Sleep tracking, wearable technology, and opportunities for research and clinical care. *Chest*. 2016;150(3):732-743. doi:10.1016/j.chest.2016.04.016

9. Meltzer LJ, Hiruma LS, Avis K, Montgomery-Downs H, Valentin J. Comparison of a commercial accelerometer with polysomnography and actigraphy in children and adolescents. *Sleep*. 2015;38(8):1323-1330. doi:10.5665/sleep.4918

10. Liu W, Ploderer B, Hoang T. In bed with technology: Challenges and opportunities for sleep tracking. In: *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction*. Parkville VIC Australia: ACM; 2015:142-151. doi:10.1145/2838739.2838742

11. Jeon L, Finkelstein J. Consumer sleep tracking devices: A critical review. *Stud Health Technol Inform*. 2015;210:458-460. doi:10.3233/978-1-61499-512-8-458

12. Tuominen J, Peltola K, Saaresranta T, Valli K. Sleep parameter assessment accuracy of a consumer home sleep monitoring ballistocardiograph Beddit Sleep Tracker: A validation study. *J Clin Sleep Med*. 2019;15(3):483-487. doi:10.5664/jcsm.7682

13. Smart Sleep Tracking Device: Market Size, Demand, Growth & Forecast 2024. Accessed at: https://www.researchnester.com/reports/smart-sleep-tracking-device-market/373.

14. Sadeh A. The role and validity of actigraphy in sleep medicine: An update. *Sleep Med Rev*. 2011;15(4):259-267. doi:10.1016/j.smrv.2010.10.001

15. Chee NI, Ghorbani S, Golkashani HA, Leong RL, Ong JL, Chee MW. Multi-night validation of a sleep tracking ring in adolescents compared with a research actigraph and polysomnography. *Nat Sci Sleep*. 2021;*13*:177-190.

16. Iber C, Ancoli-Israel S, Chesson AL, Quan SF. The new sleep scoring manual–The evidence behind the rules. *J Clin Sleep Med*. 2007;03(02):107-107. doi:10.5664/jcsm.26812

17. Kushida CA, Chang A, Gadkary C, Guilleminault C, Carrillo O, Dement WC. Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. *Sleep Med*. 2001;2(5):389-396. doi:10.1016/S1389-9457(00)00098-8

18. Lichstein KL, Stone KC, Donaldson J, et al. Actigraphy validation with insomnia. *Sleep*. 2006;29(2):232-239. doi:10.1093/sleep/29.2.232

19. Morgenthaler T, Alessi C, Friedman L, et al. Practice parameters for the use of actigraphy in the assessment of sleep and sleep disorders: An update for 2007. *Sleep*. 2007;30(4):519-529. doi:10.1093/sleep/30.4.519

20. Rijsketic, JM, Dietch JR, Wardle-Pinkston S, Taylor DJ. *Actigraphy (Actiware) Scoring Hierarchy Manual.* 2020. Retrieved from insomnia.arizona.edu/actigraphy.

21. Lauderdale DS, Knutson KL, Rathouz PJ, Yan LL, Hulley SB, Liu K. Cross-sectional and longitudinal associations between objectively measured sleep duration and body mass index: The CARDIA sleep study. *Am J Epidemiol*. 2009;170(7):805-813. doi:10.1093/aje/kwp230

22. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. *Psychiatry Res*. 1989;28(2):193-213. doi:10.1016/0165-1781(89)90047-4

23. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155-163. doi:10.1016/j.jcm.2016.02.012

24. Altman DG, Bland JM. Measurement in medicine: The analysis of method comparison studies. *Journal of the Royal Statistical Society: Series D (The Statistician)*. 1983;32(3):307-317.

25. Sackett PR, Laczo RM, Arvey RD. The effects of range restriction on estimates of criterion interrater reliability: Implications for validation research. *Personnel Psychol*. 2002;55(4):807-825. doi:10.1111/j.1744-6570.2002.tb00130.x

26. Smith MT, McCrae CS, Cheung J, Martin JL, Harrod CG, Heald JL, Carden KA. Use of actigraphy for the evaluation of sleep disorders and circadian rhythm sleep-wake disorders: An American Academy of Sleep Medicine systematic review, meta-analysis, and GRADE assessment. *J Clin Sleep Med*. 2018;14(7),1209-1230.

27. Lujan MR, Perez-Pozuelo I, Grandner MA. Past, present, and future of multisensory wearable technology to monitor sleep and circadian rhythms. *Front Dig Health*. 2021;3:721919.

28. Schade MM, Bauer CE, Murray BR, Gahan L, Doheny EP, Kilroy H, Zaffaroni A, Montgomery-Downs HE. Sleep validity of a non-contact bedside movement and respiration-sensing device. *J Clin Sleep Med*. 2019;15(7):1051-1061.

29. Toften S, Pallesen S, Hrozanova M, Moen F, Grønli J. Validation of sleep stage classification using non-contact radar technology and machine learning (Somnofy®). *Sleep Med*. 2020;75:pp.54-61.

30. Tal A, Shinar Z, Shaki D, Codish S, Goldbart A. Validation of contact-free sleep monitoring device with comparison to polysomnography. *J Clin Sleep Med*. 2017;13(3):517-522.

31. Chinoy ED, Cuellar JA, Huwa KE, Jameson JT, Watson CH, Bessman SC, Hirsch DA, Cooper AD, Drummond SP, Markwald RR. Performance of seven consumer sleep-tracking devices compared with polysomnography. *Sleep*. 2021;44(5):zsaa291.

Disclosure statement:
This was not an industry supported study. The authors declare no conflicts of interest.

Table 1. PSG in Comparison to Beddit, Actigraphy, and Self-Report Measures of Bedtime, Total Sleep Time, Sleep Efficiency, Sleep Onset Latency, and Wake After Sleep Onset

	PSG (*M, SD*)	Comparison device (*M, SD*)	Difference with PSG (*M, SD*)	Agreement with PSG (ICC, 95% CI)	Main Effect of Device	Device by Sleep Condition Interaction
Bedtime						
Beddit 3.5 (*n*=14)	00:05 am (102.93 min)	00:00 am (102.14 min)	-4.90 (3.52) min	0.999 [0.998, >0.999]	$F(1,12)=24.88$, $p<.001$, $\eta_p^2=0.68$***	$F(1,12)=0.82$, $p=.383$, $\eta_p^2=0.06$
Actigraphy (*n*=31)	11:30 pm (96.83 min)	11:30 pm (95.66 min)	0.68 (6.38) min	0.998 [0.995, 0.999]	$F(1,29)=0.10$, $p=.755$, $\eta_p^2=0.003$	$F(1,29)=1.36$, $p=.254$, $\eta_p^2=0.05$
Total Sleep Time (min)						
Beddit 3.5 (*n*=14)	375.07 (115.63)	376.64 (115.47)	1.57 (7.40)	0.998 [0.994, 0.999]	$F(1,12)=0.48$, $p=.502$, $\eta_p^2=0.04$	$F(1,12)=0.24$, $p=.633$, $\eta_p^2=0.02$
Beddit 3.0 (*n*=21)	399.40 (85.42)	444.81 (487.65)	45.40 (462.02)	0.13 [-0.31, 0.52]	$F(1,19)=0.05$, $p=.831$, $\eta_p^2=0.002$	$F(1,19)=0.85$, $p=.367$, $\eta_p^2=0.04$
Actigraphy (*n*=31)	400.34 (98.76)	398.16 (99.82)	-2.18 (28.32)	0.96 [0.92, 0.98]	$F(1,29)=0.29$, $p=.598$, $\eta_p^2=0.01$	$F(1,29)=0.29$, $p=.594$, $\eta_p^2=0.01$
Self-report (*n*=34)	384.65 (94.50)	357.35 (134.41)	-27.29 (110.24)	0.55 [0.27, 0.75]	$F(1,32)=1.93$, $p=.174$, $\eta_p^2=0.06$	$F(1,32)=0.45$, $p=.508$, $\eta_p^2=0.01$
Sleep Efficiency (%)						
Beddit 3.5 (*n*=14)	89.45 (12.21)	88.36 (11.43)	-1.10 (2.09)	0.98 [0.95, 0.995]	$F(1,12)=3.24$, $p=.097$, $\eta_p^2=0.21$	$F(1,12)=0.88$, $p=.366$, $\eta_p^2=0.07$
Beddit 3.0 (*n*=21)	89.27 (7.08)	88.24 (15.49)	-1.04 (14.69)	0.26 [-0.19, 0.61]	$F(1,19)=0.34$, $p=.564$, $\eta_p^2=0.02$	$F(1,19)=1.29$, $p=.270$, $\eta_p^2=0.06$
Actigraphy (*n*=31)	88.91 (9.64)	87.94 (8.80)	-0.97 (5.71)	0.81 [0.64, 0.90]	$F(1,29)=1.20$, $p=.282$, $\eta_p^2=0.04$	$F(1,29)=0.65$, $p=.428$, $\eta_p^2=0.02$
Self-report (*n*=34)	89.35 (9.44)	83.83 (17.05)	-5.52 (15.35)	0.38 [0.05, 0.63]	$F(1,32)=4.24$, $p=.048$, $\eta_p^2=0.12$*	$F(1,32)=0.002$, $p=.960$, $\eta_p^2<0.001$
Sleep Onset Latency (min)						
Beddit 3.5 (*n*=14)	16.96 (13.77)	23.36 (16.97)	6.39 (6.06)	0.92 [0.78, 0.98]	$F(1,12)=14.24$, $p=.003$, $\eta_p^2=0.54$**	$F(1,12)=0.02$, $p=.890$, $\eta_p^2=0.002$
Beddit 3.0 (*n*=21)	22.86 (21.52)	9.43 (9.30)	-13.43 (23.24)	0.02 [-0.41, 0.44]	$F(1,19)=5.21$, $p=.034$, $\eta_p^2=0.22$*	$F(1,19)=3.75$, $p=.068$, $\eta_p^2=017$
Actigraphy (*n*=31)	21.94 (19.45)	17.52 (23.41)	-4.42 (15.62)	0.74 [0.52, 0.86]	$F(1,29)=1.65$, $p=.209$, $\eta_p^2=0.05$	$F(1,29)=1.30$, $p=.264$, $\eta_p^2=0.04$
Self-report (*n*=34)	20.07 (18.91)	27.21 (14.96)	7.13 (19.19)	0.37 [0.04, 0.62]	$F(1,32)=4.87$, $p=.035$, $\eta_p^2=0.13$*	$F(1,32)=0.76$, $p=.390$, $\eta_p^2=0.02$
Wake After Sleep Onset (min)						
Beddit 3.5 (*n*=14)	25.86 (34.71)	47.57 (40.51)	21.71 (14.89)	0.92 [0.78, 0.97]	$F(1,12)=32.90$, $p<.001$, $\eta_p^2=0.73$***	$F(1,12)=1.72$, $p=.214$, $\eta_p^2=0.13$
Beddit 3.0 (*n*=21)	27.24 (34.91)	52.90 (46.68)	25.67 (58.96)	-0.02 [-0.44, 0.40]	$F(1,19)=6.25$, $p=.022$, $\eta_p^2=0.25$*	$F(1,19)=3.49$, $p=.077$, $\eta_p^2=0.16$
Actigraphy (*n*=31)	28.24 (36.01)	25.05 (15.61)	-3.19 (29.25)	0.44 [0.11, 0.69]	$F(1,29)=0.49$, $p=.489$, $\eta_p^2=0.02$	$F(1,29)=0.26$, $p=.612$, $\eta_p^2=0.01$
Self-report (*n*=34)	26.51 (34.82)	31.99 (44.00)	5.47 (29.61)	0.72 [0.51, 0.85]	$F(1,32)=1.17$, $p=.288$, $\eta_p^2=0.04$	$F(1,32)=0.12$, $p=.734$, $\eta_p^2=0.004$

PSG=Polysomnography; ICC=Intraclass Correlation Coefficient.

*$p<.05$, **$p<.01$, ***$p<.001$.

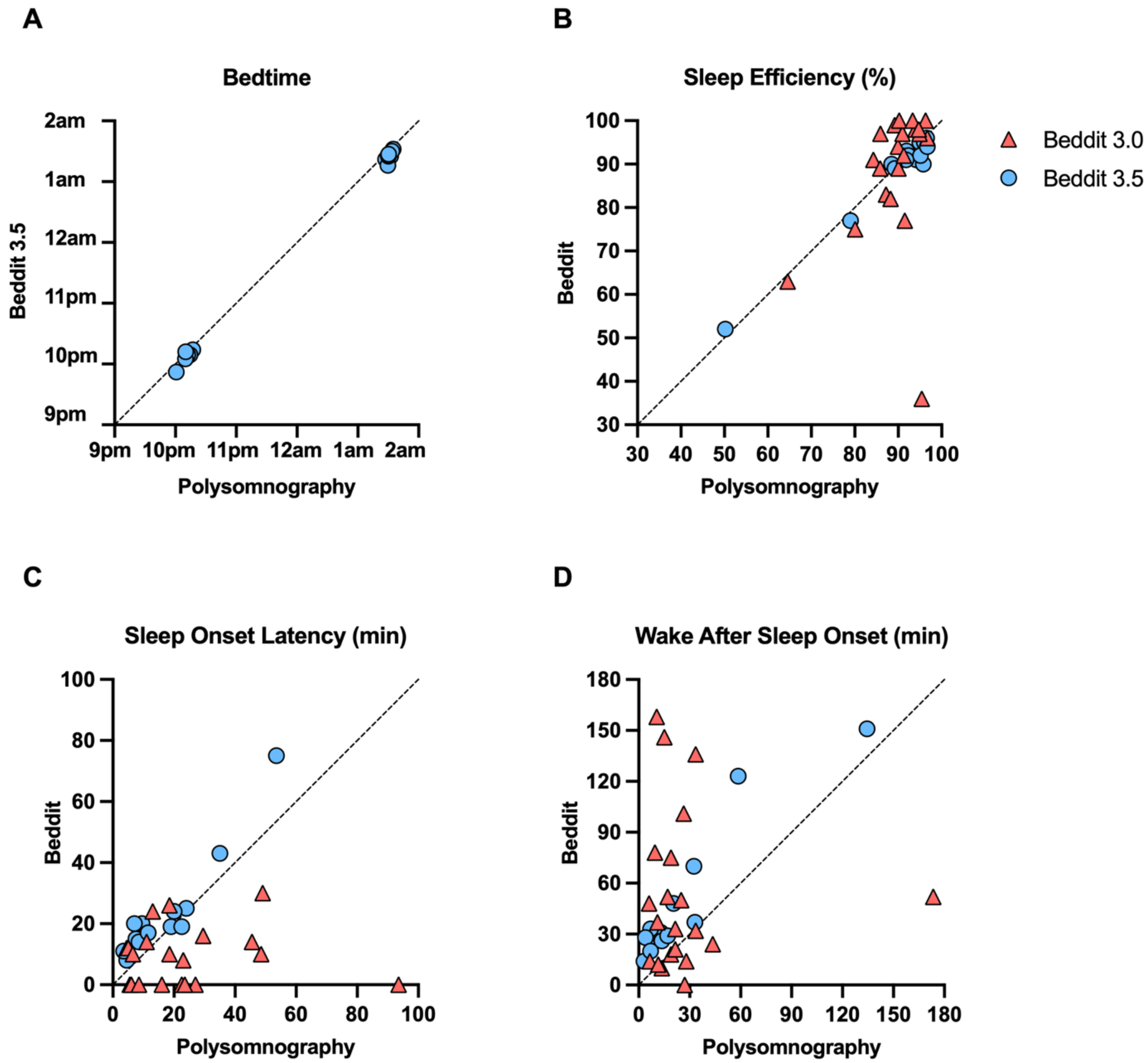


Figure 1. Comparisons of Beddit and PSG estimates of bedtime (ICC>0.99 for Beddit 3.5; A), sleep efficiency (ICC=0.98 for Beddit 3.5, ICC=0.26 for Beddit 3.0; B), sleep onset latency (ICC=0.92 for Beddit 3.5, ICC=0.02 for Beddit 3.0; C), and wake after sleep onset (ICC=0.92 for Beddit 3.5, ICC=-0.02 for Beddit 3.0; D). The diagonal lines indicate perfect agreement between Beddit and PSG.

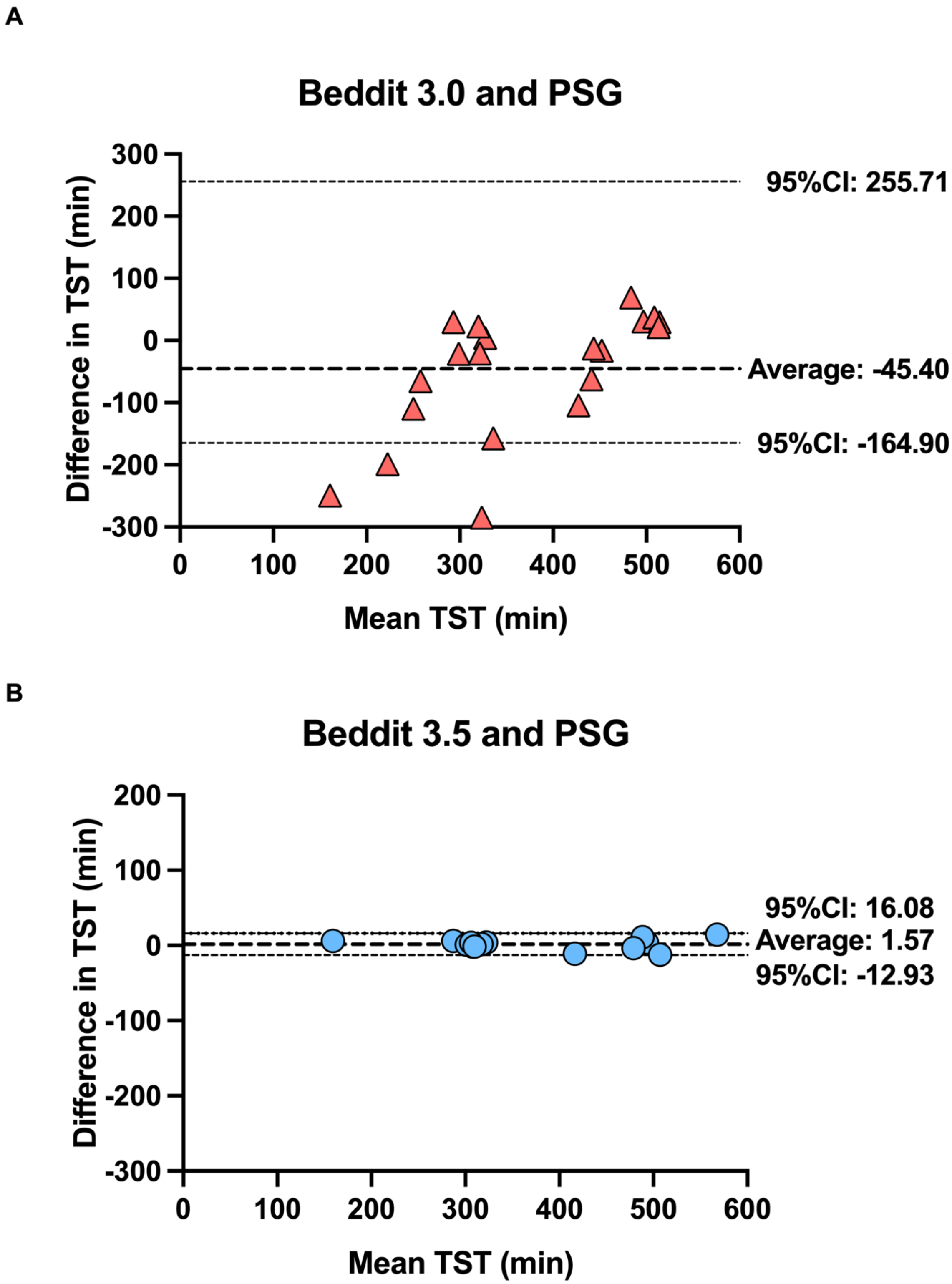


Figure 2. Bland-Altman plots for total sleep time (TST) between Beddit 3.0 and PSG (A), and Beddit 3.5 and PSG (B). Note that one outlier with TST > 40 hours was excluded in (A).

Validation of a non-wearable device in healthy adults with normal and short sleep durations

David A. Hsiou, B.S.,[1,2] Chenlu Gao, PhD,[3,4] R. Christopher Matlock,[5] & Michael K. Scullin, PhD[1]

[1]Baylor University, Department of Psychology and Neuroscience

[2]Baylor College of Medicine

[3]Division of Sleep and Circadian Disorders, Departments of Medicine and Neurology, Brigham and Women's Hospital

[4]Division of Sleep Medicine, Department of Medicine, Harvard Medical School

[5]Ascension Providence, Department of Sleep Medicine

*Corresponding author: Michael K. Scullin, Ph.D., Department of Psychology and Neuroscience, Baylor University, One Bear Place 97334, Waco, TX 76798, Phone: 254-710-2241

Email: Michael_Scullin@Baylor.edu

Supplemental Table 1. Bivariate Correlations among Different Measures for Total Sleep Time

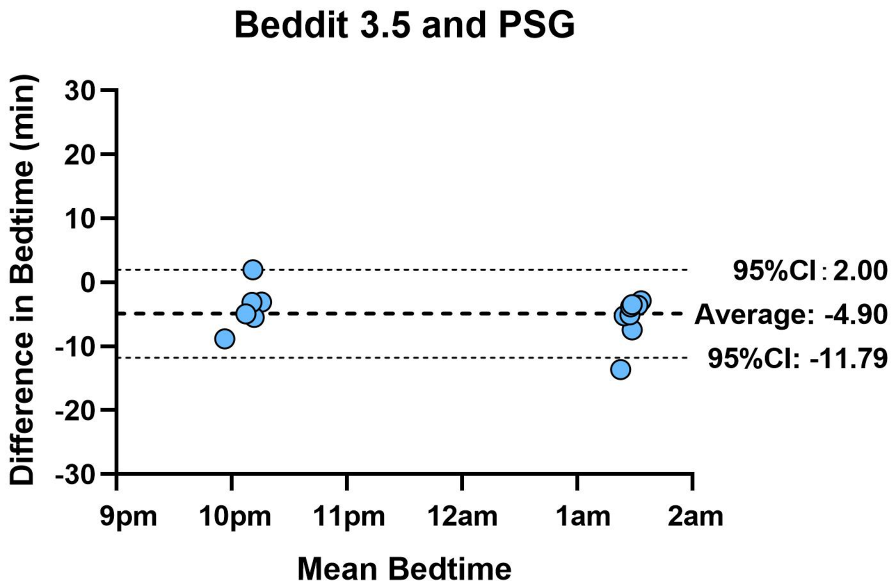
	Beddit 3.5	Beddit 3.0	Actigraphy	Self-report
Actigraphy	$r(10) = .99$, $p < .001$***	$r(17) = .42$, $p = .075$	----	----
Self-report	$r(11) = .75$, $p = .003$**	$r(19) = .20$, $p = .386$	$r(28) = .52$, $p = .003$**	----
PSG	$r(12) = .998$, $p < .001$***	$r(19) = .38$, $p = .089$	$r(29) = .96$, $p < .001$***	$r(32) = .58$, $p < .001$***

*p<.05, **p<.01, ***p<.001.

Supplemental Table 2. Bivariate Correlation among Different Measures for Sleep Efficiency

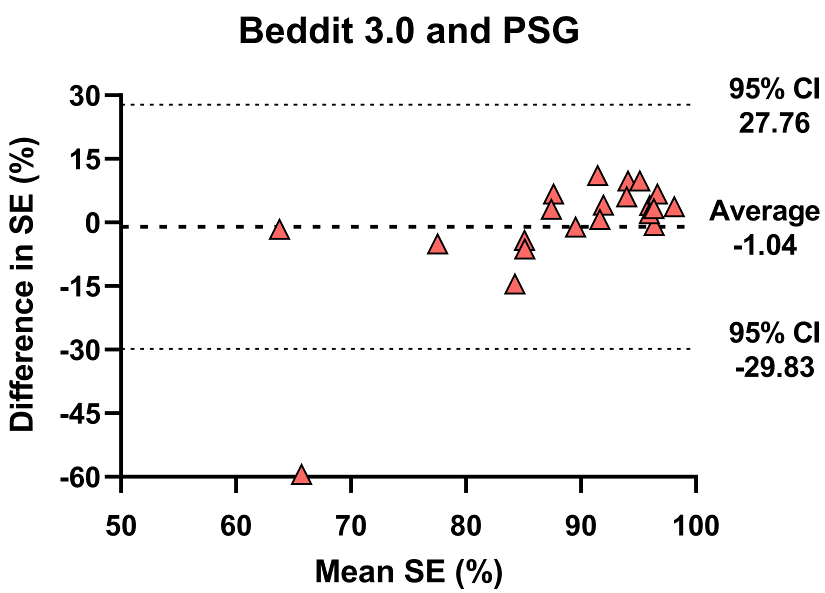
	Beddit 3.5	Beddit 3.0	Actigraphy	Self-report
Actigraphy	$r(10) = .96$, $p < .001$***	$r(17) = .60$, $p = .007$**	----	----
Self-report	$r(11) = .66$, $p = .015$*	$r(19) = -.002$, $p = .994$	$r(28) = .28$, $p = .132$	----
PSG	$r(12) = .99$, $p < .001$***	$r(19) = .34$, $p = .133$	$r(29) = .81$, $p < .001$***	$r(32) = .45$, $p = .008$**

*p<.05, **p<.01, ***p<.001.

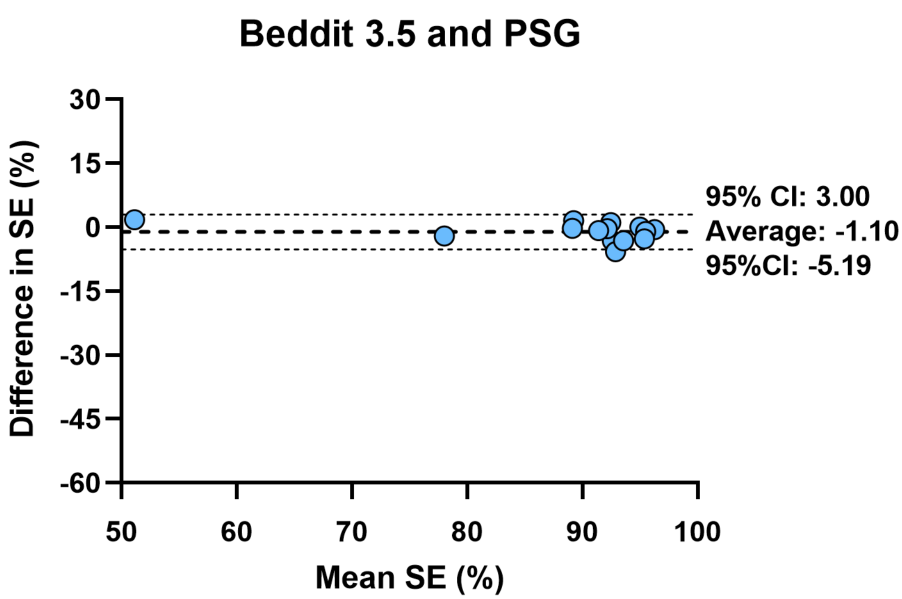


Supplemental Figure 1. Bland-Altman plot for bedtime between Beddit 3.5 and PSG.

(A)

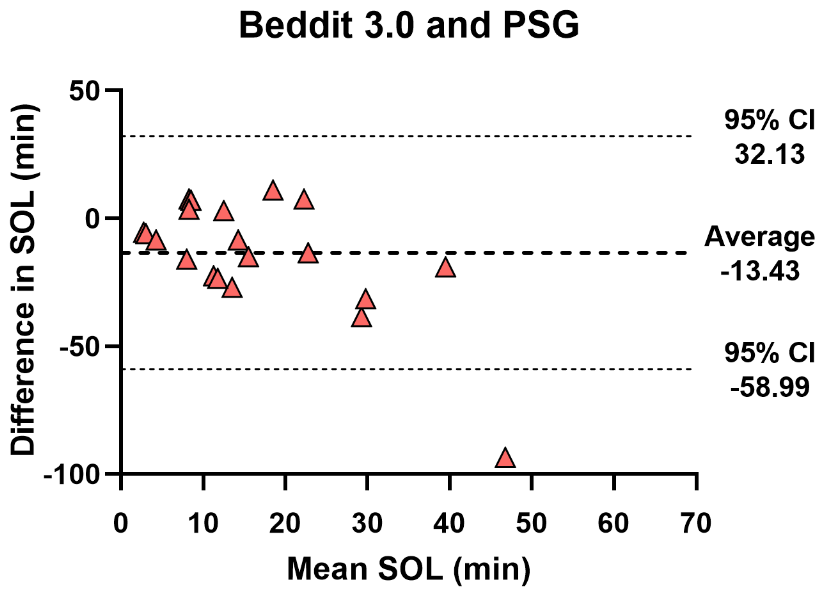


(B)

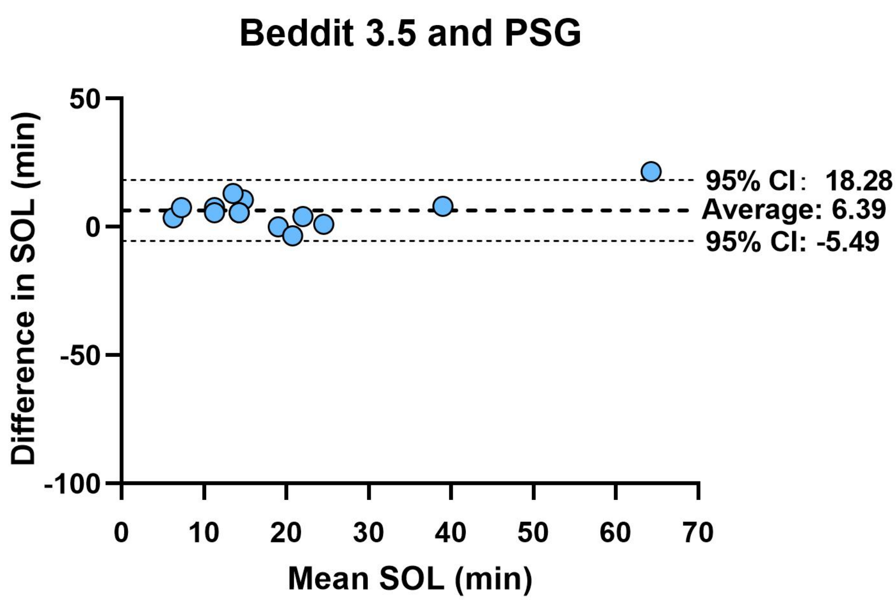


Supplemental Figure 2. Bland-Altman plots for sleep efficiency (SE) between Beddit 3.0 and PSG (A), and Beddit 3.5 and PSG (B).

(A)

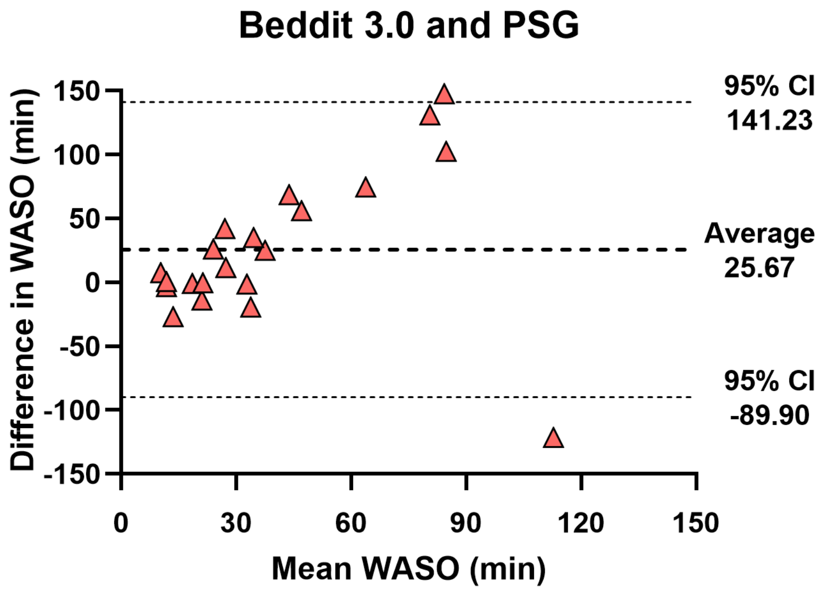


(B)

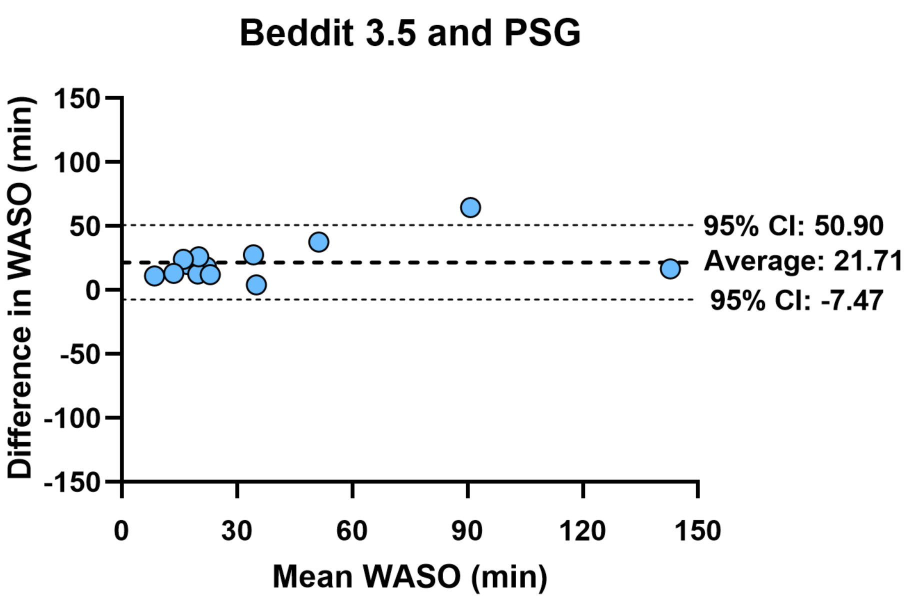


Supplemental Figure 3. Bland-Altman plots for sleep onset latency (SOL) between Beddit 3.0 and PSG (A), and Beddit 3.5 and PSG (B).

(A)



(B)



Supplemental Figure 4. Bland-Altman plots for wake after sleep onset (WASO) between Beddit 3.0 and PSG (A), and Beddit 3.5 and PSG (B).