Criterion validity of a research-based application for tracking screen time on android and iOS smartphones and tablets

Kristensen, Peter L.; Olesen, Line G.; Egebæk, Heidi K.; Pedersen, Jesper; Rasmussen, Martin G.; Grøntved, Anders

Published in:
Computers in Human Behavior Reports

DOI:
10.1016/j.chbr.2021.100164

Publication date:
2022

Document version:
Final published version

Computers in Human Behavior Reports 5 (2022) 100164

Contents lists available at ScienceDirect

Computers in Human Behavior Reports

journal homepage: www.sciencedirect.com/journal/computers-in-human-behavior-reports

Criterion validity of a research-based application for tracking screen time on android and iOS smartphones and tablets

Peter L. Kristensen [a,*], Line G. Olesen [a], Heidi K. Egebæk [a,b], Jesper Pedersen [a], Martin G. Rasmussen [a], Anders Grøntved [a]

[a] *Center for Research in Childhood Health/Unit for Exercise Epidemiology, Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, 5230, Odense M, Denmark*

[b] *University College South Denmark, Lembckesvej 7A, 6100, Haderslev, Denmark*

A R T I C L E   I N F O

Keywords:
Screen time
Validity
Reproducibility
Smartphone
Limit of agreement
Objective assessment

A B S T R A C T

This study evaluated the criterion validity of a newly developed application for assessing smartphone use on iOS and Android based devices. A sample of adult Danes (n=40; 40% female; median age= 42.0 years) had the application SDU DeviceTracker installed on their mobile phone together with a commercial application that served as a reference for comparison (Apple Screen Time (iOS) or ActionDash (Android)). Each participant collected data for at least 6 days, adding up to a total of 277 assessment days. Day-level analyses were performed and Bland-Altman limits of agreement for repeated measurements and repeated measurement correlation was calculated. A non-significant mean bias of −0.8 min/day and a correlation coefficient of 0.99 (95% CI: 0.98 to 0.99) was observed for the assessment of screen time on Android devices compared to the reference method. For iOS devices a significant mean bias of 19.3 min/day and a correlation coefficient of r=0.88 (95% CI: 0.84 to 0.91) was observed. A correction method for systematic error in the assessment of screen time on iOS devices was provided. The study concluded that SDU DeviceTracker provides an opportunity to collect highly detailed and valid information about screen usage over extended periods of time.

1. Introduction

Screen devices such as mobile phones and tablets are a central focal point in many people's lives today. The devices provide easily accessible opportunities for entertainment, learning, information retrieval, shopping and communication, and are used in almost every aspect of human life. Population-based studies show that both children and adults spend a large part of their day in front of a screen device (Folkesundhed, 2019; Friel et al., 2020; Statistik, 2018; Whiting et al., 2020), and this trend has gained ground over a very short number of years. The first iPhone was launched as late as 2007 (Apple.com, 2007) and a decade later, 88 percent of all households in Denmark owned at least one smartphone, according to a danish study from 2018 (Statistik, 2018). The possible health effects associated with this development have not yet been sufficiently studied. Experts in several health disciplines have expressed concern and put forward hypotheses that link excessive screen time with health issues such as poor mental health, physical inactivity, poor sleep, obesity and poor eating habits (Biddle et al., 2017; Domingues-Montanari, 2017; Fang et al., 2019; Hale & Guan, 2015; Lissak, 2018; Shqair et al., 2019). A general characteristic of previous studies addressing screen time is that they almost exclusively use self-reporting questionnaires for assessing screen time (Biddle et al., 2017; Domingues-Montanari, 2017; Fang et al., 2019; Hale & Guan, 2015; Lissak, 2018; Shqair et al., 2019). Questionnaires have proven valuable for collecting health-related information, but they also have certain limitations and weaknesses. Questionnaires can be prone to missing data and specific types of information bias (e.g. social desirability bias and recall bias) depending on the subject being studied. In particular, activities that are often repeated and do not leave a clear imprint in the mind, such as frequent use of a mobile phone, can be difficult to account for using a questionnaire (Schwarz & Oyserman, 2001). In line with this a recent review study designed to examine the consistency between logged and self-reported use of digital media concluded, that self-reported media use correlates only moderately with logged measurements (Parry et al., 2021), and several other investigations have reached similar conclusions (Andrews et al., 2015;

* Corresponding author. Department of Sports Science and Clinical Biomechanics University of Southern Denmark 5230 Odense M, Denmark.
E-mail address: plkristensen@health.sdu.dk (P.L. Kristensen).

https://doi.org/10.1016/j.chbr.2021.100164
Received 26 August 2021; Received in revised form 16 November 2021; Accepted 17 December 2021
Available online 20 December 2021

Verbeij, 2021). Information bias is not only a significant problem in descriptive studies or in observational studies examining the relationship between screen time and health outcomes, but also in randomized trials designed to change screen time, as it can bias the assessment of adherence to assigned intervention by trial participants.

Unlike self-reporting, objective assessment methods have the advantage that information can be obtained independently of the participants' ability to recall and explain the information. Several commercial applications for assessing screen time on mobile devices exist. For instance, there is an official Apple Screen Time application built into the operating system of Apple's iPhone, which provides an overview of daily screen time and unlockings. Google has developed a similar application named Digital Balance for Android based devices, but the application is not available on all Android devices. Alternatively, the application ActionDash can be used, which has almost the same functionalities and is available for all Android devices. For research purposes, however, these commercial applications present several problems. Among the main issues is that screen time statistics must be registered manually by the researcher, since the information cannot be automatically forwarded to a secure server, and the level of detail in the available information is limited. The primary purpose of this study was to evaluate the criterion validity of a newly developed application assessing smartphone use on iOS and Android-based devices. The application is named SDU DeviceTracker and was designed specifically for research use. The secondary purpose was to study the day-to-day reproducibility of daily time spent on smartphone usage among adults.

2. Methods

The present study is based on a convenience sample of adult Danes who collected information about daily screen time using a newly developed application as well as an already existing and recognized commercial application for the same purpose. Data were used to assess the criterion validity of the newly developed application SDU DeviceTracker.

2.1. Participants

All employees at the Department of Sports Science and Clinical Biomechanics at the University of Southern Denmark ($n = 177$, 85 women and 92 men) were invited by email to participate in the study. The invitation included a description of the project as well as legal information about the participants' rights and the procedures for handling sensitive personal information. A total of 40 individuals accepted the invitation and participated in the study.

The study was reported to and approved by the local data protection department SDU RIO (ID: 10.372) in accordance with applicable rules of the Danish Data Protection Agency.

2.2. Sampling

Convenience sampling was used to recruit participants due to logistical benefits when installing and extracting screen time information from the applications. Equally important, we reasoned that any validity issues of the SDU DeviceTracker application would be likely to show up in a convenience sample as well as in a random sample, since technical errors in applications tend to operate independently of the user.

2.3. Protocol

Data were collected between November 2019 and February 2020. An introductory meeting was scheduled with each participant to provide oral information about the project and install the necessary applications. The SDU DeviceTracker application continuously registers screen time and sends the encrypted information to a secure server at the University of Southern Denmark. For the purpose of comparison, screen time was also recorded using an existing commercial application. On Android-based phones, the application ActionDash was installed and initialized. On iOS-based devices, screen time information was provided by the Apple Screen Time application built into the iOS operating system from iOS version 12 and onwards.

After installation, a second meeting was scheduled no earlier than 1 week and no later than 1 month after installation. Participants could freely choose a meeting time within this period to ensure maximum flexibility and minimize the risk of drop out and data loss. At the second meeting, the amount of daily screen time and the number of daily unlockings were manually extracted for the last 6–7 days for Android devices and up to 18 days for iOS-based devices. The difference in procedure depending on the type of operating system was due to the fact that ActionDash only registers the last 7 days of screen time activity, while IOS devices, depending on the operating system version, allow registration further back in time.

2.4. Screen time applications

2.4.1. SDU DeviceTracker

The SDU DeviceTracker application was designed by the authors of this article and programmed by a private company called Centic (www.Centic.dk). Two different applications were developed for Android and iOS operating systems, respectively. Both versions are activated by scanning a personal Quick Response code (QR-code) containing an identification number and a specification of the length of the measurement period. The basic design principles are described below.

2.4.1.1. IOS. The iOS version of the application detects whether a device is locked or unlocked using a part of an iOS API that is not publicly available. It is usually not advisable to use a private API, partly because the behavior of a private API is not guaranteed, and the application cannot be distributed in Apple's App Store. However, in our best belief, there are no other options at present. Both manual locking and auto locking of a device is classified as a lock event in the application. Device activities where the screen is not actively used (e.g. streaming of music over prolonged periods of time) is classified as screen time until the phone automatically locks. When using the phone without unlocking it, for example when previewing notifications, no events are registered. If a device runs out of power or SDU DeviceTracker is force quit by the user, an event is also received. The program operates with the following event codes: 1 = unlock device, 0 = lock device, 2 = force quit, 3 = power loss. All events are stored in a local database on the device along with an exact timestamp with millisecond precision regardless of whether the phone is offline or online. The information is automatically sent to a secure server at the University of Southern Denmark. The application is kept alive in the background by keeping track of GPS changes, which prevents the application from entering idle mode with only a minimal impact on phone battery life. If the application is force quit by the user, it will restart itself when new GPS location data arrives.

2.4.1.2. Android. The Android version of SDU DeviceTracker uses broadcast action messages from the operating system which Android offers full access to. The primary broadcast events used are SCREEN_ON and SCREEN_OFF, and the application stores these two types of events in a local database on the device along with an exact timestamp for each event regardless of whether the phone is offline or online. Unlike the iOS application, notifications that cause the screen to turn on without unlocking the device will also be registered as events. When making phone calls or streaming music, screen time is only recorded as long as the screen is on. Every half hour, data is sent to a secure server at the University of Southern Denmark. The Android operating system allows the application to run in the background at all times and it is not possible for the user to close the application manually. It is only in the event that a user uninstalls the application that the measurements stop.

2.4.2. ActionDash and Apple Screen Time

ActionDash and Apple Screen Time were used as a basis for comparison for the SDU DeviceTracker application. An official and detailed description of the algorithms underlying these two applications has not been disclosed. Only general descriptions are available, and to ensure that the applications were comparable to SDU DeviceTracker, we investigated how screen time is defined and calculated by these applications before the study was initiated. The ActionDash application offers user-specific settings for assessing screen time and the settings were standardized on all devices to match the SDU DeviceTracker algorithm. However, notifications are not classified as screen time in ActionDash as opposed to in Android SDU DeviceTracker. To compensate, an algorithm was developed that filters out notifications in the processing of Android SDU DeviceTracker data. In contrast to ActionDash the Apple Screen Time application offers very limited options for controlling screen time calculations, but the initial testing of the application revealed no fundamental discrepancies in the definition and calculation of screen time compared to the iOS DeviceTracker algorithm.

2.4.3. SDU DeviceTracker data processing

SDU DeviceTracker data was processed using a Python program developed by the first author of this article before the study was initiated. The program summarizes screen time and unlockings on a daily and hourly basis for each user. In case of lost registration time due to force quit of the application, the time loss is calculated and reported by the program.

During the coding phase of the iOS version of SDU DeviceTracker, test data revealed that the order of events occasionally deviated from what was expected, for instance due to an unlock or lock event not being registered. After an unlock event a lock event should always follow unless the phone runs out of power or the application is force quit, but in iOS test data, an unlock event was in some cases followed by another unlock event. It was not possible to solve this issue in the coding of the SDU DeviceTracker application, as we suspect that the problem is related to the use of a private API and the options available to keep the iOS application running in the background. The problem was therefore addressed in the Python program. For this purpose, a short protocol was prepared in which the authors of this article used the iOS application for a limited period while all events (unlocks/locks) were manually registered with a precision of 1 s. This provided an opportunity to study patterns in events that did not follow the logically expected order and subsequently develop a strategy to compensate in the Python program. The Python program offers to import an optional screen time interval in case that an unlock event is not followed by a lock event or vice versa. An interval of 60 s of screen time was imputed in the present study. The program also offers to filter out bouts of screen time longer than a user-specified length to avoid long erroneous bouts that could occur if the application fails to register all events. For instance, consider the following hypothetical series of events: Unlock phone (9:30), Lock phone (9:31), Unlock phone (20:40), Lock phone (21:00). If for some reason the application fails to register the two middle events, an erroneous screen time interval from 9:30–21:00 will appear in the output file. In this study, a maximum bout length of 6 h was allowed. The number of deviations from the expected order of events and the number of bouts filtered out due to the maximum bout length criteria are counted and reported by the program for quality control purposes and to support the subsequent analysis of data.

As earlier explained, notifications were filtered out in the processing of Android data in the present study for comparative reasons. Bouts of screen time of less than 40 s with an accuracy of one decimal that recurred repeatedly (at least 3 times) and accounted for at least 1% of all event recordings were considered notifications. These criteria were selected after studying the likelihood of bouts recurring in iOS SDU DeviceTracker data, where notifications are not registered as screen time.

2.5. Statistical methods

Day-level analyses were performed and Bland-Altman limits of agreement for repeated measurements was calculated as described by Bland-Altman and Zou (Bland & Altman, 2007; Zou, 2013). There are two different situations to consider when calculating limits of agreement for replicated data (Bland & Altman, 2007). The observations for the same subject could either be a series of measurements of a quantity that does not vary over the period of observation or it could be a series of measurements of a changing quantity. In this study, the “changing quantity” situation applies since the amount of screen time and the number of unlockings vary from day to day within individuals. The assumption of normality of the differences was evaluated by visual inspection of histograms, Q-Q plots and boxplots and deemed reasonable with one exception. The differences with respect to daily unlockings was not normally distributed for the Android application. Logarithmic transformation did not solve the issue, thus a nonparametric approach of estimating limits of agreement was implemented. A trend towards proportional bias (i.e. greater variability at higher values) was observed for both Android and IOS measurements of the daily number of unlockings, however the trend was eliminated by plotting the difference expressed as a percentage of the mean. The Preiss–Fisher procedure was applied to check if the measurement range was sufficiently wide for the Bland-Altman analyses (Gerke, 2020; Preiss & Fisher, 2008), and locally weighted scatterplot smoothing (lowess) analyses were applied to study the relationship between the SDU DeviceTracker and Apple Screen Time/ActionDash data.

Repeated measurement correlation within subjects was calculated to study if an increase in SDU DeviceTracker assessed screen time within individuals was associated with an increase in the reference estimate of screen time (Bland & Altman, 1995). The assumption of linearity was checked by visual inspection. The package rmcorr in R was used for the calculations.

The day-to-day reproducibility of daily time spent on smartphone usage was assessed using variance partitioning applying a linear mixed effect model controlling for day type (weekend vs. weekday). Number of days needed to obtain a reliability (ICC) of 0.70, 0.80 or 0.90, respectively, was estimated using the Spearman Brown prophecy formula (Trost et al., 2005).

All analyses were conducted in either Stata/IC 16.1 or R version 4.0.2.

3. Results

3.1. Descriptive analyses

A total of 40 men and women accepted the invitation and participated in the study. Fig. 1 shows a flow chart illustrating the data collection process. As shown in the figure, the installation of SDU DeviceTracker failed on one Android device and ActionDash failed collecting data on one occasion. In addition, four participants had no SDU DeviceTracker data at the 2nd meeting, as the application had stopped collecting data at some point between the 1st and 2nd meeting. In all, 34 participants were included in the statistical analyses and all but four participants had at least 6 days of data. In total, the study was based on 277 individual days of measurement.

Descriptive characteristics of participants and daily smartphone use are shown in Table 1.

3.2. Data quality indicators

Table 2 provides an overview of the available data quality indicators. No inconsistencies (i.e. deviations from the logically expected order of events) were observed for the Android data, thus only iOS results are

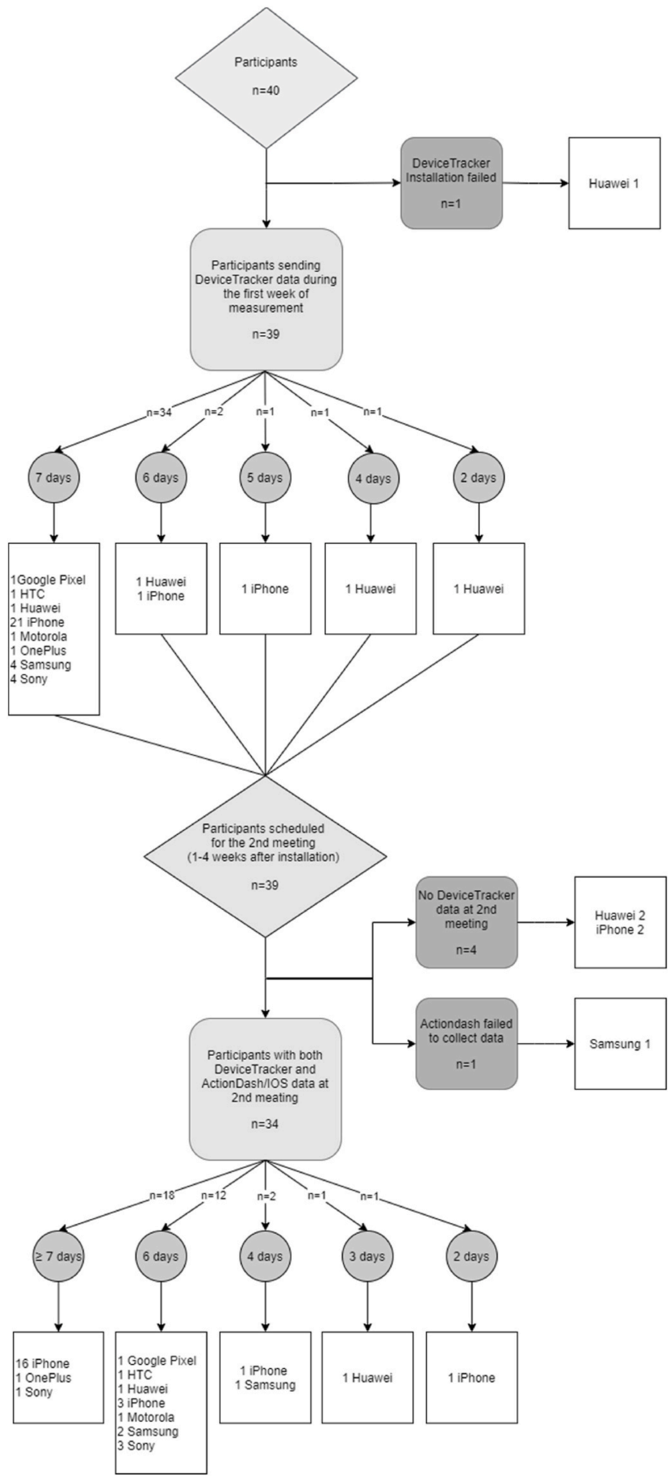

Fig. 1. Flow chart illustrating the data collection process.

presented. A force quit event was observed on 15 out of 201 days, while inconsistent events were observed on just over half of the days. In total, 56% of the inconsistent events followed a recognizable pattern observed in the programming and testing phase of the application and were handled in the Python processing code without imputations. The remaining inconsistent events were interpreted as valid but incomplete events (e.g. an unlock event lacking the associated lock event) and handled in the Python program by imputation.

Table 1
Descriptive characteristics and daily smartphone use.

	Android	iOS	Total
n	13	21	34
Sex, n (male/female)	9/4	12/9	21/13
Age, years	41.8 (37.9; 45.9)	42.2 (31.9; 51.3)	42.0 (32.8; 46.9)
ISCED, n (≤2/3–6/≥7)	0/0/34	0/0/34	0/0/34
Measurement days, total n of individual days	76	201	277
Measurement days, n per participant	6 (6; 6)	8 (7; 13)	7 (6; 10)
Smartphone usages, min/day*	123.1 (74.0; 145.8)	109.8 (77.5; 154.7)	114.4 (77.3; 149.8)
Smartphone unlocks, n/day*	34 (27; 37)	54 (42; 76)	42 (33; 66)
Bout length, sec	38.8 (11.2; 137.2)	50.6 (16.2; 137.4)	47.4 (14.6; 137.4)

Medians and 25th and 75th percentiles are presented for all variables except for *Sex, ISCED* and *Measurement days, total n of individual days*. *Smartphone usage is based on ActionDash and Apple Screen Time measurements. ISCED: International Classification of Education.

Table 2
Data quality indicators for the iOS SDU DeviceTracker application, n=201 days.

	IOS
Number of events, n/day*	74 (54; 112)
Force quit events, n of days with (0/1/2) force quit events	186/13/2
Force quit duration, minutes/day**	91.3 (15.2; 248.9)
Inconsistent events, n of days with (0/1–3/4–7/8–23) inconsistent events***	87/33/36/45
Percentage of inconsistent events per day****	6.8 (4.5; 11.5)
Number of bouts that exceeded the maximum bout length criterion*****	0

*Median (25th and 75th percentile). **Median (25th and 75th percentile) duration of lost registration time calculated for days with at least one force quit event, n=15. ***Inconsistent events are defined as events (unlock/lock/force quit/power loss) in the output file that do not follow the logically expected structure (for example, an unlock event followed by another unlock event). ****Median (25th and 75th percentile) percentage of inconsistent events calculated for days with at least one inconsistent event, n=114. ***** A maximum bout length of 6 h was used.

3.3. Repeated measurement correlation

Repeated measurement correlation was calculated between SDU DeviceTracker and Apple Screen Time/ActionDash outputs. Higher correlations were observed for the Android version of SDU DeviceTracker compared to the IOS version, and for the measurements of screen time versus unlocks. The highest observed correlation was 0.99 for the assessments of daily screen time on Android devices (See Table 3).

3.4. Bland-Altman analyses

Limits of agreement plots are shown in Figs. 2–5. According to the

Table 3
Repeated measurement correlation (r) between SDU DeviceTracker and Apple Screen Time/ActionDash measurements, day-level analysis.

	n	n days	r	Lower 95% CI	Upper 95% CI
IOS					
Screen time	21	201	0.88	0.84	0.91
Unlocks			0.58	0.48	0.67
Android					
Screen time	13	76	0.99	0.98	0.99
Unlocks			0.87	0.80	0.92

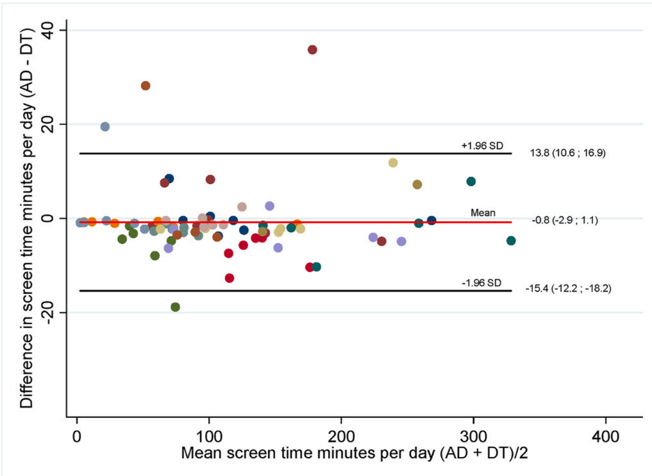

Fig. 2. Repeated measures Bland-Altman plot of intermethod agreement between SDU DeviceTracker (DT) and ActionDash (AD) measurements of daily screen time, min/day. (Android).
The plot shows the mean difference (red line) and the limits of agreement (black lines) including 95% confidence intervals in parentheses. Repeated measurements within the same individual are marked with a unique color. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

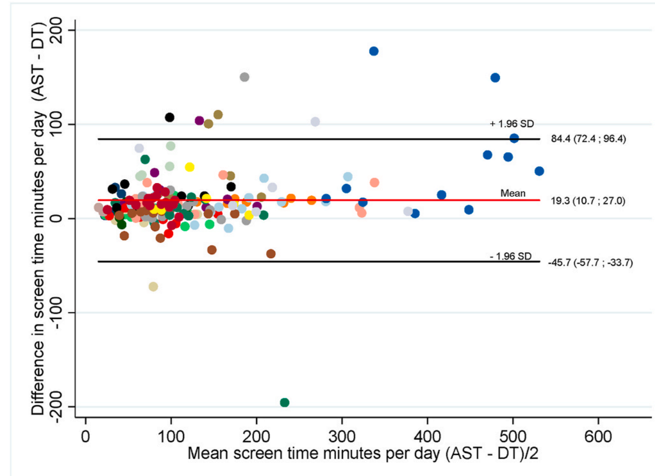

Fig. 4. Repeated measures Bland-Altman plot of intermethod agreement between SDU DeviceTracker (DT) and Apple Screen Time (AST) measurements of daily screen time, min/day. (IOS).
The plot shows the mean difference (red line) and the limits of agreement (black lines) including 95% confidence intervals in parentheses. Repeated measurements within the same individual are marked with a unique color. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

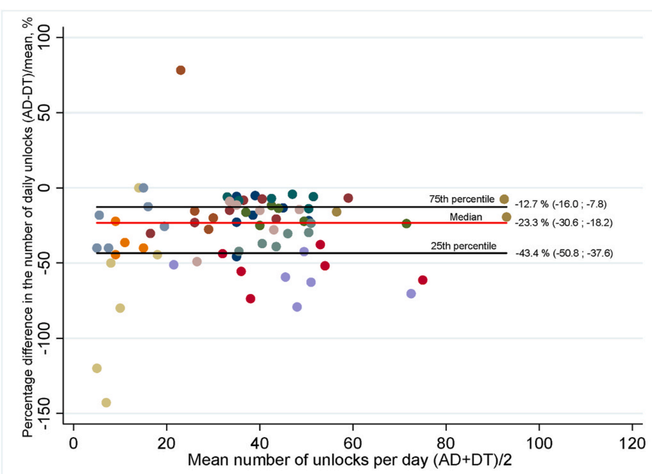

Fig. 3. Repeated measures Bland-Altman plot of intermethod agreement between SDU DeviceTracker (DT) and ActionDash (AD) measurements of the daily number of unlockings. (Android).
The plot shows the median difference (red line) and the limits of agreement (black lines) including 95% confidence intervals in parentheses. Note that medians and percentiles are displayed due to nonparametric data. Repeated measurements within the same individual are marked with a unique color. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Fig. 5. Repeated measures Bland-Altman plot of intermethod agreement between SDU DeviceTracker (DT) and Apple Screen Time (APS) measurements of the daily number of unlockings. (IOS).
The plot shows the mean difference (red line) and the limits of agreement (black lines) including 95% confidence intervals in parentheses. Repeated measurements within the same individual are marked with a unique color. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Preiss–Fisher procedure, the measurement range was sufficiently wide for all Bland-Altman analyses. The data was randomly mispaired 1000 times, and the distribution of the respective 1000 standard deviations was clearly beyond the observed standard deviation of the differences in the sample – see Fig. 6 for an example.

A non-significant mean bias of −0.8 min/day with a slight trend towards proportional bias was observed for the Bland-Altman analysis of daily screen time assessed on Android devices. Individual differences between SDU DeviceTracker and ActionDash assessments were estimated to lie within + - 15 min per day in 95 percent of cases (Fig. 2). In terms of daily unlockings, however, a significant median percentage bias of 23.3% was observed indicating that Android SDU DeviceTracker tends to detect more daily unlockings compared to ActionDash. The differences were not symmetrically distributed around the median but appeared to accumulate close to zero without exceeding zero (Fig. 3).

As to the iOS application, the Bland-Altman analyses showed a significant mean bias towards underestimation of both daily screen time and unlockings, and wider limits of agreement were observed compared to the Android application (Figs. 4 and 5). To investigate the possibility of correcting for the systematic bias, a lowess smoothing analysis was performed on the iOS data. Fig. 7 shows a scatterplot of the association between Apple Screen Time and SDU DeviceTracker assessments of daily

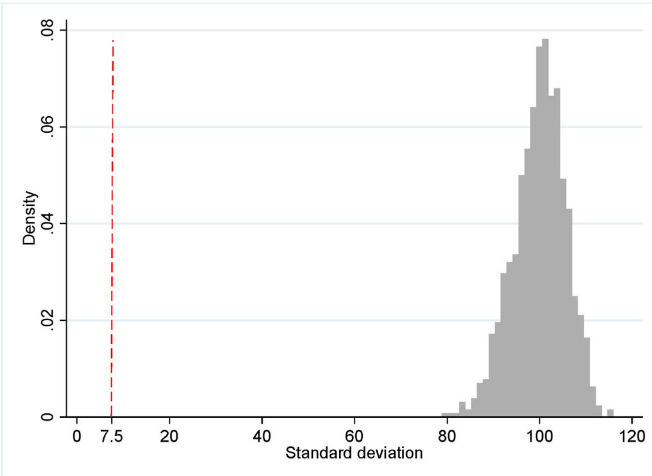

Fig. 6. Distribution of the standard deviation from 1000 random mispairings of android assessments of screen time according to the Preiss-Fisher procedure. The observed standard deviation (7.47) is clearly smaller than the minimum of the standard deviations from 1000 random mispairings (78.8).

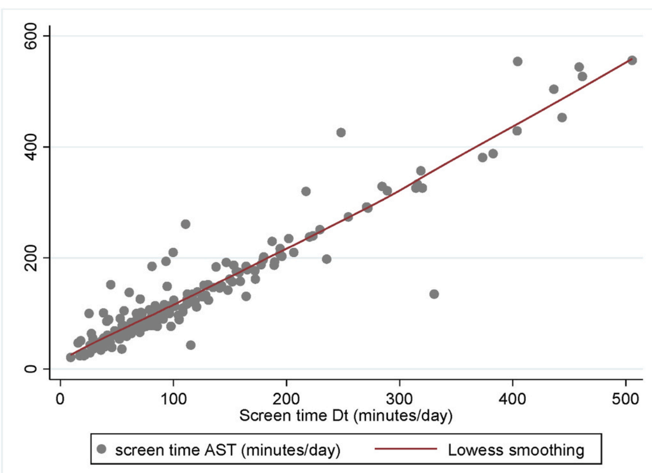

Fig. 7. Locally Weighted Scatterplot Smoothing for the association between Apple Screen Time (AST) and SDU DeviceTracker (DT) assessments of daily screen time, minutes/day.

screen time including a locally weighted smoothing line. As shown in the figure, the lowess analysis indicated a clear linear relationship suggesting that a simple linear model could be used for correcting for systematic bias. A mixed linear regression model with a random effect for subjects was therefore fitted via maximum likelihood estimation to obtain a linear function that can be used to correct iOS SDU DeviceTracker data. Gender was not significant in the model and therefore omitted. The daily number of minutes of lost registration time due to force quit of the application was included as an independent variable to correct for a potential underestimation of screen time due to force quit. Results are shown in Table 4. For daily unlockings the lowess analysis did not show an equally clear linear relationship, and for this reason we chose not to model the differences for this outcome.

3.5. *Day to day reproducibility of smartphone use*

Table 5 shows the reproducibility of SDU DeviceTracker assessed daily screen time. The reproducibility of a single day of measurement was ICC = 0.58 and analyzes showed that 2.9–6.5 days were needed to raise the reproducibility to ICC = 0.8 and 0.9, respectively.

Table 4
Mixed effect linear regression analysis of the association between Apple Screen Time minutes/day (AST) and SDU DeviceTracker (DT) assessments of daily screen time (minutes/day), IOS.

	β	95% CI	P-value
Fixed parts			
Screen time (DT)	0.94	0.88; 1.00	0.00
Daily number of force quit minutes	0.07	0.01; 0.13	0.02
Intercept	25.28	14.65; 38.00	0.00
Random parts			
N_{grp}	21		
N	201		
Subject variance (95% CI)	369.1 (155.1; 878.6)		
Residual variance (95% (CI)	754.2 (611.3; 930.5)		

Table 5
Reproducibility of the assessment of daily screen time (minutes/day) by the SDU DeviceTracker application. iOS and Android data combined.

N	n days	ICC_s	$N_{0.7}$	$N_{0.8}$	$N_{0.9}$
34	277	0.58	1.7	2.9	6.5

ICCs Intra-class correlation for a single day of measurement adjusted for day type and gender.
N_x Number of days needed to achieve an ICC of size x.

4. Discussion

In this study, the criterion validity of a newly developed application for assessing screen time on iOS and Android devices specifically designed for research purposes was studied.

4.1. *SDU DeviceTracker validity*

In terms of assessing daily screen time no significant mean bias was observed for the Android SDU DeviceTracker application. Although no pre-established acceptable limits of agreement was defined, we consider a deviation of + - 15 min per day in 95 percent of cases highly acceptable, especially in light of a median screen time usage of 115 min/day in the sample. The estimation of daily unlockings, however, did not agree to the same extent with the reference method. A likely explanation is that SDU DeviceTracker classifies notifications as screen time while ActionDash sorts out notifications – as explained in the methods section. Notifications activate the screen for a short period of time and have little potential to impact the overall screen time, but notifications can significantly increase the daily number of unlockings. Although an algorithm was implemented to filter out notifications post hoc from the SDU DeviceTracker output, not all notifications can be expected to be identified. The algorithm searched for bouts of screen time of less than 40 s with an accuracy of one decimal that tend to recur more often than expected by chance. To ensure that only notifications were filtered out, a slightly conservative approach was chosen, and notifications that were not repeated often or varied greatly in length were more likely to remain in the output file. The most accurate method would have been to filter out notifications in the Android code, but this approach was not chosen in the design of the application as it can be debated whether notifications should be considered as screen time or not. The post hoc filtering algorithm was implemented solely for the sake of comparability to the ActionDash approach.

The iOS application significantly underestimated both the daily amount of screen time and the number of unlockings. Limits of agreement were also markedly wider compared to the Android application indicating that there is a risk of considerable error on an individual level. However, as reported in the results section (Table 4), further analysis indicated that the mean bias could be accounted for by the following equation: Screen time (min/day) = 0.94*Screen time_DeviceTracker (min/day) + 0.07*Force_quit time (min/day) + 25.28. Ideally, this equation should be cross validated in a new study, and preferably for different age groups, as it cannot be ruled out that different usage patterns could impact the parameters of the equation. The adjustment for systematic bias will not affect the width of the limits of agreement, but the limits should slightly decrease if screen time is calculated as an average of several days instead of using day level estimates. In many studies an average estimate of screen time across several days would probably be preferable. As a post hoc analysis, we repeated the Bland-Altman analysis based on an average of daily screen time across all included days. Mean bias was still estimated at 19.3 min/day for IOS screen time, but the limits of agreement narrowed in from (−45.7; 84.4) to (−19.8; 58.5). Despite this improvement, the agreement is still not quite comparable to that of the Android application, and it is doubtful whether the gap can be closed unless Apple releases a public API granting developers access to the same functionalities as Apple's native "Screen Time" uses. A group of software developers have already put forward a Screen time Api proposal and hope that Apple will address this in a future update (Screen time Api proposal. *Access date 11-06-2021*). In the meantime, iOS SDU DeviceTracker appears to be a very good alternative compared to potential self-reporting solutions. There are only few available questionnaire batteries to comprehensively assess screen time use (Klakk et al., 2020), but there are several questionnaires assessing sedentary time. To the best of our knowledge no sedentary questionnaire correlates 0.88 to a gold standard reference measure, as is the case for IOS SDU DeviceTracker, and the Bland-Altman analysis of agreement for existing sedentary questionnaires rarely show an equally symmetrical result as for the IOS SDU DeviceTracker application (Cleland et al., 2014; Rivière et al., 2018; Sagelv et al., 2020; Tanaka et al., 2021). In addition, SDU DeviceTracker can provide detailed information about components of screen time that are difficult to assess by means of self-report such as the frequency and distribution of screen time bouts throughout specific periods of the day. Thus, going forward, the DeviceTracker or similar applications will enhance the possibilities of studying the potential health impact of smartphone behaviors that can be quantified - in particularly behaviors that are frequent and highly integrated into respondents' lives which makes them difficult to distinguish and retrieve accurately by memory. An example could be "frequent phone checking" throughout the day or smartphone use before bedtime, which hypothetically have been linked to mental health and poor sleep quality, respectively (Hale & Guan, 2015) (Kornberg, 2020).

As can be seen from Fig. 1 a few devices stopped collecting data before intended in this study. This was not entirely unexpected for the iOS application since iOS does not provide the same opportunities as Android for keeping the application running in the background, as described in the method section. Two out of 23 iPhones did not collect any data at the 2nd meeting, and 2 iPhones collected less than the expected minimum of six measurement days. For the Android application no data loss was expected, but out of 17 Android phones 1 Huawei failed to install the application and 2 Huawei phones did not collect any data at the 2nd meeting. Two additional phones (1 Huawei and 1 Samsung) collected less than the expected minimum of six measuring days, which, in all fairness, does not necessarily indicate a loss of data. Given the specific sample, it is possible that not all participants used their smartphone on a daily basis, especially if the application was installed on a company phone instead of a personal phone. The observed pattern of Huawei phones being more likely to stop collecting data than other phone brands may not be a coincidence and should be pursued in a future update of the SDU DeviceTracker application. It is important to point out to others who might be considering developing their own application, that the creation of an application is an ongoing work process, and it is necessary to monitor updates of the Android and iOS operating systems to adjust the application accordingly.

At present we are only aware of a few other research groups who have used an application developed specifically for research use to assess screen time (Katapally & Chu, 2019) (Andrews et al., 2015) (Geyer et al., 2021). To the best of our knowledge, an in-depth Bland-Altmann validation analysis against other objective methods has not yet been published for these applications, and as a consequence they cannot serve as a direct comparative basis for a performance evaluation of the SDU DeviceTracker application.

4.2. Day to day reproducibility of smartphone use

As a secondary aim, the reproducibility of objectively assessed screen time was studied. Screen time behavior will vary over time, and it is important to determine how many days of measurement that should be included to obtain reproducible estimates of habitual screen time. The results indicated that a reasonable reliability (i.e., intraclass correlation of approximately 0.80) could be achieved with as little as 3 days of monitoring. Compared to studies of the reproducibility of objectively assessed sedentary time among adults this is a relatively low estimate. Matthews et al. (Matthews et al., 2002) suggested seven days of monitoring to reliably assess inactivity among adults (18–79 yr) and Hart et al.(Hart et al., 2011) estimated that five days of complete data is needed to collect reproducible estimates (ICC=0.80) of sedentary behavior among older adults (55–86 yr). Due to the low level of inconvenience to the participants and due to recent studies indicating, that the method used to assess reproducibility could slightly underestimate the number of monitoring days needed (Aadland et al., 2020), we suggest 7 days of monitoring for assessing habitual screen time in future studies. Of note, we did not examine seasonal variation which might occur and could be relevant in some studies with assessments that span seasons.

4.3. Limitations

A limitation of the present study was the relatively low number of participants especially Android users. A larger sample would have provided an opportunity to study potential differences in validity across specific Android phone brands. Judging from the confidence intervals in the Bland-Altman analyzes, however, the sample size was large enough to provide acceptable certainty in the analyzes. A more heterogeneous sample including young people and different layers of society would have been preferable, as it cannot be ruled out that different usage patterns may impact the outcome of the validation. An additional limitation was the lack of complete comparability between SDU DeviceTracker and the reference applications. Different strategies have been implemented for the classification of notifications across applications, which in all probability impacted the validation of the assessment of daily unlockings negatively. At the time of development, iOS DeviceTracker was designed to run on iOS 9, iOS 10 and iOS 11, and no issues have yet been observed on newer versions of iOS. The Android SDU DeviceTracker was also tested on several versions of the operating system in the development phase, and so far, no compatibility issues have been observed with any Android version – if we disregard the results of this study, which suggest that Huawei's modified version of Android potentially causes Huawei phones to be more likely to stop collecting data unexpectedly during the collection period. However, it should be emphasized that future updates of iOS and Android may affect the stability and performance of the SDU DeviceTracker, and that the validation results reported in this study cannot be guaranteed for future versions of iOS and Android, unless the impact of these updates is continuously examined and SDU DeviceTracker is updated if necessary. A final limitation is that SDU DeviceTracker is not directly applicable to other researchers. This is because the application is coded to communicate with servers at the University of Southern Denmark and iOS DeviceTracker is distributed to research participants via Apple's Enterprise Program and not Apple's App store due to the use of a private API. However, the source code can be provided upon request if other researchers wish to use DeviceTracker as a starting point for developing their own application.

5. Conclusion

In conclusion, the SDU DeviceTracker application is a highly valid tool for assessing screen time and daily unlockings on Android smartphone devices. The iOS version of the application also showed a satisfactory degree of validity, although there is a need to correct for systematic bias, and the limits of agreement were wider compared to the Android application. Going forward, DeviceTracker will provide unique new opportunities for studying screen time and the relation to various health-related outcomes in observational studies. Also, SDU DeviceTracker provides an instrument for use in experimental studies for instance when smartphone use is an outcome or there is a need to monitor adherence to a smartphone use intervention. Besides limiting information bias in the assessment of screen time, SDU DeviceTracker provides an opportunity to collect highly detailed information about screen usage over extended periods of time.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the European Research Council (no. 716657).

References

Aadland, E., Nilsen, A. K. O., Ylvisåker, E., Johannessen, K., & Anderssen, S. A. (2020). Reproducibility of objectively measured physical activity: Reconsideration needed. *Journal of Sports Sciences, 38*(10), 1132–1139. https://doi.org/10.1080/02640414.2020.1743054

Andrews, S., Ellis, D. A., Shaw, H., & Piwek, L. (2015). Beyond self-report: Tools to compare estimated and real-world smartphone use. *PLoS One, 10*(10), Article e0139004. https://doi.org/10.1371/journal.pone.0139004

Apple.com. (2007). Press release https://www.apple.com/newsroom/2007/01/09Apple-Reinvents-the-Phone-with-iPhone/ https://www.apple.com/newsroom/2007/01/09Apple-Reinvents-the-Phone-with-iPhone/.

Biddle, S. J. H., García Bengoechea, E., Pedisic, Z., Bennie, J., Vergeer, I., & Wiesner, G. (2017). Screen time, other sedentary behaviours, and obesity risk in adults: A review of reviews. *Curr Obes Rep, 6*(2), 134–147. https://doi.org/10.1007/s13679-017-0256-9

Bland, J. M., & Altman, D. G. (1995). Calculating correlation coefficients with repeated observations: Part 1–Correlation within subjects. *BMJ, 310*(6977), 446. https://doi.org/10.1136/bmj.310.6977.446

Bland, J. M., & Altman, D. G. (2007). Agreement between methods of measurement with multiple observations per individual. *Journal of Biopharmaceutical Statistics, 17*(4), 571–582. https://doi.org/10.1080/10543400701329422

Cleland, C. L., Hunter, R. F., Kee, F., Cupples, M. E., Sallis, J. F., & Tully, M. A. (2014). Validity of the global physical activity questionnaire (GPAQ) in assessing levels and change in moderate-vigorous physical activity and sedentary behaviour. *BMC Public Health, 14*, 1255. https://doi.org/10.1186/1471-2458-14-1255

Domingues-Montanari, S. (2017). Clinical and psychological effects of excessive screen time on children. *Journal of Paediatrics and Child Health, 53*(4), 333–338. https://doi.org/10.1111/jpc.13462

Fang, K., Mu, M., Liu, K., & He, Y. (2019). Screen time and childhood overweight/obesity: A systematic review and meta-analysis. *Child: Care, Health and Development, 45*(5), 744–753. https://doi.org/10.1111/cch.12701

Folkesundhed, S. I.f. (2019). *Skolebørnsundersøgelsen 2018 - Helbred, trivsel og sundhedsadfærd blandt 11-, 13- og 15-årige skoleelever i Danmark*. https://www.sdu.dk/da/sif/rapporter/2019/skoleboernsundersoegelsen_2018.

Friel, C. P., Duran, A. T., Shechter, A., & Diaz, K. M. (2020). U.S. Children meeting physical activity, screen time, and sleep guidelines. *American Journal of Preventive Medicine, 59*(4), 513–521. https://doi.org/10.1016/j.amepre.2020.05.007

Gerke, O. (2020). Reporting standards for a bland-altman agreement analysis: A review of methodological reviews. *Diagnostics, 10*(5). https://doi.org/10.3390/diagnostics10050334

Geyer, K., Ellis, D. A., Shaw, H., & Davidson, B. I. (2021). Open-source smartphone app and tools for measuring, quantifying, and visualizing technology use. *Behavior Research Methods*. https://doi.org/10.3758/s13428-021-01585-7

Hale, L., & Guan, S. (2015). Screen time and sleep among school-aged children and adolescents: A systematic literature review. *Sleep Medicine Reviews, 21*, 50–58. https://doi.org/10.1016/j.smrv.2014.07.007

Hart, T. L., Swartz, A. M., Cashin, S. E., & Strath, S. J. (2011). How many days of monitoring predict physical activity and sedentary behaviour in older adults? *International Journal of Behavioral Nutrition and Physical Activity, 8*, 62. https://doi.org/10.1186/1479-5868-8-62

Katapally, T. R., & Chu, L. M. (2019). Methodology to derive objective screen-state from smartphones: A smart platform study. *International Journal of Environmental Research and Public Health, 16*(13). https://doi.org/10.3390/ijerph16132275

Klakk, H., Wester, C. T., Olesen, L. G., Rasmussen, M. G., Kristensen, P. L., Pedersen, J., & Grøntved, A. (2020). The development of a questionnaire to assess leisure time screen-based media use and its proximal correlates in children (SCREENS-Q). *BMC Public Health, 20*(1), 664. https://doi.org/10.1186/s12889-020-08810-6

Kornberg, J. (2020). Left to our own devices: Smartphone use, mental health, and academic psychiatry. *Academic Psychiatry, 44*(4), 483–486. https://doi.org/10.1007/s40596-020-01258-1

Lissak, G. (2018). Adverse physiological and psychological effects of screen time on children and adolescents: Literature review and case study. *Environmental Research, 164*, 149–157. https://doi.org/10.1016/j.envres.2018.01.015

Matthews, C. E., Ainsworth, B. E., Thompson, R. W., & Bassett, D. R., Jr. (2002). Sources of variance in daily physical activity levels as measured by an accelerometer. *Medicine & Science in Sports & Exercise, 34*(8), 1376–1381. https://doi.org/10.1097/00005768-200208000-00021

Parry, D. A., Davidson, B. I., Sewall, C. J. R., Fisher, J. T., Mieczkowski, H., & Quintana, D. S. (2021). A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nat Hum Behav*. https://doi.org/10.1038/s41562-021-01117-5

Preiss, D., & Fisher, J. (2008). A measure of confidence in Bland-Altman analysis for the interchangeability of two methods of measurement. *Journal of Clinical Monitoring and Computing, 22*(4), 257–259. https://doi.org/10.1007/s10877-008-9127-y

Rivière, F., Widad, F. Z., Speyer, E., Erpelding, M. L., Escalon, H., & Vuillemin, A. (2018). Reliability and validity of the French version of the global physical activity questionnaire. *Journal Sport Health Science, 7*(3), 339–345. https://doi.org/10.1016/j.jshs.2016.08.004

Sagelv, E. H., Hopstock, L. A., Johansson, J., Hansen, B. H., Brage, S., Horsch, A., Ekelund, U., & Morseth, B. (2020). Criterion validity of two physical activity and one sedentary time questionnaire against accelerometry in a large cohort of adults and older adults. *BMJ Open Sport Exercise Medicine, 6*(1), Article e000661. https://doi.org/10.1136/bmjsem-2019-000661

Schwarz, N., & Oyserman, D. (2001). Asking questions about behavior: Cognition, communication, and questionnaire construction. *American Journal of Evaluation, 22*(2), 127–160. https://doi.org/10.1016/S1098-2140%2801%2900133-3

Screen time Api proposal. Access date 11-06-2021. (Access date 11-06-2021). Retrieved 11-06-2021 from https://screentimeapi.com/.

Shqair, A. Q., Pauli, L. A., Costa, V. P. P., Cenci, M., & Goettems, M. L. (2019). Screen time, dietary patterns and intake of potentially cariogenic food in children: A systematic review. *Journal of Dentistry, 86*, 17–26. https://doi.org/10.1016/j.jdent.2019.06.004

Statistik, D. (2018). *It-anvendelse i befolkningen 2018*.

Tanaka, R., Fujita, K., Makimoto, K., Yakushiji, K., Tanaka, S., Motomura, G., Fujii, M., Ikemura, S., & Nakashima, Y. (2021). Development and evaluation of the sedentary behavior and light-intensity physical activity questionnaire. *Journal of Physical Therapy Science, 33*(2), 125–131. https://doi.org/10.1589/jpts.33.125

Trost, S. G., McIver, K. L., & Pate, R. R. (2005). Conducting accelerometer-based activity assessments in field-based research. *Medicine & Science in Sports & Exercise, 37*(11 Suppl), S531–S543. https://doi.org/10.1249/01.mss.0000185657.86065.98

Verbeij, T., Loes Pouwels, J., Beyens, I., & Valkenburg, P. (2021). The accuracy and validity of self-reported social media use measures among adolescents. *Computers in Human Behavior Reports*. https://doi.org/10.1016/j.chbr.2021.100090

Whiting, S., Buoncristiano, M., Gelius, P., Abu-Omar, K., Pattison, M., Hyska, J., Duleva, V., Musić Milanović, S., Zamrazilová, H., Hejgaard, T., Rasmussen, M., Nurk, E., Shengelia, L., Kelleher, C. C., Heinen, M. M., Spinelli, A., Nardone, P., Abildina, A., Abdrakhmanova, S., … Breda, J. (2020). *Physical Activity, Screen Time, and Sleep Duration of Children Aged 6-9 Years in 25 Countries: An Analysis within the WHO European Childhood Obesity Surveillance Initiative (COSI) 2015-2017* (pp. 1–13). Obes Facts. https://doi.org/10.1159/000511263

Zou, G. Y. (2013). Confidence interval estimation for the Bland-Altman limits of agreement with multiple observations per individual. *Statistical Methods in Medical Research, 22*(6), 630–642. https://doi.org/10.1177/0962280211402548