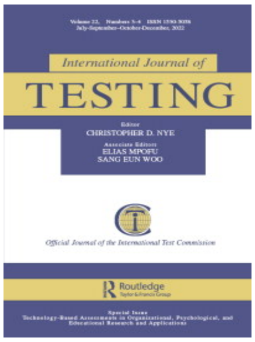

International Journal of Testing

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hijt20

Mobile sensing in psychological and educational research: Examples from two application fields

Efsun Birtwistle, Ramona Schoedel, Florian Bemmann, Astrid Wirth, Christoph Sürig, Clemens Stachl, Markus Bühner & Frank Niklas

To cite this article: Efsun Birtwistle, Ramona Schoedel, Florian Bemmann, Astrid Wirth, Christoph Sürig, Clemens Stachl, Markus Bühner & Frank Niklas (2022) Mobile sensing in psychological and educational research: Examples from two application fields, International Journal of Testing, 22:3-4, 264-288, DOI: 10.1080/15305058.2022.2036160

To link to this article: https://doi.org/10.1080/15305058.2022.2036160

Published online: 18 Nov 2022.

INTERNATIONAL JOURNAL OF TESTING
2022, VOL. 22, NOS. 3–4, 264–288
https://doi.org/10.1080/15305058.2022.2036160

OPEN ACCESS Check for updates

Mobile sensing in psychological and educational research: Examples from two application fields

Efsun Birtwistle[a], Ramona Schoedel[a], Florian Bemmann[b], Astrid Wirth[a], Christoph Sürig[c], Clemens Stachl[d], Markus Bühner[a], and Frank Niklas[a]

[a]Department of Psychology, LMU Munich, Munich, Germany; [b]Institute of Informatics, LMU Munich, Munich, Germany; [c]Department of Informatics, Technical University of Munich, Munich, Germany; [d]Institute of Behavioral Science and Technology, University of St. Gallen, Gallen, Switzerland

ABSTRACT
Digital technologies play an important role in our daily lives. Smartphones and tablet computers are very common worldwide and are available for everybody from a very early age. This trend offers the opportunity to track digital usage data for psychological and educational research purposes. The current paper introduces two research projects, the *PhoneStudy* and *Learning4Kids* that both use mobile sensing software to collect ecologically valid data on the usage of applications installed on smartphones and tablets. This usage data is used for statistical analyses, for a reward system, and to provide feedback to the study participants. The advantages and challenges of using mobile sensing compared to conventional forms of assessments, and the potential applications of mobile sensing in psychological and educational research are discussed.

KEYWORDS
digital assessment; log files; mobile apps; mobile sensing; phone study app; psychological and educational research

Introduction

Consumer electronics have become an important part of our daily lives with their features to support communication, entertainment, information acquisition, and transfer. Nowadays, almost every household has a variety

CONTACT Efsun Birtwistle Efsun.Birtwistle@lmu.de Department of Psychology, LMU Munich, Munich, Germany.

of intelligent devices from smartphones, tablets to laptop computers (Chaudron et al., 2015; Livingstone et al., 2015). In today's digital devices, there are several mobile *applications (apps)* that are available to facilitate our personal demands. Through these app tools, we can contact with others, play games, or even learn a new language.

As we use these helpful tools in daily life, we leave huge amounts of digital footprints, reflecting our actions, preferences and personal traits (Kosinski et al., 2013). At the intersection of computer science and psychology, researchers have begun to realize the potential of these data for investigating behaviors in everyday contexts with high ecological validity (Harari et al., 2016; 2018).

The methodology and the technical background of mobile sensing

Running the right software on consumer electronics allows for the automatic and unobtrusive tracking of digital user data for research purposes. This methodology enables the assessment of actual usage behavior and is called *mobile sensing* (Harari et al., 2016; Lane et al., 2010). Mobile sensing is one of the data collection tools summarized under the umbrella of *ambulatory assessment* which more generally describes the investigation of naturally occurring user behavior in a wide range of real-life contexts (Conner & Mehl, 2015). In addition to passive logging, by means of mobile sensing, active logging, by means of ecological momentary assessment, or daily diaries is becoming increasingly popular in research (Conner & Mehl, 2015).

Although mobile sensing has been described as a highly ecologically valid method (Miller, 2012), it is not yet widely used in educational and social sciences. Instead, previous research often applied self-report questionnaires to assess digital media usage (e.g., Broughton et al., 2019). Similarly, in experimental designs with digital media, often intention-to-treat analyses technique (i.e., analysis of all study groups after a randomized assignment of participants regardless of participant characteristics, the intervention type, and intervention fidelity) is applied to avoid biased comparisons among the groups that are receiving treatment (e.g., Maertens et al., 2016). However, self-reports or dairies have been found to be biased in terms of response tendencies such as social desirability, incorrect memory processes or other biases (Ziegler & Buehner, 2009). Per-protocol analyses that consider how often specific apps had actually been used provide a much clearer picture of intervention fidelity and thus of the effectiveness of interventions (cf. O'Donnell, 2008).

Custom research apps represent the core of mobile sensing. These tracking apps can be installed on participants' portable devices, such as tablets, smartphones, or laptop computers which offer access to a wide range of logging data produced by its user that can be unobtrusively collected in the background (Harari et al., 2016). Tracking apps have been developed for diverse operating systems among which Android and iOS are the most frequent ones (e.g., Beierle et al., 2018; Montag et al., 2014; Stachl et al., 2020; Wang et al., 2014).

Depending on data types and the programming of the specific tracking app, data can be logged at regular intervals (e.g., GPS position every 15 minutes) or whenever changes occur [1] (e.g., Wi-Fi connection was lost). In addition, device-own usage statistics can be used [2] (e.g., to track app *usage data*, in other words *usage time data*, which provides information on how often the apps were used by means of time such as minutes). Logged data can be encrypted and synchronized on a regular basis to the backend server. Alternatively, smarter, more energy efficient synchronization modes such as transferring data only if Wi-Fi is available and/or during night are also possible. Before the logged raw data can be used for any analysis, it has to be pre-processed. One option is on-device pre-processing and it means that raw data is pre-processed instantly on the device before it is transferred to the backend server. For example, audio files can be classified into silence versus noise versus voice instead of saving the raw audio information (e.g., Harari et al., 2020). In contrast, offline pre-processing describes transferring raw data to the backend server and retrospectively extracting meaningful variables such as exact usage times.

Mobile sensing in field research

Mobile sensing can benefit different areas of research focusing human behavior in daily life. Various behavioral and situational data can be obtained through mobile sensing technology (Harari et al., 2015; 2017). Some of the examples include obtaining; 1) daily activities, such as physical activity through accelerometer, GPS, Wi-Fi, or barometric measures, 2) social-interaction behaviors through microphones, Bluetooth, or phone logs (SMS, phone calls), or 3) daily location routines or mobility patterns through cell phones, e.g., traveling, spending time at work or a restaurant (Berke et al., 2011, Wahle et al., 2016; Farrahi & Gatica-Perez, 2008; Harari et al., 2015; Miluzzo et al., 2008). These behavioral and situational information collected via mobile sensing have been suggested

[1]https://developer.android.com/guide/components/broadcasts

[2]https://developer.android.com/reference/android/app/usage/UsageStatsManager

to be useful in different research areas ranging from computer science to personality science or clinical psychology.

Research in computer science

Mobile sensing technology requires an app that is running in the background of a mobile device such as smartphone or tablet. It reads internal or external sensors of the device and reports acquired data to a web server system. Such an app ideally does not interfere with other running apps in the operation system of this device (Lane et al., 2010). Further, as the sensor app is continuously running in the background of the device and tracks device activities, user security and energy consumption of the device are considered as critical technical challenges (Christin et al., 2013; Macias et al., 2013). Some theoretical and applicable solutions were proposed to overcome these problems (security, energy consumption: Macias et al., 2013; Ben Abdesslem, Phillips, & Henderson, 2009; Wang et al., 2018).

Research in personality psychology

Mobile sensing was used to understand to what extend self-reported personality traits (e.g., Big Five personality traits) and individual differences in user behaviors can be predicted and explained (Harari et al., 2020; 2020; Mønsted et al., 2018; Stachl et al., 2017; 2020). Self-reports of personality traits (e.g., Big Five Structure Inventory, BFSI; Arendasy, 2009) as well as fluid intelligence and demographic characteristics (e.g., age and gender) were compared with behavioral measures obtained by an Android logging app that was designed to track user activities on smartphones (Stachl et al., 2017). The activities included the usage frequency of several apps (e.g., communication, photography, games, music, etc.). The results showed that increasing and decreasing app usage behavior is associated with personality traits, fluid intelligence and demographic factors. For instance, users with high extraversion scores (i.e., people who have high scores of being outgoing and social) showed a more frequent use of photography and communication apps, whereas users with high conscientiousness scores (i.e., people who have high scores of being self-disciplined and organised) showed a lower usage of gaming apps.

Research in clinical psychology

The evolution of mobile sensing also allows the collection of data in the context of the health and well-being industry (Cornet & Holden, 2018).

In their review, Cornet and Holden (2018) described the application of mobile sensing in mental health studies (e.g., depression, bipolar, schizophrenia) and other well-being factors (e.g., sleep quality, stress levels). Whereas some studies integrated mobile sensing into personalized feedback to support well-being, other studies used mobile sensing to monitor health and well-being behavior (Saeb et al., 2016; Thomée, 2018). Through the sensors of wearable devices such as activity trackers or smartphone applications, it is possible to monitor or collect mental health data, which has the potential to understand clinical samples (i.e., patients who show depression or social anxiety symptoms) and support personalized interventions (Chow et al., 2017; Wahle et al., 2016).

Overall, mobile sensing enables the collection of generalizable and ecologically valid data and may have the potential to benefit different fields of research. Although technological tools are slowly replacing traditional methods in educational research, up until now the usage of mobile sensing has not been applied as a widely used assessment tool.

Application of mobile sensing in psychological research: *the PhoneStudy*

Within the framework of an ongoing interdisciplinary research project in psychology, computer science, and computational statistics, Stachl et al. (2020) have developed an Android app called *PhoneStudy* to collect usage data on Android devices (e.g., smartphones and tablets). Using this approach, three studies were conducted between 2014 and 2018 (Schoedel et al., 2018, Schuwerk et al., 2019; Stachl et al., 2017). In total, 743 adults participated in these studies for at least 30 days.

After the installation of the app, smartphone events such as app and phone usage, notifications, screen, device, and connectivity status were logged with respective timestamps whenever they happened. Depending on the actual Android version, GPS was logged every 10 to 20 minutes. If participants were connected to Wi-Fi, the logged data was transferred to the server which is encrypted through Secure Sockets Layer (SSL). The resulting log files were in tabular data format, which means that each usage event was written in one row and specified by the respective timestamp, the participant's ID, and further details depending on the event.

For instance, notifications and apps were specified by the app's name, calls were specified as incoming, outgoing, or missed, and GPS data was specified by longitude and latitude. Pre-processing of data was conducted offline. For each research question at hand, variables were defined and extracted from the abstract logging data. Based on the variety of resulting behavioral variables, individual differences in smartphone usage were investigated by using both classical and algorithmic modeling techniques (Schoedel et al., 2018, Schuwerk et al., 2019; 2020; Stachl et al., 2017). For instance, the number of incoming calls for each of the 30 study days was computed and then averaged to the daily mean of incoming calls. In summary, about 16,000 individual smartphone usage behaviors were extracted, which could be assigned to the categories of communication, music, app usage, mobility and day-night patterns (Stachl et al., 2020). In particular, the interest was on whether self-reported personality traits manifest in smartphone usage behaviors (Stachl et al., 2017) and whether smartphone usage, in return, can be used to predict personality traits such as the big five or sensation seeking (Schoedel et al., 2018; Stachl et al., 2020). Consequently, the *PhoneStudy* project highlights that mobile sensing can be used in psychological research to collect ecologically valid data on mobile device usage and to investigate the associations with psychological variables such as personality traits. Table 1 shows a summary of findings of the *PhoneStudy* project and reveals how real-world behavioral data can be collected through mobile sensing by highlighting the potential use of this methodology for new insights into inter- and intra-individual differences in psychological research.

Application of mobile sensing in educational research: *Learning4Kids*

Mobile apps have become popular tools to foster children's learning and development. There is a growing interest to download educational apps for children especially in the age category of toddlers and preschoolers (Shuler, 2012). Early learning apps were reported to be the most popular selection among all educational apps and importantly, preschool children could benefit and learn from these apps (e.g., improvement in literacy skills: Chiong & Shuler, 2010; Judge et al., 2015). The exponential growth of digital media usage in households and at schools is also shifting the focus of educational researchers from using traditional assessment methods to more technological methods. A promising approach is to integrate mobile sensing into educational assessments. However, the application of mobile sensing has not been widely used in educational research so far. With its potential to provide structured quantitative data about the usage times of educational apps, and a feedback system that is based on actual user data, two research projects, *Learning4Kids* and *PhoneStudy* came together to apply a new approach and a new methodology in educational research setting.

Young children often engage with computers and internet-based activities. Parents and educators consider digital education important in their support of young children's learning (Papadakis & Kalogiannakis, 2017).

Table 1. Overview of studies in the *PhoneStudy* project.

References	Research Question(s)	Behavior(s) under Investigation	Central Personality Construct(s) under Investigation	Summary of Findings
Stachl et al. (2017)	Investigation of the degree personality traits predict frequency and duration of app usage behaviors.	Frequency and duration of categorial app usage (e.g., communication, games, transportation).	Big five personality factors (openness, conscientiousness, extraversion, agreeableness, emotional stability) and respective facets, demographics, and fluid intelligence; assessed via self-reports.	Self-reported personality traits on factor level and facet level, fluid intelligence, and demographic variables are manifested in a range of different categorial app usage behaviors on smartphones.
Schoedel et al. (2018)	Prediction of a biopsychological personality trait based on behaviors extracted from smartphone data using machine-learning models.	222 behavioral markers of sensation seeking that were derived from existing research literature and preregistered.	Sensation seeking; assessed via self-reports.	One of several compared machine-learning models was able to predict individual sensation seeking scores on average slightly above chance.
Schuwerk et al. (2019) *	Investigation of autistic traits and their relation to daily mentalizing and social interaction.	Social interaction (e.g., social media app use, calling, text messages).	Autistic traits; assessed via self-reports.	Autistic traits were negatively correlated with communication behavior. No associations were found between autistic traits and social media usage, or social network size.
Harari et al. (2020) +	Investigation of individual differences in daily social behavior patterns.	Frequency and duration of messaging app use, social media app use, text messages, and calls.	Extraversion; assessed via self-reports.	Daily calling and texting tendencies were positively correlated. Extraversion was related to texting behaviors.
Stachl et al. (2020)	Prediction of personality traits based on behavioral patterns extracted from smartphone data using machine-learning models.	About 16,000 behavioral patterns in six different domains: communication and social behavior, music consumption, app usage, mobility, overall phone activity, day-/night activity patterns.	Big five personality factors (openness, conscientiousness, extraversion, agreeableness, emotional stability) and respective facets; assessed via self-reports.	Machine-learning models were able to predict openness, conscientiousness, and extraversion on a factor- and facet-level, and emotional stability on some facets. Agreeableness could not be predicted.

(*Continued*)

Table 1. Continued.

References	Research Question(s)	Behavior(s) under Investigation	Central Personality Construct(s) under Investigation	Summary of Findings
Schoedel et al. (2020)	Three-part research question around individual differences in day-night-behavior patterns:			
	1) Manifestation of "morning larks" and "night owls" in smartphone usage patterns.	Behavioral indicators for circadian preferences.	Chronotype (as it was not assessed as self-report, an unsupervised machine-learning approach was applied).	Detection of non-discrete groups of individuals with similar smartphone usage patterns.
	2) Relationship between day-night patterns and personality traits.	Behavioral indicators for sleep-wake timing (e.g., last/first smartphone usage event of the day; alarm clock app usage).	Big five personality factors (openness, conscientiousness, extraversion, agreeableness, emotional stability).	Associations between smartphone-sensed indicators for day-night behavior and conscientiousness were found.
	3) Relationship between day-night patterns during the week and personality traits and day-night patterns on weekends.	Duration of smartphone usage inactivity during nights on working and weekend days.		Individuals' nightly inactivity on weekends was mainly related to their average level of nightly inactivity. Intra-individual or personality effects were not found.

Note. * The scope of this study is much broader and includes ecological momentary assessment besides mobile sensing; not to go beyond the scope, the presented information is therefore restricted to the mobile sensing parts. + In this study, the *PhoneStudy* app was only used for parts of the data collection; the presented information is therefore restricted to this part.

Even though mobile media devices (i.e., tablets, smartphones) and apps are not educational themselves, they may be used to promote children's education by creating playful, engaging and meaningful learning environments (cf. Hirsh-Pasek et al., 2015).

In *Learning4Kids* (ERC Starting Grant 801980), such educational apps that target early learning experiences for literacy (e.g., vocabulary, phonological awareness; Niklas & Schneider, 2013), numeracy (e.g., counting abilities, number knowledge; Niklas et al., 2016), and general cognitive competencies (e.g., memory, executive functioning) are provided for families (see also Niklas et al., 2020). Here, app usage times are collected via mobile sensing for later analysis to be able to understand how much time children and their families actually spend using each of the provided apps. The usage time data provides the potential to understand child/ family app selection and the amount of time spent in apps that have different educational purposes.

Learning4Kids is designed to provide several apps (games, ebook reader and music player) to children and their families each month for the duration of about ten months (see Niklas et al., 2020). The *PhoneStudy* app was adapted to run constantly in the background on Android 9 and 10 tablets to track the usage time in the context of the home learning environment (Niklas & Schneider, 2017). It was configured automatically to log the events on the tablet together with a timestamp and the name of the application the event originates from.

In particular, two events were deemed important: 1) an application activity is moved to the tablet foreground, and 2) an application activity returns to the background again. These event logs are then transferred to a server via a secure connection every hour. In the next step, app usage times were used to provide feedback to parents and children about their tablet usage (for the importance of feedback in educational contexts, see Hattie, 2009). For instance, once a week, all logs stored on the server were aggregated, and feedback was provided to the study participants. During this process, the collected events were used to calculate the actual usage time in this week per pseudonym, and per app as the first step. Afterwards, the aggregated data was used to compute the average usage time of all pseudonyms in this week in total, and per application. Finally, these data were used to generate personalized feedback reports about the last seven days as .pdf documents, which were sent back via the server system to the tablets and could be accessed by the participating families.

In addition, the usage data was also used for an independent reward system. For every 30 minutes of tablet usage, a sticker award in the form of an animal was given to the participating children in a reward app. As from an educational perspective (Feierabend et al., 2014; 2016) young children are not recommended to use tablets longer than one hour a day, this information was explicitly stated in the personalized feedback report provided to the parents and the number of awards was restricted to a maximum of two awards per day. Here, the reward app used the *usage time data* that was measured by the *PhoneStudy* app in order to provide children with their rewards.

Here, we report on the initial usage time data in the Learning4Kids study that had been extracted. From the beginning of the intervention, the *PhoneStudy* app recorded daily user activities and app usage times separately for each app and each participant tablet. In Table 2, we provide the initial descriptive findings of *usage time data* per app in detail. As can be seen, tablet interaction was traced through mobile sensing and information about how long or how often children had been using the various intervention apps was recorded. In addition, children's competencies development will be assessed in this longitudinal study (e.g., literacy, numeracy, cognitive) and the usage data will also be used to analyze the potential effect of the tablet intervention on children's competencies development. With this approach, it will be possible to check whether regular/irregular app usage is associated with better/worse child competency scores and whether a certain dosage of app usage is necessary to find intervention effects.

Mobile sensing as an assessment tool

An important question is whether mobile sensing can be used as a single assessment method to replace traditional self-reports or whether it should be used as a supportive measurement tool together with self-reports. In our view, mobile sensing may already be used as a single assessment method, however, still many challenges need to be addressed in terms of security, privacy, reliability, validity and replication. With the current developments in mobile sensing technology and rising interest in using this technology as an assessment tool among researchers, we believe that these challenges will soon be addressed.

Using mobile sensing data to extract behavioral measures and latent variables

With mobile sensing, behavioral characteristics such as reaction times (RTs), app usage times, or error rates (ERs) can directly be extracted to understand the manifestation of behavior in real-life situations. Collecting data in traditional experiments or settings can be difficult, time consuming, and potentially be expensive (e.g., software licences). In comparison, the sampling of behavior through mobile sensing provides the potential to collect data in real time from participant's daily life. It avoids

Table 2. Overview of descriptive data of app usage times in ***Learning4Kids***.

Intervention group (*N* = 152)	Apps+	Usage times* Min	Max	Mean	Std. Deviation
Literacy (*N* = 57)	Memory (letters)	0.57	366.97	101.09	92.04
	Letter sorting	0.27	394.23	89.63	76.98
	Letter drawing	0.35	226.68	78.65	54.18
	Painting with letters	17.05	554.68	177.44	121.62
	Letter-Domino	0.00	162.78	29.04	38.02
	Initial letter sounds	0.00	302.55	84.28	80.59
	Snakes & Ladders (letters)	0.00	271.90	91.65	76.60
	Finding pairs (rhymes)	0.00	160.87	33.06	36.32
	Sentence understanding	0.00	150.25	38.12	41.85
	Find the vowels	0.00	357.93	37.33	56.60
	Word puzzle	0.00	290.62	30.70	65.52
	Magic potion (sounds)	0.00	111.37	28.01	29.29
	Pixi-a day of adventure	0.00	77.67	12.02	18.15
	Pixi-farm	0.00	434.67	38.06	68.91
	Pixi-start reading	0.00	71.57	16.91	19.30
	Pixi-Ottokar	0.00	481.02	72.82	109.34
Numeracy (*N* = 60)	Memory (numbers)	0.00	815.63	89.81	143.67
	Number drawing	0.45	617.57	92.95	102.62
	Painting with numbers	13.82	818.05	216.80	189.43
	Number sorting	0.00	808.37	85.41	121.91
	Build a number rocket	4.38	699.68	89.33	101.05
	Collecting nuts (numbers)	0.00	304.60	50.36	60.91
	Snakes & Ladders (numbers)	0.00	1124.93	157.68	221.71
	Tap it! Numbers	0.00	491.95	45.04	71.81
	Mathemarmite	0.00	1449.37	153.44	251.61
	Measurement app	0.00	556.82	84.36	114.12
	Finding pairs (numbers)	0.00	167.77	24.08	33.60
	Count and compare	0.00	334.58	33.71	62.15

subjective biases when compared to asking participants about their prior behaviors. Further, the data can be collected repeatedly under several different situations and analyzed across varying contexts (e.g., time of the day, day of the week, at home, at work, etc.) and more precise behavioral data can be collected in comparison to self-reports. On the other hand, mobile sensing comes with potential limitations. For example, participants may still fake behavior or be inattentive while using apps. Further, the awareness of mobile sensing usage could lead to participants affecting the data acquisition intentionally or unintentionally by being over- or underactive with their mobile device (please see section: *Potential limitations as an assessment tool*).

In comparison, latent psychological variables cannot be directly obtained through mobile sensing but through extracted behavioral data. For example, tracking the usage of social network apps, activity apps, or social interaction through messaging may give insights about constructs such as social tendencies and personality traits (Harari et al., 2020; 2020; Stachl et al., 2017).

Table 2. (Continued).

Intervention group ($N = 152$)	Apps+	Usage times* Min	Max	Mean	Std. Deviation
	Counting the balloons	0.00	224.95	18.12	36.44
	Number-domino	0.00	434.25	43.56	78.56
	Connect the number dots	0.00	213.93	23.10	41.08
	Learn the clock	0.00	276.30	23.16	48.25
	Count and sort objects	0.00	772.42	43.38	116.78
	Categorize numbers	0.00	153.03	16.29	29.65
Control ($N = 35$)	Memory (colors)	0.00	177.83	53.62	45.71
	Color sorting	0.00	131.95	17.87	28.35
	Color-domino	0.00	1778.22	114.29	342.45
	Build a color rocket	6.17	295.26	65.15	55.75
	Fish-maze	0.00	821.83	89.67	157.85
	Sagomini Forest Flyer	0.00	495.97	118.61	132.07
	Sagomini Friends	0.00	715.20	124.77	174.48
	Shape drawing	1.28	160.72	54.37	45.92
	Snakes & Ladders (colors)	0.00	345.10	67.29	81.30
	Connect the color dots	0.00	56.97	11.44	17.53
	Collecting nuts (colors)	0.00	126.77	38.96	38.30
	Piano	0.00	78.62	10.23	20.27
	Finding pairs (colors)	0.00	105.15	22.90	27.85
	Animal maze	0.00	131.95	17.88	28.36
	Bird tower	0.00	1778.22	114.29	342.46
	Animal puzzle	8.10	1142.23	132.68	204.50
	Tap it! Colors	0.00	96.37	19.68	26.50
	Painting with colors	19.92	872.42	186.30	186.88

Note. Usage time duration is a total of 168 days (Learning4Kids project-phase 1); the unit of time is minutes.

+*All apps were available for a different duration. The apps that are uploaded in the first month (e.g., memory (letters), memory (numbers), memory (colors)) remained longer to-be-played than the apps uploaded in the last months (e.g., magic potion sounds, learn the clock, animal maze).*

Mobile sensing data offers a great variety of data types, complex data modeling techniques such as machine learning may be used to build models that cluster data and provide hints at latent variables. With this approach, a number of study variables can be tested via several apps in different contexts (e.g., reading skills at home, social media usage while traveling, and so forth) and algorithmic modeling techniques can be used to infer psychological variables (Schoedel et al., 2018; Stachl et al., 2017). Currently, supervised machine learning is used for such modeling approaches, however, it is still in discussion whether this approach is better than classical questionnaires or not (see Stachl et al., 2020).

Measurement quality

Reliability and validity are complex issues to address in mobile sensing. Reliability measurement is of great importance to evaluate behavioral outcomes. Here, inter-device reliability (i.e., the quality of measurements across different study devices) and intra-device reliability (i.e., the quality of repeated measurements within the same device) are important aspects to consider. In their systematic review, Evenson et al. (2015) evaluated inter-device reliability and validity of wearable trackers. They found high validity of certain constructs such as counted steps and lower validity for energy expenditure and sleep, and reported high inter-device reliability for wearable sensing devices. However, intra-device reliability was difficult to assess due to the dynamic structure and variability of everyday activities. Consequently, variations over time may be linked to error variance, but also to actual variability in the assessed behavior (Rat Für Sozial- und Wirtschaftsdaten, 2020). Here, it would be helpful if the manufacturers and users of sensors provided information about the reliability. Further, when mobile sensing data such as audio or video data is coded, interrater-reliability should be provided.

Various facets of validity should be considered in mobile sensing research. Face validity may be assumed (e.g., communication and socialization behavior as indicators of sociability) or may be fairly difficult to establish (e.g., very complex patterns of behavior as predictors of a certain construct). Here, it is important to identify the extent to which the obtained mobile sensing data are actually grasping the latent construct of interest (e.g., construct validity) and whether they are fully representative of it (e.g., do features on calling, texting, and social app use actually represent the construct sociability; content validity). Researchers need to consider whether these data are associated with related (i.e., convergent validity) and unrelated (i.e., divergent validity) constructs, and with relevant outcomes (i.e., criterion validity) in the expected direction and magnitude. In a similar way, measures of incremental validity could be obtained by comparing the additional predictive value these metrics provide over traditional measures of a construct (e.g., sensed sociability data in addition to self-reports as predictors of job success). Further, correlational patterns between the individual indicators and mobile sensing data may be investigated for an initial estimate of structural validity.

Constructs of interest in mobile sensing can be information about *personality* by using app usage frequency of social apps (Stachl et al., 2017; 2020), information about *anxiety or stress* through heart rates when measured by sensor-based devices (Shcherbina et al., 2017), and information about *math and reading abilities* and further progress through behavioral data (e.g., RTs, ERs, usage times) obtained from learning apps that focus on academic skills. Although there are further constructs that can readily be assessed with mobile sensing, other constructs such as emotions, decision making, memory processes, and parent-child relationship will be difficult to measure with mobile sensing.

In general, mobile sensing may provide ecological valid data from natural everyday life of participants, but the quality of physical measurements such as the sensors of the device plays an essential role. In order to address technical issues, a more controlled laboratory or observational validation is still necessary (Shcherbina et al., 2017). Further, selective usage of the device (e.g., not using the pedometer while playing soccer) or shared usage with other people may have a negative impact on external and internal validity.

Further potential applications of mobile sensing

Mobile sensing as a new method for data collection has only lately become applicable in psychological research and the growing number of empirical studies using this method underline its potential usage and benefit (Harari et al., 2016; 2020; Miller, 2012). More and more researchers in different fields have become aware of using mobile sensing as a methodological tool for assessments. Digital devices such as smartphones are meanwhile widely spread in most age groups and social classes around the world and offer new opportunities for educational and psychological assessment, some of which are introduced in the following.

Application 1: Assessment of personality development and individual differences in young children

Assessing and identifying young children's personality and individual differences is one of the hot topics in child psychology that will support our understanding of children's socio-emotional, cognitive and behavioral development. Young children's personality assessment tools are traditionally based on qualitative observations, child statements, or surveys that are given to teachers, parents, or other caregivers and that are subjected to biased interpretations and answers (Eder, 1990; Halverson et al., 2003; Wilson et al., 2013). Previously, Lamb et al. (2002) had shown that the big-five personality traits and personality development of young children can be assessed in a longitudinal study design. As mobile sensing technology yielded its usage to asses personality traits for adults, we believe that one application would be to assess behavioral indicators of young children's personality development through tablet usage time at home or at schools. Extending on the findings of Stachl et al. (2020; see above), studies can be conducted to assess personality traits of young children and the frequency of app usage and app/game preference, to understand the association between children's developing personality and activity levels in certain apps/games. However, one limitation of this approach is that children are not recommended to spend too much time with digital media (Feierabend et al., 2014, 2016), and mobile sensing should only be used together with traditional assessment methods to achieve a more comprehensive picture.

Application 2: Assessment of socio- emotional development and media exposure

Children's socio-emotional competence development is considered to be an important factor for their future academic competencies and success as individuals in daily life (Denham & Brown, 2010; Wirth et al., 2020). Rasmussen et al. (2019) showed that mobile apps that provide prosocial content positively influence children's socio-emotional skills. Playing with app-based prosocial games helps children to regulate their emotions. Children also use emotional regulation strategies as a result of their experience and knowledge derived by such apps. Consequently, mobile sensing methodology can be used to investigate the associations between the usage time of the apps that are based on prosocial behavior and socio-emotional competence development in young children. For instance, studies may investigate the interaction of children with such apps (e.g., based on usage times and RTs) and their association with socio-emotional competency measures.

Application 3: Monitoring academic competencies in classrooms

Integrating digital media in learning environments is becoming more and more popular in education, in particular, since the worldwide COVID-19 pandemic. Currently, tablet technology is often used to "assist" children's formal learning activities with the advantage of flexibility of movement and its potential to provide individualized content according to children's knowledge and abilities (Walling, 2014). One practical usage of mobile sensing in educational context would be to monitor and assess children's classroom performance individually and in learning groups through tablet activities and give individualized or formative feedback to children and their parents based on the usage data that is derived from mobile sensing technology. It is, however important for such technology to design data visualization and illustration for academic curriculum in pedagogical settings and also take data privacy of students into account.

Application 4: Assessment of social media addiction

Social media has changed our way of communicating and our media usage. Research investigating high levels of web- or app-based media activity is getting attention in recent years (Leung & Chen, 2021). Here, self-reports or diagnostic questionnaires are commonly used methods to assess a potential social media addiction. In their review, Leung and Chen (2021) reported that tools to accurately diagnose media addiction in terms of media usage duration (i.e., whether the addiction behavior is long-lasting or temporary) and the precise cutoff point is crucial for future assessments and to get further insights about social media addiction. Mobile sensing may be used as a potential assessment tool to tackle diagnostic limitations of usage time duration and cutoff point in media addiction research. However, in such research, the privacy issues and the possibility of participant's sense of being tracked should be considered.

These four potential applications of mobile sensing in psychological or educational research are only some examples of further potential applications in which mobile sensing could be applied.

Advantages of mobile sensing as an assessment tool

Traditional assessment methods are based on self-reports that are evaluating behaviors. Self-reports include surveys or questionnaires that assess people's attitudes, thoughts, feelings, and beliefs. Even though these assessments are useful to understand attitudes and are widely used in research, they also have disadvantages. As they are based on subjective self-evaluation, the validity may be low (Hough et al., 1990) due to participants' biased or socially acceptable answers (i.e., social desirability), and potential misinterpretations of the written questions. In order to tackle such problems, mobile sensing can be used as a new state-of-the-art methodology to capture people's experiences through the sensors of mobile devices. Correlating data of traditional self-report methods and mobile sensing data through monitoring people's activity everyday has the potential to improve accurate and objective measurement and reduce the noise derived from biased responses.

In the following, we provide some examples for potential advantages: 1) Mobile sensing provides a direct assessment of participant activities and responses. For instance, as it could possibly happen in traditional assessment methods of behavior, instead of asking a parent to report, "how often does your child play with a smartphone or tablet during the week" and receiving an answer that would be based on parents' observations or positive/negative bias of their children's weekly interaction with the device, mobile sensing would measure the actual weekly usage time. 2) Mobile sensing may help to overcome wrong answers that participants may give due to their false memories. For instance, study participants may be asked about the times when they used the device most often, which content they used most often, etc. Here, mobile sensing would provide accurate data that is free from misjudgement. 3)

Mobile sensing may be used for behavioral data acquisition such as RTs and ERs. To understand how active or passive participants were by using each app. RTs (e.g., tapping/drag & drop responses) and/or ERs analyses from the onset until the offset of one level may be performed and thus faking behavior may be identified (e.g., through very short or very long RTs or high ERs). 4) Within-subject progress of the same apps can be recorded: Assessing usage time data and ERs within one app and per subject would give the possibility to follow the app progress of participants over time. This assessment would provide information about the engagement of participants and may also be used for adaptive testing and training.

Potential limitations of mobile sensing as an assessment tool

Some limitations apply to mobile sensing just as for any other assessment tool such as social desirability or faking (Paulhus, 2017). For instance, participants may intentionally use the study device for longer or shorter hours than they usually would in daily life and manipulate the recorded data. In adult research, participants give their consent to be tracked while using the digital devices, and therefore some questions may emerge about 1) how often they are consciously aware that their device activities are being tracked at the background, and 2) whether the sense of being tracked would reduce or disappear over the course of the study, and 3) whether faking tendencies would increase or reduce based on participant's sense of being tracked. For young children, the sense of being tracked by others while using the internet or other technological systems/devices can be considered to be very low (Chaudron et al., 2015; Edwards et al., 2018). However, the risk of misuse for adults (Montag et al., 2014) is considered as an unresolved limitation. Consequently, the potential effects of social desirability or faking in mobile sensing should be investigated and elaborated in future studies amongst different age groups.

The necessity to store and process large amounts of data is also an important limitation of mobile sensing. To apply mobile sensing, specific apps and servers need to be installed and large amounts of data that are generated need to be safely stored and processed.

Further, due to the battery consumption with active and rapid sensing apps that are running in the background, the battery of smartphones/tablets needs to be charged regularly. If they are not charged in time, the devices are more likely to shut down and no data can be accessed until the user decides to use the device actively again.

Another limitation is potential technical problems that may require immediate attention and solution with remote technical support. Researchers should prepare their resources to provide remote support to the participants and avoid technical problems that may affect the data collection process.

Unresolved questions in mobile sensing application

There are several unresolved questions and concerns for the usage of mobile sensing. The first one is on the reliability of extracted measures. The accuracy of test questions and the consistency of assessment results are important factors contributing to the reliability of an assessment. It is not yet clear whether traditional notions of reliability apply to mobile sensing, and if yes, how reliability could be estimated (e.g., with intra-class-correlations; cf. Harari et al., 2020, or with test re-test reliability; cf. Brouillette et al., 2013).

The second one is the validity of mobile sensing when compared to traditional measurements and whether sensing based measurement reflects the tested construct or encompasses other elements. Montag et al. (2014) revealed external validation and replication of the tracked personality variables (e.g., call behavior and extraversion) as the first validation data of "psycho-informatic" research suggesting the generalizability of the results. However, to ensure validity, a consistent measurement and generalizability of mobile sensing data across different mobile devices or within the same device are important technical issues to address in future studies. In addition, for each new application of mobile sensing more traditional assessment methods should also be used to validate the data.

The third one is the concerns on biased samples and replication problems. Harari et al. (2020) pointed out that the use of unrepresentative samples that are potentially resulting from self-selection biases, and the replicability of research findings are important challenges that remain to be addressed for the use of mobile sensing methodology. Further, the discussion about personal data security applies also, and particularly, to mobile sensing as here, all kind of personal usage data of private devices is assessed and stored in apps and on server systems. Consequently, best practices need to be applied to treat such data appropriately (cf. Harari et al., 2020).

The forth one concerns the ethics about security and safety in sensing-based research, especially for studies collecting data from vulnerable populations (e.g., children, clinical patients; Fuller et al., 2017). Breslin et al. (2019) and Fuller et al. (2017) suggested that ethics should be clarified in terms of the security and privacy during data collection, analyses and interpretation. For example, throughout the data collection and transmission process, device security should be provided with encrypted data. Active monitoring should be avoided and participants should be able to "pause" data collection if they want to. Legally bound disclosures should be included and clearly explained in the consent form, and technical support services should be remotely and easily accessible for the participants through the device they are using. In terms of privacy and the notion of surveillance of private life, no personal data should be stored in the tablets to avoid private data leakage (e.g., in case of compromised Wi-Fi networks, stolen tablets), and the data must be stored according to specific regulations such as the EU General Data Protection Regulation (GDPR)[3]. Importantly, participants must be informed about where and how their data is stored and how adherence to the GDPR is ensured (i.e., whether the researcher institution complies with the guidelines of data privacy for data storage).

Future prospects and conclusion

In our view, the application of mobile sensing in empirical research offers many prospects in different research fields as this novel methodological approach will provide researchers with more ecologically valid data. Further, in intervention studies that use electronic media, it may support intervention fidelity (i.e., whether an intervention is carried out as its intended) and thus may help researchers to come to more accurate findings and implications. Consequently, we believe that despite all the limitations and challenges discussed above, this approach will open a new page in psychological and educational research, and we hope that the information provided in this paper may inspire more social scientists to use mobile sensing methodology in their own studies.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Learning4Kids has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no 801980).

[3]General Data Protection Regulation (GDPR). https://gdpr-info.eu/. Accessed 14 Dec 2021; Regulation, G. D. P. (2016). Regulation EU 2016/679 of the European Parliament and of the Council of 27 April 2016. Official Journal of the European Union.

ORCID

Efsun Birtwistle http://orcid.org/0000-0002-6958-588
Ramona Schoedel http://orcid.org/0000-0001-7275-0626
Astrid Wirth http://orcid.org/0000-0003-1126-141X
Clemens Stachl http://orcid.org/0000-0002-4498-3067
Frank Niklas http://orcid.org/0000-0002-3777-7388

References

Arendasy, M. (2009). *BFSI: Big-five Struktur-Inventar (Test & Manual)*. SCHUHFRIED GmbH.

Ben Abdesslem, F., Phillips, A., & Henderson, T. (2009). Less is more: energy-efficient mobile sensing with senseless. Proceedings of the 1st ACM workshop on Networking, systems, and applications for mobile handhelds. *Barcelona, Spain, Association for Computing Machinery*, 6162. https://doi.org/10.1145/1592606.1592621

Beierle, F., Tran, V. T., Allemand, M., Neff, P., Schlee, W., Probst, T., … Zimmermann, J. (2018). TYDR–Track your daily routine. Android App for tracking smartphone sensor and usage data. In *2018 IEEE/ACM 5th International Conference on Mobile Software Engineering and Systems (MOBILESoft)* (pp. 72–75). IEEE.

Berke, E. M., Choudhury, T., Ali, S., & Rabbi, M. (2011). Objective measurement of sociability and activity: Mobile sensing in the community. *Annals of Family Medicine*, *9*(4), 344–350. https://doi.org/10.1370/afm.1266

Breslin, S., Shareck, M., & Fuller, D. (2019). Research ethics for mobile sensing device use by vulnerable populations. *Social Science & Medicine (1982)*, *232*, 50–57. https://doi.org/10.1016/j.socscimed.2019.04.035

Broughton, A., Daly, M., Marx, N. J., Nieuwoudt, M., Le Roux, D. B., & Parry, D. A. (2019 *An exploratory investigation of online and offline social behaviour among digital natives* [Paper presentation]. The South African Institute of Computer Scientists and Information Technologists 2019, (pp. 1–10). https://doi.org/10.1145/3351108.3351146

Brouillette, R. M., Foil, H., Fontenot, S., Correro, A., Allen, R., Martin, C. K., Bruce-Keller, A. J., & Keller, J. N. (2013). Feasibility, reliability, and validity of a smartphone based application for the assessment of cognitive function in the elderly. *PLoS One*, *8*(6), e65925. https://doi.org/10.1371/journal.pone.0065925

Chaudron, S., Beutel, M. E., Donoso Navarrete, V., Dreier, M., Fletcher-Watson, B., Heikkilä, A. S., … Wölfling, K. (2015). *Young children (0-8) and digital technology: A qualitative exploratory study across seven countries* (pp. 13–18). *JRC; ISPRA, Italy*. https://doi.org/10.2760/294383

Chiong, C., & Shuler, C. (2010). Learning: Is there an app for that. In *Investigations of young children's usage and learning with mobile devices and apps* (pp. 13–20). The Joan Ganz Cooney Center at Sesame Workshop.

Chow, P. I., Fua, K., Huang, Y., Bonelli, W., Xiong, H., Barnes, L. E., & Teachman, B. A. (2017). Using mobile sensing to test clinical models of depression, social anxiety, state affect, and social isolation among college students. *Journal of Medical Internet Research*, *19*(3), e62. https://doi.org/10.2196/jmir.6820

Christin, D., Roßkopf, C., Hollick, M., Martucci, L. A., & Kanhere, S. S. (2013 IncogniSense: An anonymity-preserving reputation framework for participa-

tory sensing applications [Paper presentation]. *Pervasive and Mobile Computing*, 9(3), 353–371. https://doi.org/10.1109/PerCom.2012.6199860
Conner, T. S., & Mehl, M. R. (2015). Ambulatory assessment: Methods for studying every- day life. In R. A. Scott, M. C. Buchmann, & S. M. Kosslyn (Eds.), *Emerging trends in the social and behavioral sciences* (pp. 1–15). John Wiley & Sons. https://doi.org/10.1002/9781118900772.etrds0010
Cornet, V. P., & Holden, R. J. (2018). Systematic review of smartphone-based passive sensing for health and wellbeing. *Journal of Biomedical Informatics*, *77*, 120–132. https://doi.org/10.1016/j.jbi.2017.12.008
Denham, S. A., & Brown, C. (2010). Plays nice with others": Social–emotional learning and academic success. *Early Education & Development*, *21*(5), 652–680. https://doi.org/10.1080/10409289.2010.497450
Eder, R. A. (1990). Uncovering young children's psychological selves: Individual and developmental differences. *Child Development*, *61*(3), 849–863. https://doi.org/10.2307/1130969
Edwards, S., Nolan, A., Henderson, M., Mantilla, A., Plowman, L., & Skouteris, H. (2018). Young children's everyday concepts of the internet: A platform for cyber-safety education in the early years. *British Journal of Educational Technology*, *49*(1), 45–55. https://doi.org/10.1111/bjet.12529
Evenson, K. R., Goto, M. M., & Furberg, R. D. (2015). Systematic review of the validity and reliability of consumer-wearable activity trackers. *International Journal of Behavioral Nutrition and Physical Activity*, *12*(1), 1–22. https://doi.org/10.1186/s12966-015-0314-1
Farrahi, K., & Gatica-Perez, D. (2008). *What did you do today?* [Paper presentation]. October). Discovering daily routines from large-scale mobile data. In Proceedings of the 16th ACM international conference on Multimedia (pp. 849–852).
Feierabend, S., Plankenhorn, T., & Rathgeb, T. (2014). *KIM-Studie. 2014. Jugend, Information,(Multi) Media [JIM-Study 2014. Youth, Information, and (Multi) Media]*. Medienpädagogischer Forschungsverband Südwest.
Feierabend, S., Plankenhorn, T., & Rathgeb, T. (2016). *KIM-Studie. 2016. Kindheit, Internet, Medien. Basisuntersuchung zum Medienumgang. [KIM-Study 2016. Childhood, Internet, and Media. Base assessments of media usage]*. Medienpädagogischer Forschungsverband Südwest.
Fuller, D., Shareck, M., & Stanley, K. (2017). Ethical implications of location and accelerometer measurement in health research studies with mobile sensing devices. *Social Science & Medicine (1982)*, *191*, 84–88. https://doi.org/10.1016/j.socscimed.2017.08.043
Halverson, C. F., Havill, V. L., Deal, J., Baker, S. R., Victor, J. B., Pavlopoulos, V., Besevegis, E., & Wen, L. (2003). Personality structure as derived from parental ratings of free descriptions of children: The inventory of child individual differences. *Journal of Personality*, *71*(6), 995–1026. https://doi.org/10.1111/1467-6494.7106005
Harari, G. M., Gosling, S. D., Wang, R. U. I., & Campbell, A. T. (2015). Capturing situational information with smartphones and mobile sensing methods. *European Journal of Personality*, *29*(5), 509–511. https://doi.org/10.1002/per.2032
Harari, G. M., Lane, N. D., Wang, R., Crosier, B. S., Campbell, A. T., & Gosling, S. D. (2016). Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *11*(6), 838–854. https://doi.org/10.1177/1745691616650285

Harari, G. M., Müller, S. R., Aung, M. S., & Rentfrow, P. J. (2017). Smartphone sens- ing methods for studying behavior in everyday life. *Current Opinion in Behavioral Sciences*, *18*, 83–90. https://doi.org/10.1016/j.cobeha.2017.07.018

Harari, G. M., Müller, S. R., Gosling, S. D., Harari, G. M., Müller, S. R., & Gosling, S. D. (2018). Naturalistic assessment of situations using mobile sensing methods. In *The Oxford handbook of psychological situations*. Oxford University press. https://doi.org/10.1093/oxfordhb/9780190263348.013.14

Harari, G. M., Müller, S. R., Stachl, C., Wang, R., Wang, W., Bühner, M., Rentfrow, P. J., Campbell, A. T., & Gosling, S. D. (2020). Sensing sociability: Individual differences in young adults' conversation, calling, texting, and app use behaviors in daily life. *Journal of Personality and Social Psychology*, *119*(1), 204–228. https://doi.org/10.1037/pspp0000245

Harari, G. M., Vaid, S. S., Müller, S. R., Stachl, C., Marrero, Z., Schoedel, R., Bühner, M., & Gosling, S. D. (2020). Personality sensing for theory development and assessment in the digital age. *European Journal of Personality*, *34*(5), 649–669. https://doi.org/10.1002/per.2273

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., & Kaufman, J. (2015). Putting education in "educational" apps: Lessons from the science of learning. *Psychological Science in the Public Interest: A Journal of the American Psychological Society*, *16*(1), 3–34. https://doi.org/10.1177/1529100615569721

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology*, *75*(5), 581–595. https://doi.org/10.1037/0021-9010.75.5.581

Judge, S., Floyd, K., & Jeffs, T. (2015). Using mobile media devices and apps to promote young children's learning. In *Young children and families in the information age* (pp. 117–131). Springer. doi: https://doi.org/10.1007/978-94-017-9184-7_7

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(15), 5802–5805. https://doi.org/10.1073/pnas.1218772110

Lamb, M. E., Chuang, S. S., Wessels, H., Broberg, A. G., & Hwang, C. P. (2002). Emergence and constructvalidation of the big five factors in early childhood: a longitudinal analysis of their ontogeny in Sweden. *Child Development*, *73*(5), 1517–1524. https://doi.org/10.1111/1467-8624.00487

Lane, N. D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., & Campbell, A. T. (2010). A survey of mobile phone sensing. *IEEE Communications Magazine*, *48*(9), 140–150. https://doi.org/10.1109/MCOM.2010.5560598

Leung, L., & Chen, C. (2021). A review of media addiction research from 1991 to 2016. *Social Science Computer Review*, *39*(4), 648–665. https://doi.org/10.1177/0894439318791770

Livingstone, S., Mascheroni, G., Dreier, M., Chaudron, S., & Lagae, K. (2015). *How parents of young children manage digital devices at home: The role of income, education and parental style*. EU Kids Online, LSE.

Macias, E., Suarez, A., & Lloret, J. (2013). Mobile sensing systems. *Sensors (Basel, Switzerland)*, *13*(12), 17292–17321. https://doi.org/10.3390/s131217292

Maertens, B., De Smedt, B., Sasanguie, D., Elen, J., & Reynvoet, B. (2016). Enhancing arithmetic in pre-schoolers with comparison or number line estimation training: Does it matter? *Learning and Instruction*, *46*, 1–11. https://doi.org/10.1016/j.learninstruc.2016.08.004

Miller, G. (2012). The smartphone psychology manifesto. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *7*(3), 221–237. https://doi.org/10.1177/1745691612441215

Miluzzo, E., Lane, N. D., Fodor, K., Peterson, R., Lu, H., Musolesi, M., … Campbell, A. T. (2008 *Sensing meets mobile social networks* [Paper presentation]. The 6th ACM Conference, November). the design, implementation and evaluation of the cenceme application. In *Proceedings of on Embedded network sensor systems* (pp. 337–350). https://doi.org/10.1145/1460412.1460445

Mønsted, B., Mollgaard, A., & Mathiesen, J. (2018). Phone-based metric as a predictor for basic personality traits. *Journal of Research in Personality*, *74*, 16–22. https://doi.org/10.1016/j.jrp.2017.12.004

Montag, C., Błaszkiewicz, K., Lachmann, B., Andone, I., Sariyska, R., Trendafilov, B., Reuter, M., & Markowetz, A. (2014). Correlating personality and actual phone usage. *Journal of Individual Differences*, *35*(3), 158–165. https://doi.org/10.1027/1614-0001/a000139

Niklas, F., Annac, E., & Wirth, A. (2020). App-based learning for kindergarten children at home (Learning4Kids): study protocol for cohort 1 and the kindergarten assessments. *BMC Pediatrics*, *20*(1), 1–14. https://doi.org/10.1186/s12887-020-02432-y

Niklas, F., Cohrssen, C., & Tayler, C. (2016). Improving preschoolers' numerical abilities by enhancing the home numeracy environment. *Early Education and Development*, *27*(3), 372–383. https://doi.org/10.1080/10409289.2015.1076676

Niklas, F., & Schneider, W. (2013). Home literacy environment and the beginning of reading and spelling. *Contemporary Educational Psychology*, *38*(1), 40–50. https://doi.org/10.1016/j.cedpsych.2012.10.001

Niklas, F., & Schneider, W. (2017). Home learning environment and development of child competencies from kindergarten until the end of elementary school. *Contemporary Educational Psychology*, *49*, 263–274. https://doi.org/10.1016/j.cedpsych.2017.03.006

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research*, *78*(1), 33–84. https://doi.org/10.3102/0034654307313793

Papadakis, S., & Kalogiannakis, M. (2017). Mobile educational applications for children: What educators and parents need to know. *International Journal of Mobile Learning and Organisation*, *11*(3), 256–277. https://doi.org/10.1504/IJMLO.2017.085338

Paulhus, D. L. (2017). Socially desirable responding on self-reports. In V. Zeigler-Hill, T. K. Shackleford (Eds.), *Encyclopedia of Personality and Individual Differences* (pp. 1–5). https://doi.org/10.1007/978-3-319-28099-8_1349-1

Rasmussen, E. E., Strouse, G. A., Colwell, M. J., Russo Johnson, C., Holiday, S., Brady, K., Flores, I., Troseth, G., Wright, H. D., Densley, R. L., & Norman, M. S. (2019). Promoting preschoolers' emotional competence through prosocial TV and mobile app use. *Media Psychology*, *22*(1), 1–22. https://doi.org/10.1080/15213269.2018.1476890

Rat Für Sozial- und Wirtschaftsdaten. (2020). Datenerhebung mit neuer Informationstechnologie: Empfehlungen zu Datenqualität und -management, Forschungsethik und Datenschutz [Data collection with new information technology: Recommendations on data quality and management, research ethics, and data privacy]. *Ratswd Output*, *6*(6), 13–20. https://www.ratswd.de/dl/RatSWD_Output6.6_Datenerhebung-neueIT.pdf

Saeb, S., Lattie, E. G., Schueller, S. M., Kording, K. P., & Mohr, D. C. (2016). The relationship between mobile phone location sensor data and depressive symptom severity. *PeerJ*, *4*, e2537. https://doi.org/10.7717/peerj.2537

Schoedel, R., Au, Q., Völkel, S. T., Lehmann, F., Becker, D., Bühner, M., Bischl, B., Hussmann, H., & Stachl, C. (2018). Digital footprints of sensation seeking. *Zeitschrift Für Psychologie*, *226*(4), 232–245. https://doi.org/10.1027/2151-2604/a000342

Schoedel, R., Pargent, F., Au, Q., Völkel, S. T., Schuwerk, T., Bühner, M., & Stachl, C. (2020). To challenge the morning lark and the night owl: Using smartphone sensing data to investigate day–night behaviour patterns. *European Journal of Personality*, *34*(5), 733–752. https://doi.org/10.1002/per.2258

Schuwerk, T., Kaltefleiter, L. J., Au, J. Q., Hoesl, A., & Stachl, C. (2019). Enter the wild: Autistic traits and their relationship to mentalizing and social interaction in everyday life. *Journal of Autism and Developmental Disorders*, *49*(10), 4193–4208. https://doi.org/10.1007/s10803-019-04134-6

Shcherbina, A., Mattsson, C., Waggott, D., Salisbury, H., Christle, J., Hastie, T., Wheeler, M., & Ashley, E. (2017). Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. *Journal of Personalized Medicine*, *7*(2), 3. https://doi.org/10.3390/jpm7020003

Shuler, C. (2012). *iLearn II: An Addendum an analysis of the games category of Apple's App Store*. New York.

Stachl, C., Au, Q., Schoedel, R., Gosling, S. D., Harari, G. M., Buschek, D., Völkel, S. T., Schuwerk, T., Oldemeier, M., Ullmann, T., Hussmann, H., Bischl, B., & Bühner, M. (2020). Predicting personality from patterns of behavior collected with smartphones. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(30), 17680–17687. https://doi.org/10.1073/pnas.1920484117

Stachl, C., Hilbert, S., Au, J. –Q., Buschek, D., De Luca, A., Bischl, B., Hussmann, H., & Bühner, M. (2017). Personality traits predict smartphone usage. *European Journal of Personality*, *31*(6), 701–722. https://doi.org/10.1002/per.2113

Thomée, S. (2018). Mobile phone use and mental health. A review of the research that takes a psychological perspective on exposure. *International Journal of Environmental Research and Public Health*, *15*(12), 2692. https://doi.org/10.3390/ijerph15122692

Wahle, F., Kowatsch, T., Fleisch, E., Rufer, M., & Weidt, S. (2016). Mobile sensing and support for people with depression: a pilot trial in the wild. *JMIR mHealth and uHealth*, *4*(3), e111. doi: https://doi.org/10.2196/mhealth.5960

Walling, D. R. (2014). *Designing learning for tablet classrooms: Innovations in instruction*. Springer Science & Business Media.

Wang, J., Wang, Y., Zhang, D., & Helal, S. (2018). Energy saving techniques in mobile crowd sensing: Current state and future opportunities. *IEEE Communications Magazine*, *56*(5), 164–169. https://doi.org/10.1109/MCOM.2018.1700644

Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., … Campbell, A. T. (2014). *Studentlife* [Paper presentation]. Assessing mental health, academic performance and behavioral trends of college students using smartphones. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 3–14. https://doi.org/10.1145/2632048. 2632054

Wilson, S., Schalet, B. D., Hicks, B. M., & Zucker, R. A. (2013). Identifying early childhood personality dimensions using the California Child Q-Set and prospective associations with behavioral and psychosocial development. *Journal of Research in Personality*, *47*(4), 339–350. https://doi.org/10.1016/j.jrp.2013.02.010

Wirth, A., Ehmig, S. C., Drescher, N., Guffler, S., & Niklas, F. (2020). Facets of the early home literacy environment and children's linguistic and socioemotional competencies. *Early Education and Development*, *31*(6), 892–909. https://doi.org/10.1080/10409289.2019.1706826

Ziegler, M., & Buehner, M. (2009). Modeling socially desirable responding and its effects. *Educational and Psychological Measurement*, *69*(4), 548–565. https://doi.org/10.1177/0013164408324469