A Review of Current Ambulatory Assessment Studies in Adolescent Samples and Practical Recommendations

Eeske van Roekel (iD), Loes Keijsers, and
Joanne M. Chung (iD)
Tilburg University

The use of ambulatory assessment (AA) and related methods (experience sampling, ecological momentary assessment) has greatly increased within the field of adolescent psychology. In this guide, we describe important practices for conducting AA studies in adolescent samples. To better understand how researchers have been implementing AA study designs, we present a review of 23 AA studies that were conducted in adolescent samples from 2017. Results suggest that there is heterogeneity in how AA studies in youth are conducted and reported. Based on these insights, we provide recommendations with regard to participant recruitment, sampling scheme, item selection, power analysis, and software choice. Further, we provide a checklist for reporting on AA studies in adolescent samples that can be used as a guideline for future studies.

Ambulatory assessment (AA) is a research methodology that uses a variety of data sources to better understand people's thoughts, feelings, and behaviors in their natural environment. AA is typically implemented through the repeated administration of brief questionnaires and the monitoring of activity over a period of time, for instance, through smart phone apps or through wearables. AA allows researchers to study people outside of the laboratory, making this methodology more ecologically valid than other, traditional methodologies.

One of the earliest AA studies among adolescents was Larson and Csikszentmihalyi's (1983) work examining the socio-emotional lives of teenagers. Youth were given a packet of questionnaires and an electronic pager, from which they received signals several times a day. When beeped, adolescents completed a brief survey with questions about mood, peers and other relationship partners, and their environment. Through studies like these, AA has provided rich insights into the psychology

of adolescents at a level that is unprecedented (Larson, 1983; Larson & Csikszentmihalyi, 1983; Larson, Csikszentmihalyi, & Graef, 1980). However, obtaining such data has required much effort in the past; for example, researchers had to rely on the use of electronic pagers and paper-and-pencil questionnaires, and would often require a number of personnel to successfully carry out the study design.

Yet, many of the practical hurdles for studying adolescents in a more naturalistic, ecologically valid way have since lessened. Smartphones and wearables have become an integral part of adolescent life. For instance, adolescents report using technology an average of 9.25 hr each day (Katz, Felix, & Gubernick, 2014). Additionally, a multitude of mobile applications and software packages now exist for the sole purpose of helping researchers conduct AA studies more efficiently. For example, on a smartphone, features such as push notifications can alert participants that an assessment is ready, and health and social activities can be assessed through global positioning system (GPS) scans, accelerometer activity, and text message logs. Moreover, these applications often allow for the careful tracking of the study's progress in real time from a researcher's own computer. When preparing the data for analysis, content coding of open-ended questions may be done in part automatically, for instance, by translating the structured

coding scheme to a study-specific code using available software, such as R (R Core Team, 2017). Not surprisingly, the increased availability of AA tools provides exciting opportunities for researchers who are interested in conducting psychological research with adolescents "in the wild" and has resulted in greater applications of AA study designs in psychological research (Hamaker & Wichers, 2017), including in the field of adolescent psychology (see Figure 1 for study numbers based on a structured search in PubMed for AA studies in adolescents).

Given the increase in popularity of AA designs within the field of adolescent psychology, we think the time is right to provide concrete guidelines for conducting AA studies specifically with adolescent samples. In this guide, we share our knowledge of practices for conducting AA studies in adolescent populations. We first report the results of a structured review of AA studies that use adolescent samples published in 2017 and describe current standards within the field of adolescent psychology. Building on the valuable content of earlier guidelines and reviews in both youth (Heron, Everhart, McHale, & Smyth, 2017; Wen, Schneider, Stone, & Spruijt-Metz, 2017) and adult samples (Christensen, Barrett, Bliss-Moreau, Lebo, & Kaschub, 2003; Scollon, Prieto, & Diener, 2009), we offer insights that focus on three topics: study design, technical issues, and practical issues. Within these three topics, we discuss current standards based on the literature review we conducted and provide suggestions tailored to AA research with adolescent samples based on our own experiences with collecting such data (Keijsers, Hillegers, & Hiemstra, 2015; van Roekel et al., 2013).

A STRUCTURED REVIEW OF CURRENT PRACTICES OF AA IN ADOLESCENT PSYCHOLOGY

To gain some insight into the current practices of AA studies in adolescent samples, we conducted a structured review of all AA studies in adolescent samples published in 2017. Although some excellent reviews on AA studies in youth up to 2016 have been published (Heron et al., 2017; Wen et al., 2017), the pace of technological possibilities and knowledge of good research practices has expanded. Therefore, we provide a summary of the most recent practices and methods of reporting. These studies were not reported in previous reviews on youth (Heron et al., 2017; Wen et al., 2017).

Method

We identified studies that used AA in adolescent samples by conducting a search through PubMed. We used the search terms "ecological momentary," "experience sampling," "ambulatory assessment*," "momentary assessment*," "EMA," "ESM" in combination with "adolescen*" and "youth." To be included in the review, the study must: (1) include empirical data, (2) use >1 ambulatory assessment per day (i.e., diary studies were excluded), (3) assess participants who were between 10 and 18 years old, and (4) have a publication date of 2017 (the last search conducted was on November 22, 2017). We used a coding scheme to assess the relevant information from each study (see Table S1 in the online Supporting Information), using

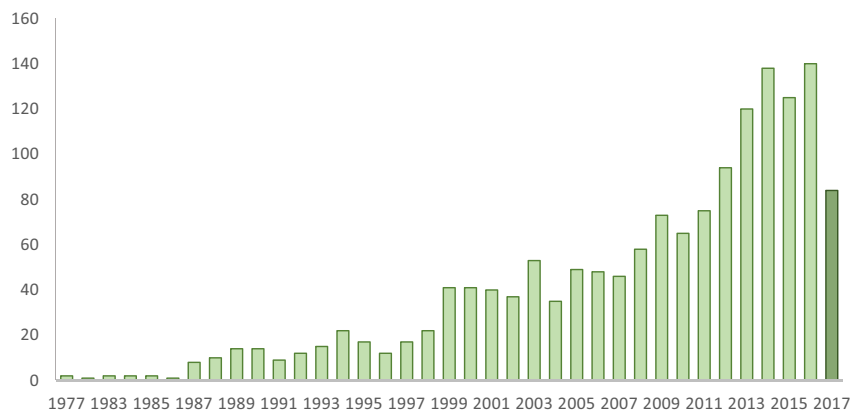FIGURE 1   Number of studies using AA in adolescent samples until November 22, 2017. Studies of 2017 are summarized in Table S1 in the online Supporting Information. *Note.* We used the search terms "ecological momentary," "experience sampling," "ambulatory assessment*," "momentary assessment*," "EMA," "ESM" in combination with "adolescen*" and "youth." [Color figure can be viewed at wileyonlinelibrary.com]

categories from a previous review as a starting point (Heron et al., 2017). We added the following categories: study purpose, number of items administered at each assessment, questionnaire duration, mobile sensor use, sampling during school hours, time allotted for questionnaire completion, and incentives. All publications resulting from the search were checked with regard to inclusion criteria by two independent coders.

Results

Our search resulted in 86 publications that could potentially be included in our review. Of these, 12 were not empirical, 27 did not include self-reported momentary assessments, and 15 included samples of participants that were younger than 10 or older than 18 years. Additionally, one study could not be accessed online and was therefore excluded. Our final selection consisted of 31 publications. Some of these publications used the same data set and therefore were combined in our review, resulting in 23 unique studies.

Of these 23 unique studies, seven were conducted in clinical samples (i.e., adolescents in treatment). As shown in Table S1 in the online Supporting Information, these recent AA studies on adolescents covered various topics, including nonsuicidal self-injury, stress, alcohol use, sleep, passionate experiences, marijuana use, and emotion differentiation. Furthermore, we coded features of the study design, including type of sampling, number of measurements, compliance, and implementation procedures. We will incorporate results from the review in detail in each section below.

In general, we found that many studies did not report details with regard to the study design and data collection, such as power calculations, number of items, questionnaire duration, and the extent to which any problems were encountered. This is problematic, as it makes it difficult to replicate findings and overcome similar methodological issues in future research. We now turn to guidelines for setting up an AA study based on our structured review, and our own experiences.

STUDY DESIGN

Should I Conduct an AA Study?

Ambulatory assessment is an exciting method for studying adolescents in a naturalistic manner as they go about their daily lives. Participants are often asked to report in the moment or to reflect on their thoughts, feelings, and behaviors over a short period of time, reducing recall bias. The repeated measures, longitudinal design also allows for more reliable estimates of the psychological process at hand (Myin-Germeys et al., 2009). Additionally, the collection of behavioral data that is often employed in AA studies contributes an alternative data source that can complement self-reports. Although the aforementioned reasons can result in a study's increased ecological validity, AA can be considered intrusive and intensive for participants (Hufford, 2007). Therefore, before setting up an AA study, the first question for the interested researcher is, do I really need an AA design?

Box 1. Collaborating with schools (tips and tricks)

- In order to form a strong research alliance, it is important to actively inform school boards, teachers, parents, and the adolescents themselves about the goals and relevance of the study and the additional value for the school.
- One way of increasing the benefits for the schools is to examine questions that are of interest to school administrators (e.g., motivations to do well in courses, reasons for frequent absenteeism, factors that impact adolescents' well-being), and present the study results to the school, especially in a way that the school finds most useful (e.g., via a policy report, a presentation for teachers).
- If schools have a policy that forbids the use of smartphones on campus, researchers could: (1) create identification cards for adolescents who are participating in the study, to show to teachers that they have permission to use their smartphone during classes. In doing so, other adolescents will not be able to use the study as an excuse to use their smartphones in class; (2) disable phone functionalities outside of those required by the study design, if possible.
- To foster a partnership with schools, it is essential to include teachers, parents, and adolescents in activities related to the research project (from design and recruitment to report). This can be accomplished by providing short presentations at parent and teacher meetings. Adolescents may be recruited through personal presentations in classes by highlighting the advantages of participation, and through attractive advertisements. Linking the study to student projects can also increase participation.

Or, does the burden for the participant outweigh the benefits?

Recruitment of Participants

If the research question requires an AA design, the first step in designing an AA study is to decide on the characteristics of the sample that the researcher will recruit, but also on the feasibility and desirability of conducting an AA study in the sample of interest. For example, does the research question require a clinical or nonclinical sample? What age group? In which cultural context?

In our experience, when the objective is to examine normative development, one efficient way to set up a study is to connect with adolescents via the school administration. Not all schools will be equally willing to participate, or will allow their students to participate. At the same time, we have experienced that it is feasible to create enthusiasm for the study by reaching out to school administrators, and ensuring that the school administrators are treated as active and equal partners in the research project, such that both researchers and schools benefit from the study outcomes. We share some of our favorite approaches for collaborating with schools in Box 1 (which, at least in the Netherlands, provided good compliance rates; van Roekel et al., 2013)

Ambulatory assessment studies are not limited to adolescents with a typical developmental trajectory. In fact, in our review of recent AA studies, of the 23 unique studies, seven were conducted in clinical samples (e.g., youth in treatment, youth diagnosed with physical or psychiatric disorders), including adolescents with Duchenne muscular dystrophy (Bray, Bundy, Ryan, & North, 2017), bipolar disorder (Andrewes, Hulbert, Cotton, Betts, & Chanen, 2017a), and anorexia nervosa (Kolar et al., 2017). In such clinical samples, the process of recruitment is slightly different. In these seven studies (Andrewes et al., 2017a; Bray et al., 2017; Kolar et al., 2017; Kranzler et al., 2017; Rauschenberg et al., 2017; Ross et al., 2018; Wallace et al., 2017), all clinical samples were recruited through clinical institutions such as mental health care institutions, health clinics, and hospitals, sometimes with additional measures such as flyers or advertisements (Kranzler et al., 2017; Wallace et al., 2017). As with schools, a strong alliance between the researcher and the institution is essential for success.

Concerns have been raised in the literature with regard to the impact of AA studies on vulnerable youth in clinical samples. Thus, collecting AA data in clinical samples requires careful consideration of different aspects, especially with regard to ethical concerns. First, it is often assumed that reporting multiple times per day about symptoms could worsen the problems. Yet, multiple studies suggest that frequent reporting on symptoms does not negatively affect depressive symptoms (Broderick & Vikingstad, 2008; Kramer et al., 2014), anhedonic symptoms (van Roekel et al., 2017), and pain levels (Cruise, Broderick, Porter, Kaell, & Stone, 1996). Second, filling out multiple questionnaires a day for a period of time can be a burden on participants, which may be more problematic in clinical samples. However, research has shown that it is feasible to collect AA data in clinical samples, and often compliance is higher in clinical samples than in normative samples (see, e.g., Ebner-Priemer & Trull, 2009).

In our experience, in order to create a strong research alliance with clinical institutions, it is helpful to highlight the potential advantages that AA may have for adolescents with clinical symptoms. Participating in AA studies may have benefits for clinical samples, as reporting on symptoms, moods, and activities multiple times per day may provide self-insight into one's symptoms and what elicits these symptoms (Kramer et al., 2014). One possibility is that AA can be used as the basis for low-cost interventions for clients on a waiting list for treatment. For example, research in late adolescents has shown that it is both feasible and effective to use momentary assessments as a tool to provide personalized feedback (van Roekel et al., 2017). Additionally, in adult samples, first steps have already been taken toward integrating AA in clinical practice, in which AA data are discussed as part of the treatment (Kroeze et al., 2017). These applications of AA in clinical practice are highly relevant for adolescent samples as well and may help to collaborate with clinical practice in a way that is fruitful for the institution, the participant, and the researcher.

Sampling Scheme

As with any study in psychology, in AA studies the study design should be closely aligned with the research questions that one aims to answer. Aspects that may be more specific to AA study design are the time window in which sampling occurs, the type of sampling (i.e., event-based—filling out an assessment after an event has occurred, or time-based—at specific times), the intervals

between assessments, and the number of days and assessments. We now turn to a discussion of the considerations regarding the sampling scheme.

Time window.   One question to address is the time window in which the sampling occurs. Because adolescents spend a great deal of their time in schools, the researcher must decide whether or not to sample during school hours. Of the recent studies described in Table S1 in the online Supporting Information, about half sampled during school hours (54.2%). One of the main advantages of sampling during school hours is that it will provide a more comprehensive picture of adolescents' daily lives. However, schools and teachers have to agree to this, which may be difficult as many schools enforce anti-smartphone policies. We have included some tips and tricks for sampling during school hours in Box 1. An alternative would be to only sample during regularly scheduled breaks (e.g., during the mid-morning break, lunch, and afternoon break). In addition, it is important to take care when deciding on the first and last assessments of each day for each participant. This decision largely depends on the research question and whether the variables of interest occur during the early morning or late evening. In our experience, tailoring the first and last assessments to correspond to the adolescents' sleep and school schedule is an effective way to increase participant compliance. When individual schedules are not possible due to software constraints, finding a time window that is feasible for all adolescents would be the second-best option.

Types of sampling.   There are three ways of collecting experience sampling data, each with its own unique strengths: interval-contingent sampling, event-contingent sampling, and signal-contingent sampling. Interval-contingent sampling refers to sampling that occurs when participants provide self-reports after a predetermined amount of time (e.g., the participant reports on her mood at the end of each hour). Event-contingent sampling refers to sampling that occurs when participants provide self-reports following a specific event (e.g., a participant indicates how satisfied he is with his relationships immediately following a social interaction). Signal-contingent sampling refers to sampling that occurs when participants provide self-ratings following a notification (e.g., a participant provides self-reports after receiving a push notification on their phone) that is either fixed (e.g., at 9 a.m., 12 p.m., 3 p.m., 6 p.m.) or random (e.g., 5 random time points throughout the day). A specific strength of the interval-contingent approach is that there are equal intervals in the data, which allows the use of discrete time methods for modeling the data, something that may not be always possible with event-contingent or signal-contingent sampling (but see de Haan-Rietdijk, Voelkle, Keijsers, & Hamaker, 2017). Yet, this advantage only holds when discrete time methods accurately deal with the interval between the last assessment of one day to the first assessment the next day. This has become possible using new analytic techniques, such as Dynamic Structural Equation Modeling (DSEM). In fact, the analytical advantage of using equal intervals in one's study design is disappearing as several statistical analyses packages can now handle unequal time intervals (e.g., PROC MIXED procedure in SAS, and the tinterval option in the DSEM package in Mplus). Event-contingent sampling is most often used when researchers are interested in specific behaviors that may be rare or irregular, such as nonsuicidal self-injury or substance use. Event-contingent sampling can also be used combined with mobile sensing technology. For example, an assessment can be triggered when participants enter a specific location (GPS), or when participants are highly active or inactive (actigraphy). A specific strength of signal-contingent sampling is that it allows to provide a random subset of behaviors and moods as they occur throughout the day. The main advantages of random sampling are that it (1) decreases the possibility that adolescents change their daily life behaviors because they are not able to predict when the next signal will occur, and (2) decreases the possibility that adolescents will be in the same context at every occasion. For youth in schools, random sampling may be difficult as the signal might occur at inconvenient times (e.g., during tests, presentation). Further, if sampling during lessons is not possible, fixed time points during break times or user-initiated assessments may be useful; this can only be done when it is not a problem that assessments occur in the same contexts (e.g., during breaks).

As can be seen in Table S1, most recent studies use signal-contingent sampling with random time points (58%), followed by fixed time points (20.8%), a combination of random time points with event-contingent sampling (12.5%), and interval-contingent sampling (4.2%). Which method is preferred will depend, among others, on the research question, the prevalence of the behaviors, but also the desired analytical strategy once the data are collected. For instance, it is important to align the

study design to the "speed of the process" under examination. That is, to observe emotional episodes we may need to measure every minute, to assess mood, we may rely on hourly or daily measures, and to assess relatively stable temperament traits we may need yearly intervals (Lewis, 2000).

Items

Number of days/assessments/items. Making a decision regarding the number of days, number of assessments per day, and number of items per assessment requires careful consideration for the data needed to answer the research question, while minimizing the expected burden on participants. In recent studies (see Table S1), the total number of assessments ranged between 12 and 147 ($M = 49.05$, $SD = 29.95$), with on average 5.65 assessments per day ($SD = 3.01$, range between 2 and 15 assessments) and 12.30 days ($SD = 10.78$, range between 2 and 42 days). Unfortunately, most studies do not report how many items were administered in total and how long it took adolescents to fill out the questionnaire. The only study that reported both showed that filling out five items took between 10 and 60 seconds (D'Amico et al., 2017). We were able to calculate these numbers for our own data (see Table 1 for study details). We have shared our data and syntax for all reported analyses on OSF, which can be found at https://osf.io/u9cqp/. Filling out a 37-item questionnaire on a smartphone, including five open-ended questions, took on average 6 min ($SD = 2.8$) (van Roekel et al., 2013), whereas filling out 23 items, including one open-ended question, took on average 2 min ($SD = 6.2$) (Keijsers et al., 2015).[1] We also checked whether survey completion time decreased when adolescents became more familiar with the questions. We therefore performed multilevel analyses in Mplus 8, to examine whether the number of the assessment and survey completion time were associated. We found small significant associations in both studies ($B = -.01$, $p < .001$ for study 1; $B = -.40$, $p < .001$ for study 2). Although these effects are small, this indicates that adolescents became slightly quicker in filling out the assessments when they became more familiar with the questions.

One possibility for reducing the burden posed on participants by having a large number of items

is to use a planned missingness design (for an elaborate discussion of the pros and cons see: Silvia, Kwapil, Walsh, & Myin-Germeys, 2014), by allowing researchers to exclude items at each assessment. Different designs are possible, such as the anchor test design (e.g., when the specific item "sad" is always shown, and other items like "blue" or "unhappy" are sometimes shown) and the matrix design (e.g., each item is combined with every other item a third of the time, and participants see only two of the three items at each assessment). Combined with a multilevel latent variable approach, in which several items are used as indicators of one construct (e.g., ratings of "sad," "blue," and "unhappy" are used as manifest indicators of an overarching latent construct of sadness) this may allow one possibility to ask for more constructs, or more items per construct without increasing the burden for participants.

Characteristics of Items

It is not evident that scales constructed for adult populations can readily be used in adolescent samples; therefore, new items or items that are derived from measures used with adult samples should be carefully piloted or discussed in focus groups. Below we discuss considerations with regard to choosing item formats and answer scales.

Type of items (open vs. closed). Items can have both open-ended and closed formats. For example, one can choose to ask participants "What are you doing right now?" and allow them to answer freely, or provide a list of categories that pertain to different activities. The primary advantages of closed questions are that it takes less time and effort for participants and that the responses can be used directly in analyses, without requiring qualitative coding. Open-ended questions may provide more variation and could offer insights regarding the research question that the researchers did not consider beforehand. Still, there is another, unexpected, advantage of open-ended questions in adolescent samples that we have encountered in our research (van Roekel et al., 2013). We have found that answers to the open-ended questions could be used as a check for careless responding. For example, assessments in which the current activity was described with "poop" and "who[ever] reads this is dumb" were judged as careless responders and such data were removed. In general, however, our advice would be to avoid open-ended questions unless (1) you want to know what people think or feel without forcing

---

[1] We excluded assessments that took more than 100 min to fill out, as these were likely assessments that adolescents left open and completed at a later time point during the day.

TABLE 1
Sample Characteristics of Example Data

	Sample	Design	Items	Monitoring	Compliance
Study 1: Swinging Moods; (van Roekel et al., 2013)	303 adolescents, M_{age} = 14.19. 59% female	6 days, 9 random assessments per day	37 items	Real-time, contacted after more than 2 missings	68.7% (37.10 out of 54)
Study 2: Grumpy or Depressed (Keijsers et al., 2015)	241 adolescents, M_{age} = 13.81, 62% female	7 days, 8 random assessments per day	23 items	When data were uploaded, approximately once per day	47.6% (25.71 out of 54)

categories on them; or (2) your research question is largely exploratory and you do not yet know what the potential categories might be.

Answer scales (Likert, VAS, categorical). When using closed questions, researchers can choose between different types of answer scales: categorical answers with one forced choice, categorical answer with multiple choices, continuous answers with Likert scales, or continuous answers with visual analogue scales (VAS). Typically, researchers have used Likert scales, with, for example, seven answer options. Given the technological possibilities however, more researchers have started using VAS as well. The main advantage of using VAS is that it is a more sensitive measure, as participants are able to answer on a scale ranging from 0 to 100 (McCormack, Horne, & Sheather, 1988). Moreover, it may be that adolescents prefer to answer VAS scales compared to Likert scale question formats (Tucker-Seeley, 2008). At the same time, not all factor structures and mean levels may replicate when Likert scales or VAS scales are used (Hasson & Arnetz, 2005; Tucker-Seeley, 2008), and a careful examination is needed in order to establish whether or not this is the case for specific instruments of interest (Byrom et al., 2017). When piloting a study, it is important to test whether the technical aspects of the device (e.g., Apple vs. Android, the size of the device) can impact the participants' responses, especially when differences between individuals are of interest.

Moreover, in adolescent samples, we have found that it is crucial to provide instructions on how to complete the items. For example, during a pilot for one of our studies (van Roekel et al., 2013), we realized that the majority of adolescents always reported the lowest possible score for negative emotions (i.e., 1 = *not at all*) and the highest possible score for positive emotions (i.e., 7 = *a lot*). Encouraging adolescents to make full use of the

range of the scale can help to ameliorate these problems related to limited variance.

Consequences of Design Choices

Choices made with regard to the different study design features described above can impact the quality of the data. Below we discuss some important consequences of design choices on compliance rates, analytic choices, and power.

Compliance. Compliance is one important quality marker for AA studies. When it comes to designing a study, the number of assessments may impact the burden on participants (Hufford, 2007; Wen et al., 2017). In our review of recent studies (see Table S1), we checked whether study-level or individual-level compliance was reported and how it was reported. Most often, studies report compliance at the study level, as either a percentage of the total number of assessments that was completed, or the average number of assessments that was filled out per individual. Based on these numbers, results showed that the compliance rate varied substantially across studies between 51.56% and 92.00% (M = 74.00, SD = 11.50). What is generally missing however, is more accurate insights into individual-level compliance, that is, for instance by adding measures or indicators for spread around the study-level compliance (e.g., SD, histogram, patterns of individual-level compliance rates). In general terms, this average study-level compliance rate is similar to what has been found in adult samples (Hufford, Shiffman, Paty, & Stone, 2001), yet to our knowledge there are no (recent) reviews or meta-analyses available on compliance in adults. In these recent studies, no clear patterns appeared between the total number of assessments in a study and compliance rates (see Figure 2; r = −.03, p = .90, N = 18). In order to increase knowledge on compliance for future studies, we
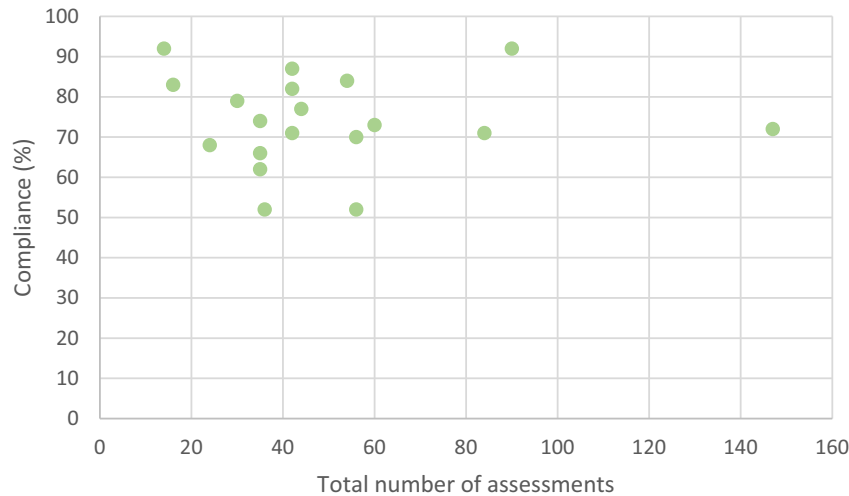
FIGURE 2   Relation between total number of assessments and compliance rate (N = 19). Note. Out of the total of N = 23 unique data sets, N = 4 did not report compliance rates. Those studies could not be included in the Figure. [Color figure can be viewed at wileyon linelibrary.com]

have included recommendations on what to report in future AA studies (see Box 5).

To give some insight into (1) which situations adolescents were most inconvenienced by in the assessment and (2) individual characteristics that may be associated with compliance, we conducted additional analyses on our own data on early adolescents (Keijsers et al., 2015; van Roekel et al., 2013).

For the first question, we measured the extent to which individuals were inconvenienced by each assessment with the item "I was inconvenienced by this beep," rated on a 7-point scale (1 = *not at all* to 7 = *very much*). On average, adolescents were moderately inconvenienced by the assessment in Study 1 (M = 4.05, SD = 1.16), and only slightly inconvenienced by the assessment in Study 2 (M = 2.18, SD = 1.86). This difference is rather large and may be due to a less intrusive notification beep in Study 2. Further, compliance rates were generally lower in Study 2, which might indicate that adolescents may have missed the notification at more inconvenient moments. To calculate differences in level of inconvenience between different contexts, we conducted multilevel analyses in Mplus 8, using dummy variables to examine the effects of different contexts and locations. We added the dummy variables as random effects, which allows for individual variation around the fixed effects. Detailed results can be found in Tables A1 and A2 (see the appendix). We found no differences in the level of inconvenience between assessments collected on weekdays and on weekends in the first sample, but we did find that adolescents were more

inconvenienced by the assessment on weekends in the second sample. In both samples, adolescents were less inconvenienced when with company compared to when alone. With regard to type of company, in the first sample significant differences were found between all types of company. Adolescents were most inconvenienced when with acquaintances (e.g., teammates, colleagues; M = 4.78), followed by friends (M = 4.34), family (M = 3.91), and lastly classmates (M = 3.71). Although this pattern was similar in Sample 2, fewer significant differences between social contexts were found. Adolescents were more inconvenienced when with friends (M = 2.26) compared to family (M = 2.16) and were more inconvenienced when with others (M = 2.44) compared to classmates (M = 2.04). With regard to location, in the first sample, adolescents were most inconvenienced when they were in public places (M = 4.49), followed by at home (M = 4.04) and school (M = 3.82). In the second sample, no significant differences were found between locations. The finding that adolescents are least inconvenienced when with classmates (both samples) or at school (Study 1) is interesting, as it indicates that sampling during school hours is at least feasible for adolescents themselves, as they are least inconvenienced by the assessment at school. We have not addressed, however, to which extent teachers or other companions were bothered by the adolescents' phone use.

For the second question, we examined associations between demographic characteristics and the number of completed assessments by conducting t

tests (gender), analyses of variance (ANOVAs) (educational level), and correlations (age). Compliance was higher for girls than for boys (Study 1: t (301) = −2.24, p = .03; Study 2: $t(241)$ = −3.16, p = .002), and for higher educational levels, compared to lower educational levels in Study 1 (Study 1: $F(2, 297)$ = 8.26, $p < .001$; Study 2: $t(241)$ = −0.30, p = .76). Age was not associated with compliance (Study 1: r = −.06, p = .27; Study 2: r = −.12, p = .06). Thus, these findings suggest that the sample characteristics such as gender and education level may partially affect the compliance rate that is feasible within a study.

Analytic Choices and Power

One of the challenges of conducting AA studies is appropriately taking the complex data structure into account when determining one's analytical strategy *after* the data have been collected (see also Keijsers & van Roekel, 2018). It is more optimal to consider one's analytical choices *while* designing the study. There are two important considerations in choosing the appropriate analytical strategy in AA studies. First, we need to account for the nested nature of the data (i.e., observations clustered within individuals), for instance, to avoid ecological fallacies in one's interpretation. We can do this by using multilevel modeling, as most studies in our review did. Second, one other aspect of the data is that time plays an important role. Measurements taken on Monday, for instance, are typically more closely associated with measurements taken on Tuesday, than to the measurements on the subsequent Friday. To date, only a handful of studies have been able to take the time-dynamic structure of the data into account, by, for example, using time series analyses in which univariate or multivariate lagged associations are also included. Fortunately, due to recent methodological developments, it has become possible to examine such lagged, dynamic associations in multiple software packages (e.g., DSEM in Mplus; Asparouhov, Hamaker, & Muthén, 2018) in a relatively user-friendly way. In models for lagged associations, the time that elapsed between assessments plays a role as well. Some techniques assume equal distances (e.g., Discrete Time Vector Autoregressive modeling in R), whereas DSEM in Mplus or Continuous Time Structural Equation Modeling in the CT-SEM package in R, for instance, are more flexible in accurately dealing with unequal intervals between assessments (de Haan-Rietdijk et al., 2017). Analyticaldevelopments for AA data are thus rapidly evolving, making it feasible to better match the structure of the data with the analytical design and to obtain a valid answer to theoretical questions from a complex data structure.

However, even though analytical techniques are evolving rapidly—and are increasingly able to deal with the complex nature of intensive longitudinal data, they cannot compensate for lack of power. As with any research design, having enough statistical power to answer the research question is a fundamental issue in determining the design of an AA study. Unique to AA studies, power may come from the number of subjects in the study (N) or the number of repeated assessments (T). In our review of recent studies, sample sizes ranged between 31 and 996 (M = 166, SD = 215, Median = 99) and the average number of assessments was 49 (SD = 30, Median = 42). Importantly, although we explicitly looked for power analyses in these manuscripts, none of the studies reported power calculations. Some studies did report small sample size as a limitation, but none of these claims were substantiated by power calculations specifically reported in the manuscript.

It is challenging to define general rules of thumb on power, and it will always depend on the exact nature of the hypothesis and the desired analytic design. Yet, some studies do provide some rules of thumb or insights. For instance, when the purpose is to estimate time series on n = 1, it is a recommendation to have 50 or even 100 time points (Chatfield, 2004; Voelkle, Oud, von Oertzen, & Lindenberger, 2012).

When the analytic model has a multilevel structure, power to estimate parameters can be "borrowed" for both levels, the number of assessments at level 1 as well as the number of participants at level 2, but the choice for more T or more N is not arbitrary. For instance, a recent study on Dynamic Structural Equation Modeling (Schultzberg & Muthén, 2017) examined the power of univariate two-level autoregressive models of order one (i.e., an AR(1) model). In general terms, this study demonstrated that N = 200 gave very good performance, even for T = 10 in some models for estimating autoregressive effects. Generally, a larger N thus seems able to compensate better for small T, than the reverse (i.e., large T being able to compensate for small N) in a multilevel autoregressive model (Schultzberg & Muthén, 2017). Moreover, with more complicated models, the requirements went up and a larger T was needed to obtain the same power.

Recently, Monte Carlo simulations were conducted on a two-level confirmatory factor model on data structures with planned missingness. These have shown that models converge well and lead to minimally biased parameter estimates when the N includes at least 100 people and 30 assessments (Silvia et al., 2014). At the same time, level 1 standard errors increased compared to a design without missing data, and to the best of our knowledge it is yet to be tested to which extent missing data designs also perform well when heterogeneity in the level 1 estimates is examined (e.g., different factor models for different individuals).

Even though these studies provide some guidance in setting up a study, whether or not these estimated sample sizes apply to other studies in adolescent psychology is an empirical question that can only be answered per individual study. We highly recommend researchers setting up new EMA studies to conduct a priori power analyses, for example by using Monte Carlo simulations in software programs like Mplus (Muthén & Muthén, 2002, 1998–2017) and R (R Core Team, 2017). Concrete guidelines for conducting these simulations for different research questions can be found in Bolger & Laurenceau (2013).

TECHNICAL ISSUES

Type of Device

Recent statistics concerning smartphone use show that smartphones are now so integrated in adolescent lives that it seems to be the most logical device to use in AA studies. For example, in Western countries, around 95% of teens own a smartphone (e.g., Netherlands, 96% of 13–18 year olds [Kennisnet, 2015]; Australia, 94% of 14–17 year olds [Roy Morgan Research, 2016]; UK, 96% of 16–24 year olds [Statista, 2017]; USA, 89% of 12–17 year olds [eMarketer, 2016]). For most countries, iPhones are more popular among teens than Android phones (e.g., 82% of all US teens own an iPhone [Jaffray, 2018] versus 58% of all Australian teens [Roy Morgan Research, 2016]).

In recent studies, 52.2% of adolescents were provided with a phone, whereas 16.7% used their own phone. Further, 25% of the studies used another device (e.g., PDAs) and one study used nondigital methods (i.e., paper-and-pencil). To our knowledge, there are no studies examining differences in compliance rates for these different devices (e.g., nondigital vs. digital or differences between different types of devices). Although using one's own phone has clear benefits, there are some challenges to consider when adolescents use their own phone. For example, some apps only run on specific platforms (Android vs. IOS), and older phones may not have the necessary specifications needed to run AA software. When comparing response styles, one study, in which a careful comparison was made between different device types and device sizes with a paper-and-pencil method among adults, minimal differences were found in VAS scale responses (Byrom et al., 2017), suggesting that in terms of how people fill out items, similar results can be obtained whether paper-or-pencil or digital devices are used. Yet, it is important to note that when using paper-and-pencil methods, it is not possible to check for backward or forward filling, which is an important disadvantage of this method. Given the advantages of using digital devices and the high levels of smartphone use in most countries, encouraging adolescents to use their own smartphones seems to be feasible for future studies.

Recommendations for Software

As the number of AA studies grows, new applications are continuously being developed, which makes it difficult to provide an up-to-date overview of potential software. This is further complicated by our finding that most studies do not explicitly report which software is used. Therefore, we provide an overview of important characteristics and requirements that we think should be considered when deciding on which software package to use in Box 2.

In addition to commercially available apps (e.g., Movisens, Illumivu, EthicaData), there are also a number of open-source alternatives such as ExperienceSampler (Thai & Page-Gould, 2017) or formr.org (Arslan, 2013). In our experience, in order to set the first steps in AA research, it may be most convenient to rely on an existing package, as developing new software or applications is highly time- and money-consuming. At the same time, it does require that the researcher thoroughly examines safety and security issues related to collaboration with an external party. Advice from legal and ethical experts may be needed on this issue.

Mobile Sensing Possibilities

As Table S1 shows, there are not yet many studies that combine AA questionnaires with mobile

```
┌─────────────────────────────────────────┐
│        Box 2. Requirements for software  │
│                                          │
│ • Does the application work on different │
│   types of smartphones? Given that both  │
│   Apple and Android platforms are used   │
│   by adolescents (with a slightly higher │
│   prevalence of Apple; see numbers       │
│   reported earlier), the app should      │
│   preferably work on both platforms.     │
│ • Is it possible to set notifications    │
│   and reminders?                         │
│ • Is real-time monitoring of incoming    │
│   data possible?                         │
│ • Are missing assessments registered in  │
│   the resulting datafile?                │
│ • Are assessments time-stamped?          │
│ • Are items time-stamped (to check       │
│   duration of filling out one            │
│   assessment)?                           │
│ • Is identifying information collected   │
│   from participants (e.g., IP addresses) │
│ • Who own the data? Are they safely      │
│   stored?                                │
└─────────────────────────────────────────┘
```

sensing. Only one study reported using GPS measures, and three studies used separate actigraphy devices in addition to the phone provided for the study. As there are excellent reviews available on mobile sensing possibilities (Harari et al., 2016), we refer to those for more information. Still, a relevant consideration for adolescents may be whether they can fully grasp what it means to consent to mobile sensing data collection. Do they understand what it means to provide their sensor data? This is an issue that might be relevant to check in focus groups.

PRACTICAL ISSUES DURING DATA COLLECTION

Instructions for Participants

We have found that the key to obtaining reliable data is to instruct participants on how to participate in an AA study. In our review of studies, we noticed that most recent studies do not report how they instructed participants, and what the specific instructions were. In order to make these instructions more explicit, we describe below what we feel are good practices for future studies (based on van Roekel et al., 2013). In order to obtain reliable data, researchers can put in effort to make sure that participants correctly interpret all items. Further, as mentioned earlier, explaining how to use the answer options can avoid getting highly skewed responses. Therefore, our advice is to have personal individual meetings with participants or in small groups, as this

makes it possible to thoroughly check participants' understanding of all procedures. Although we do not know of studies using video instructions, this may also be an effective, low-cost, and appealing medium to provide instructions to adolescents. We have included a checklist of what to include in instructions for participants (see Box 3).

Monitoring Scheme

Motivating participants to comply with the sampling procedures can be challenging, but it is a key indicator for the quality of the data collection. There are several best practices to increase compliance among adolescents (see Box 4). In our experience, the most effective practices in this age group are (1) providing cumulative incentives based on compliance, and (2) real-time monitoring and personal contact to stimulate participants.

RECOMMENDATIONS AND CONCLUSION

In this practical guide, based on our experiences and a review of recent AA studies, we described the most common issues with collecting AA data in adolescent samples and provided suggestions on how to deal with these issues. Moreover, in our review of recent studies, we noticed that many current studies on AA in youth lack details about the practicalities of data collection, such as how participants were instructed, how many items the total questionnaire comprised, how data were monitored, how much time participants had to complete an assessment. Apart from limiting the possibilities to replicate research findings, this information is essential to derive firm conclusions on what practices are effective in this specific age group. In order to improve AA research in adolescence and fine tune best practices recommendations, we have compiled a checklist on how to report on AA studies for researchers to use (see Box 5[2]). We encourage researchers to be open and transparent by reporting all steps and choices that were made in the research process, including pre-registration, material, and data. If there are space limitations in the manuscript, we encourage researchers to provide information about project details in

---

[2]A previous version of this checklist was shared on social media (i.e., Twitter and Facebook) to receive feedback from expert researchers in the field. Suggestions from colleagues were included in this checklist.

Box 3. Checklist for instructions

- Check whether participant has mobile Internet contract
- Check whether app works and provides notifications
- Train participants in using the app
- Instruct participants to keep smartphone near them during the study period, and to not use silent or do-not-disturb mode
- Explain in which situations participant is excused from filling out the momentary assessments (in traffic, during examinations, etc.)
- Highlight the importance of participant compliance
- Inform participants of the consequences of low/ high compliance
- Walk through all items:

  o Have adolescents explain the items themselves, to check whether they truly understand them
  o Explain difficult items
  o Explain that it is important that they really think about how they feel; and that it is important to realize that you can use the whole scale; and use the extremes only for, for example, "during occasions in which you have never felt happier"

Box 4. Best practices to increase compliance

- Increasing incentives; incentives based on minimum compliance
- Automated reminders for each assessment
- Real-time monitoring of compliance: contact participants after certain number of missings (e.g., 3 in a row)
- Catch-up days (i.e., providing the opportunity for participants to continue participation for some extra days to increase the total number of assessments)
- Frequent contact (school visits, individual instructions)
- Raffles for additional rewards among participants with high compliance (e.g., gift vouchers, iPads)
- Make the research fun and interactive by including game components or by creating groups of participants that can compete with each other

supplemental information files that can be stored on the Open Science Framework and can be referred to in the manuscript.

In this article, we have summarized some of the essentials of setting up and reporting about an AA study in adolescents. Yet, future methodological and theoretical research is needed to establish the best practices for studying youth in "the wild." First, we need to further examine how psychological processes can be best studied in daily life. With regard to construct validation, additional studies are needed to develop, test, and establish instruments that are brief, yet reliable and valid at the between-person and within-person level (e.g., see Adolf, Schuurman, Borkenau, Borsboom, & Dolan, 2014; Brose, Schmiedek, Koval, & Kuppens, 2014; Schuurman & Hamaker, 2018). Relatedly, more work needs to be done on how best to assess reliability and validity for measures that are used in intensive longitudinal designs. For instance, tools are needed that allow researchers to control for and deal with different sources of measurement error, including the person and the occasion (e.g., Hamaker, Schuurman,

& Zijlmans, 2017; Vogelsmeier, Vermunt, Van Roekel, & De Roover, 2018). A strong alliance between methodologists and applied researchers may be a fruitful approach, allowing methodologists to invest their time in developing techniques that can aid the advancement of psychological theories, and applied researchers to learn and apply the most innovative methods before they are implemented in standard software. Finally, at a more fundamental level, psychological theories need to account for the issue of timing when examining psychological processes, as different processes may operate at long versus short timescales (e.g., Granic & Patterson, 2006). Further, researchers need to think not only about how to apply theories to the individual (e.g., what relation holds for whom), but also how to best synthesize research findings from studies using different timescales into current theories of adolescent development.

The future of AA studies in adolescents—in addition to how much we can learn from this innovative methodology—will depend on further research into best practices, open and transparent reporting in AA publications, and strong alliances across a wide range of different disciplines and people, including researchers, school administrators, adolescents, clinicians, software developers, statisticians, and methodologists. We hope this review has provided some thoughts on how to build these bridges successfully.

Box 5. Checklist for reporting on AA studies

This checklist provides you with what we consider good practices of reporting in AA studies, above and beyond what is typically required in reporting in the Method section of empirical studies (e.g., APA).

Participants

☐ Report on specific recruitment methods (e.g., effective strategies to ensure school participation)
☐ A priori power analysis, based on sample size, number of assessments, and smallest effect size of interest

   ☐ Open Science: Share Monte Carlo simulation syntaxes and output files

Procedure

Technology

☐ Devices (including versions), when relevant (e.g., % of participants who use an IOS vs. Android smartphone)
☐ Software

Design of Study

☐ Prompt design (i.e., signal-contingent, interval-contingent, event-contingent; random vs. fixed intervals)
☐ Study duration
☐ Response window (i.e., how much time do the participants have to complete a questionnaire?)
☐ Total number of items per assessment
☐ Number of assessments per day

Participant Inclusion and Monitoring Protocol

☐ Exclusion or inclusion criteria
☐ The instructions that were given to participants
☐ Incentive structure (i.e., what compensation was provided to participants?)
☐ Monitoring scheme (i.e., if, how many, and when automatic reminders were sent; whether and under which circumstances participants were contacted, which messages were sent)
☐ Any problems during data collection
☐ Adjustments to protocol

Compliance

☐ Questionnaire duration (i.e., average questionnaire duration as well as measures of variability, e.g., SD, CI).
☐ Overall compliance (i.e., average number and percentage of completed assessments, including measure of variability such as SD, or a plot visualizing this variability)
☐ Reasons for noncompliance (e.g., technical problems, response window passed, illness reported)
☐ Time lag between prompt and completed assessment (i.e., is compliance based on assessments completed within a certain time window or on all assessments?)
☐ Patterns of noncompliance and missing data
☐ Were participants excluded for analyses based on compliance rates? If so, what cut-off was used?
☐ If relevant: Compliance after exclusion of participants

Materials

☐ Scale construction and transformation (including centering)
☐ Are participants asked about their current state (in-the-moment) or about the past hour(s)/day?
☐ Psychometric properties of scales (e.g., within-person reliability)

   ☐ Open Science: Share all items and syntaxes for scale construction and testing psychometric properties

Results

☐ Open science: Share anonymized data via open repositories
☐ Open science: Share all scripts for analyses

APPENDIX

TABLE A1
Differences in Level of Inconvenience Between Different Contexts in Study 1

Predictor	Mean reference group	Fixed effect	Random variance
Day of week			
Weekend (0 = week, 1 = weekend)	4.04 (.07)***	0.05 (.06)	.60 (.09)***
Company			
Alone (0 = alone, 1 = company)	4.19 (.07)***	−0.22 (.04)***	.13 (.05)***
Company (0 = friends, 1 = Family)	4.34 (.09)***	−0.42 (.08)***	.47 (.12)***
Company (0 = friends, 1 = classmates)	4.34 (.09)***	−0.63 (.08)***	.43 (.10)***
Company (0 = friends, 1 = acquaintances)	4.34 (.09)***	0.44 (.16)**	.71 (.33)***
Company (0 = family, 1 = classmates)	3.91 (.08)***	−0.20 (.07)***	.60 (.10)***
Company (0 = family, 1 = acquaintances)	3.91 (.08)***	0.86 (.15)***	.67 (.33)***
Company (0 = classmates, 1 = acquaintances)	3.71 (.07)***	1.09 (.15)***	.78 (.35)***
Location			
Location (0 = public, 1 = home)	4.49 (.08)***	−0.45 (.06)***	.36 (.07)***
Location (0 = public, 1 = school)	4.49 (.08)***	−0.68 (.07)***	.45 (.08)***
Location (0 = home, 1 = school)	4.04 (.07)***	−0.22 (.06)***	.58 (.09)***

***$p < .001$.

TABLE A2
Differences in Level of Inconvenience Between Different Contexts in Study 2

Predictor	Mean reference group	Fixed effect	Random variance
Day of week			
Weekend (0 = week, 1 = weekend)	2.17 (.08)***	.18 (.07)**	.47 (.09)***
Company			
Alone (0 = alone, 1 = company)	2.35 (.09)***	−.18 (.06)**	.20 (.06)***
Company (0 = friends, 1 = family)	2.26 (.09)***	−.10 (.05)*	.13 (.04)***
Company (0 = friends, 1 = classmates)	2.26 (.09)***	−.19 (.11)	.19 (.12)***
Company (0 = friends, 1 = acquaintances)	2.26 (.09)***	.18 (.18)	.31 (.30)***
Company (0 = family, 1 = classmates)	2.16 (.09)***	−.08 (.11)	.24 (.13)***
Company (0 = family, 1 = acquaintances)	2.16 (.09)***	.32 (.18)	.30 (.29)***
Company (0 = classmates, 1 = acquaintances)	2.04 (.12)***	.40 (.19)***	.34 (.31)***
Location			
Location (0 = other, 1 = home)	2.22 (.10)***	.00 (.06)	.13 (.05)***
Location (0 = other, 1 = school)	2.22 (.10)***	−.06 (.07)	.21 (.06)***
Location (0 = home, 1 = school)	2.23 (.09)***	−.07 (.06)	.29 (.06)***

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

REFERENCES

Adolf, J., Schuurman, N. K., Borkenau, P., Borsboom, D., & Dolan, C. V. (2014). Measurement invariance within and between individuals: A distinct problem in testing the equivalence of intra- and inter-individual model structures. *Frontiers in Psychology*, 5, 1–14. https://doi.org/10.3389/fpsyg.2014.00883

Andrewes, H. E., Hulbert, C., Cotton, S. M., Betts, J., & Chanen, A. M. (2017a). An ecological momentary assessment investigation of complex and conflicting emotions in youth with borderline personality disorder. *Psychiatry Research*, 252, 102–110. https://doi.org/10.1016/j.psychres.2017.01.100

Andrewes, H. E., Hulbert, C., Cotton, S. M., Betts, J., & Chanen, A. M. (2017b). Ecological momentary assessment of nonsuicidal self-injury in youth with borderline personality disorder. *Personality Disorders*, 8, 357–365. https://doi.org/10.1037/per0000205

Arslan, R. C. (2013). formr.org: v0.2.0. *Zenodo*. https://doi.org/10.5281/zenodo.33329

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 25, 359–388. https://doi.org/10.1080/10705511.2017.1406803

Bjorling, E. A., & Singh, N. (2017). Exploring temporal patterns of stress in adolescent girls with headache.

Stress and Health, *33*(1), 69–79. https://doi.org/10.1002/smi.2675

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. New York, NY: Guilford Press.

Bray, P., Bundy, A. C., Ryan, M. M., & North, K. N. (2017). Can in-the-moment diary methods measure health-related quality of life in Duchenne muscular dystrophy? *Quality of Life Research*, *26*, 1145–1152. https://doi.org/10.1007/s11136-016-1442-z

Broderick, J. E., & Vikingstad, G. (2008). Frequent assessment of negative symptoms does not induce depressed mood. *Journal of Clinical Psychology in Medical Settings*, *15*(4), 296–300. https://doi.org/10.1007/s10880-008-9127-6

Brose, A., Schmiedek, F., Koval, P., & Kuppens, P. (2014). Emotional inertia contributes to depressive symptoms beyond perseverative thinking. *Cognition and Emotion*, *29*, 1–12. https://doi.org/10.1080/02699931.2014.916252

Byrnes, H. F., Miller, B. A., Morrison, C. N., Wiebe, D. J., Woychik, M., & Wiehe, S. E. (2017). Association of environmental indicators with teen alcohol use and problem behavior: Teens' observations vs. objectively-measured indicators. *Health and Place*, *43*, 151–157. https://doi.org/10.1016/j.healthplace.2016.12.004

Byrom, B., Doll, H., Muehlhausen, W., Flood, E., Cassedy, C., McDowell, B., . . . McCarthy, M. (2017). Measurement equivalence of patient-reported outcome measure response scale types collected using bring your own device compared to paper and a provisioned device: Results of a randomized equivalence trial. *Value in Health*, *21*, 581–589. https://doi.org/10.1016/j.jval.2017.10.008

Chatfield, C. (2004). *The analysis of time series: An introduction*. Boca Raton, FL: Chapman & Hall/CRC.

Christensen, T. C., Barrett, L. F., Bliss-Moreau, E., Lebo, K., & Kaschub, C. (2003). A practical guide to experience-sampling procedures. *Journal of Happiness Studies*, *4*(1), 53–78. https://doi.org/10.1023/A:1023609306024

Collins, R. L., Martino, S. C., Kovalchik, S. A., D'Amico, E. J., Shadel, W. G., Becker, K. M., & Tolpadi, A. (2017). Exposure to alcohol advertising and adolescents' drinking beliefs: Role of message interpretation. *Health Psychology*, *36*, 890–897. https://doi.org/10.1037/hea0000521

Cruise, C. E., Broderick, J., Porter, L., Kaell, A., & Stone, A. A. (1996). Reactive effects of diary self-assessment in chronic pain patients. *Pain*, *67*(2), 253. https://doi.org/10.1016/0304-3959(96)03125-9

D'Amico, E. J., Martino, S. C., Collins, R. L., Shadel, W. G., Tolpadi, A., Kovalchik, S., & Becker, K. M. (2017). Factors associated with younger adolescents' exposure to online alcohol advertising. *Psychology of Addictive Behaviors*, *31*(2), 212–219. https://doi.org/10.1037/adb0000224

de Haan-Rietdijk, S., Voelkle, M. C., Keijsers, L., & Hamaker, E. L. (2017). Discrete- vs. continuous-time modeling of unequally spaced experience sampling method data. *Frontiers in Psychology*, *8*, 1849. https://doi.org/10.3389/fpsyg.2017.01849

Ebner-Priemer, U. W., & Trull, T. J. (2009). Ecological momentary assessment of mood disorders and mood dysregulation. *Psychological Assessment*, *21*, 463–475. https://doi.org/10.1037/a0017075

eMarketer. (2016). *Teens' ownership of smartphones has surged*. Retrieved January 2, 2018, from https://www.emarketer.com/Article/Teens-Ownership-of-Smartphones-Has-Surged/1014161

George, M. J., Russell, M. A., Piontak, J. R., & Odgers, C. L. (2017). Concurrent and subsequent associations between daily digital technology use and high-risk adolescents' mental health symptoms. *Child Development*, *89*, 78–88. https://doi.org/10.1111/cdev.12819

Granic, I., & Patterson, G. R. (2006). Toward a comprehensive model of antisocial development: A dynamic systems approach. *Psychological Review*, *113*(1), 101–131. https://doi.org/10.1037/0033-295X.113.1.101

Griffith, J. M., Silk, J. S., Oppenheimer, C. W., Morgan, J. K., Ladouceur, C. D., Forbes, E. E., & Dahl, R. E. (2018). Maternal affective expression and adolescents' subjective experience of positive affect in natural settings. *Journal of Research on Adolescence*, *28*, 537–550. https://doi.org/10.1111/jora.12357

Hamaker, E. L., Schuurman, N. K., & Zijlmans, E. A. O. (2017). Using a few snapshots to distinguish mountains from waves: Weak factorial invariance in the context of trait-state research. *Multivariate Behavioral Research*, *52*(1), 47–60. https://doi.org/10.1080/00273171.2016.1251299

Hamaker, E. L., & Wichers, M. (2017). No time like the present. *Current Directions in Psychological Science*, *26*(1), 10–15. https://doi.org/10.1177/0963721416666518

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*, *11*, 838–854. https://doi.org/10.1177/1745691616650285

Hasson, D., & Arnetz, B. B. (2005). Validation and findings comparing VAS vs. Likert scales for psychosocial measurements. *International Electronic Journal of Health Education*, *8*, 178–192.

Hennig, T., Krkovic, K., & Lincoln, T. M. (2017). What predicts inattention in adolescents? An experience-sampling study comparing chronotype, subjective, and objective sleep parameters. *Sleep Medicine*, *38*, 58–63. https://doi.org/10.1016/j.sleep.2017.07.009

Hennig, T., & Lincoln, T. M. (2018). Sleeping paranoia away? An actigraphy and experience-sampling study with adolescents. *Child Psychiatry and Human Development*, *49*, 63–72. https://doi.org/10.1007/s10578-017-0729-9

Heron, K. E., Everhart, R. S., McHale, S. M., & Smyth, J. M. (2017). Using mobile-technology-based ecological momentary assessment (EMA) methods with youth: A

systematic review and recommendations. *Journal of Pediatric Psychology*, 42, 1087–1107. https://doi.org/10.1093/jpepsy/jsx078

Hufford, M. R. (2007). Special methodological challenges and opportunities in ecological momentary assessment. In A. A. Stone, S. Shiffman, A. A. Atienza, & L. Nebeling (Eds.), *The science of real-time data capture: Self-reports in health research* (pp. 54–75). Oxford, UK: Oxford University Press.

Hufford, M. R., Shiffman, S., Paty, J., & Stone, A. A. (2001). Ecological momentary assessment: Real-world, real-time measurement of patient experience. In J. Fahrenberg & M. Myrtek (Eds.), *Progress in ambulatory assessment: Computer-assisted psychological and psychophysiological methods in monitoring and field studies* (pp. 69–92). Ashland, OH: Hogrefe & Huber.

Jaffray, P. (2018, October 4). *Taking stock with teens*. Retrieved September 18, 2018, from https://www.businessinsider.com/apple-iphone-popularity-teens-piper-jaffray-2018-4?international=true&r=US&IR=T

Katz, R. L., Felix, M., & Gubernick, M. (2014). Technology and adolescents: Perspectives on the things to come. *Education and Information Technologies*, 19, 863–886. https://doi.org/10.1007/s10639-013-9258-8

Keijsers, L., Hillegers, M. H. J., & Hiemstra, M. (2015). *Grumpy or Depressed research project* (Utrecht University Seed Project).

Keijsers, L., & van Roekel, E. (2018). Longitudinal methods in adolescent psychology: Where could we go from here? And should we? In L. B. Hendry & M. Kloep (Eds.), *Reframing adolescent research: Tackling challenges and new directions* (pp. 56–77). London, UK: Routledge. https://doi.org/10.4324/9781315150611-12

Kennisnet. (2015). *Monitor Jeugd en media (Youth and media)*. Retrieved October 26, 2017, from https://www.kennisnet.nl/publicaties/monitor-jeugd-en-media/

Kirchner, T., Magallón-Neri, E., Ortiz, M. S., Planellas, I., Forns, M., & Calderón, C. (2017). Adolescents' daily perception of internalizing emotional states by means of smartphone-based ecological momentary assessment. *Spanish Journal of Psychology*, 20, E71. https://doi.org/10.1017/sjp.2017.70

Klipker, K., Wrzus, C., Kauers, A., Boker, S. M., & Riediger, M. (2017a). Within-person changes in salivary testosterone and physical characteristics of puberty predict boys' daily affect. *Hormones and Behavior*, 95, 22–32. https://doi.org/10.1016/j.yhbeh.2017.07.012

Klipker, K., Wrzus, C., Rauers, A., & Riediger, M. (2017b). Hedonic orientation moderates the association between cognitive control and affect reactivity to daily hassles in adolescent boys. *Emotion*, 17, 497–508. https://doi.org/10.1037/emo0000241

Kolar, D. R., Huss, M., Preuss, H. M., Jenetzky, E., Haynos, A. F., Buerger, A., & Hammerle, F. (2017). Momentary emotion identification in female adolescents with and without anorexia nervosa. *Psychiatry Research*, 255, 394–398. https://doi.org/10.1016/j.psychres.2017.06.075

Kramer, I., Simons, C. J. P., Hartmann, J. A., Menne-Lothmann, C., Viechtbauer, W., Peeters, F., . . . Wichers, M. (2014). A therapeutic application of the experience sampling method in the treatment of depression: A randomized controlled trial. *World Psychiatry*, 13(1), 68–77. https://doi.org/10.1002/wps.20090

Kranzler, A., Fehling, K. B., Lindqvist, J., Brillante, J., Yuan, F., Gao, X., . . . Selby, E. A. (2017). An ecological investigation of the emotional context surrounding nonsuicidal self-injurious thoughts and behaviors in adolescents and young adults. *Suicide and Life-Threatening Behavior*, 48, 149–159. https://doi.org/10.1111/sltb.12373

Kroeze, R., van der Veen, D. C., Servaas, M. N., Bastiaansen, J., Voshaar, R. O., Borsboom, D., . . . Riese, H. (2017). Personalized feedback on symptom dynamics of psychopathology: A proof-of-principle study. *Journal of Person-Oriented Research*, 3(1), 1–10. https://doi.org/10.17505/jpor

Larson, R. W. (1983). Adolescents' daily experience with family and friends: Contrasting opportunity systems. *Journal of Marriage and Family*, 45, 739–750. https://doi.org/10.2307/351787

Larson, R. W., & Csikszentmihalyi, M. (1983). The experience sampling method. *New Directions for Methodology of Social and Behavioral Science*, 15, 41–56.

Larson, R. W., Csikszentmihalyi, M., & Graef, R. (1980). Mood variability and the psychological adjustment of adolescents. *Journal of Youth and Adolescence*, 9, 469–490. https://doi.org/10.1007/BF02089885

Lennarz, H. K., Lichtwarck-Aschoff, A., Finkenauer, C., & Granic, I. (2017a). Jealousy in adolescents' daily lives: How does it relate to interpersonal context and well-being? *Journal of Adolescence*, 54, 18–31. https://doi.org/10.1016/j.adolescence.2016.09.008

Lennarz, H. K., Lichtwarck-Aschoff, A., Timmerman, M. E., & Granic, I. (2017b). Emotion differentiation and its relation with emotional well-being in adolescents. *Cognition and Emotion*, 32, 1–7. https://doi.org/10.1080/02699931.2017.1338177

Lewis, M. D. (2000). Emotional self-organization at three time scales. In M. D. Lewis & I. Granic (Eds.), *Emotion development and self-organization* (pp. 37–69). New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9780511527883.004

Lipperman-Kreda, S., Gruenewald, P. J., Grube, J. W., & Bersamin, M. (2017). Adolescents, alcohol, and marijuana: Context characteristics and problems associated with simultaneous use. *Drug and Alcohol Dependence*, 179, 55–60. https://doi.org/10.1016/j.drugalcdep.2017.06.023

McCormack, H. M., Horne, D. J., & Sheather, S. (1988). Clinical applications of visual analogue scales: A critical review. *Psychological Medicine*, 18, 1007–1019. https://doi.org/10.1017/S0033291700009934

Moeller, J., Dietrich, J., Eccles, J. S., & Schneider, B. (2017). Passionate experiences in adolescence: Situational variability and long-term stability. *Journal of*

Research on Adolescence, 27, 344–361. https://doi.org/10.1111/jora.12297

Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling: A Multidisciplinary Journal, 9,* 599–620. https://doi.org/10.1207/S15328007SEM0904_8

Myin-Germeys, I., Oorschot, M., Collip, D., Lataster, J., Delespaul, P., & Van Os, J. (2009). Experience sampling research in psychopathology: Opening the black box of daily life. *Psychological Medicine, 39,* 1533–1547. https://doi.org/10.1017/S0033291708004947

Odgers, C. L., & Russell, M. A. (2017). Violence exposure is associated with adolescents' same- and next-day mental health symptoms. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 58,* 1310–1318. https://doi.org/10.1111/jcpp.12763

Piontak, J. R., Russell, M. A., Danese, A., Copeland, W. E., Hoyle, R. H., & Odgers, C. L. (2017). Violence exposure and adolescents' same-day obesogenic behaviors: New findings and a replication. *Social Science and Medicine, 189,* 145–151. https://doi.org/10.1016/j.socscimed.2017.07.004

R Core Team. (2017). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from November 6, 2018, http://www.R-project.org/

Rauschenberg, C., van Os, J., Cremers, D., Goedhart, M., Schieveld, J. N. M., & Reininghaus, U. (2017). Stress sensitivity as a putative mechanism linking childhood trauma and psychopathology in youth's daily life. *Acta Psychiatrica Scandinavica, 136,* 373–388. https://doi.org/10.1111/acps.12775

Ross, C. S., Brooks, D. R., Aschengrau, A., Siegel, M. B., Weinberg, J., & Shrier, L. A. (2018). Positive and negative affect following marijuana use in naturalistic settings: An ecological momentary assessment study. *Addictive Behaviors, 76,* 61–67. https://doi.org/10.1016/j.addbeh.2017.07.020

Roy Morgan Research. (2016). *9 in 10 Aussie teens now have a mobile.* Retrieved November 6, 2017, from http://www.roymorgan.com/findings/6929-australian-teenagers-and-their-mobile-phones-June-2016-201608220922

Salvy, S.-J., Feda, D. M., Epstein, L. H., & Roemmich, J. N. (2017a). Friends and social contexts as unshared environments: a discordant sibling analysis of obesity- and health-related behaviors in young adolescents. *International Journal of Obesity, 41,* 569–575. https://doi.org/10.1038/ijo.2016.213

Salvy, S.-J., Feda, D. M., Epstein, L. H., & Roemmich, J. N. (2017b). The social context moderates the relationship between neighborhood safety and adolescents' activities. *Preventive Medicine Reports, 6,* 355–360. https://doi.org/10.1016/j.pmedr.2017.04.009

Schultzberg, M., & Muthén, B. O. (2017). Number of subjects and time points needed for multilevel time series

analysis: A Monte Carlo study of dynamic structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 25*(4), 495–515. https://doi.org/10.1080/10705511.2017.1392862

Schuurman, N. K., & Hamaker, E. L. (2018). Measurement error and person-specific reliability in multilevel autoregressive modeling. *Psychological Methods.* Advance online publication. https://doi.org/10.1037/met0000188

Scollon, C. N., Prieto, C.-K., & Diener, E. (2009). Experience sampling: Promises and pitfalls, strength and weaknesses. In E. Diener (Ed.), *Assessing well-being* (pp. 157–180). Dordrecht, The Netherlands: Springer. https://doi.org/10.1007/978-90-481-2354-4

Shiyko, M. P., Perkins, S., & Caldwell, L. (2017). Feasibility and adherence paradigm to ecological momentary assessments in urban minority youth. *Psychological Assessment, 29,* 926–934. https://doi.org/10.1037/pas0000386

Silvia, P. J., Kwapil, T. R., Walsh, M. A., & Myin-Germeys, I. (2014). Planned missing-data designs in experience-sampling research: Monte Carlo simulations of efficient designs for assessing within-person constructs. *Behavior Research Methods, 46*(1), 41–54. https://doi.org/10.3758/s13428-013-0353-y

Statista. (2017). *UK: Smartphone ownership by age 2017.* Retrieved November 6, 2017, from https://www.statista.com/statistics/271851/smartphone-owners-in-the-united-kingdom-uk-by-age/

Thai, S., & Page-Gould, E. (2017). ExperienceSampler: An open-source scaffold for building smartphone apps for experience sampling. *Psychological Methods, 23,* 729–739. https://doi.org/10.1037/met0000151

Treloar, H., & Miranda, R. (2017). Craving and acute effects of alcohol in youths' daily lives: Associations with alcohol use disorder severity. *Experimental and Clinical Psychopharmacology, 25,* 303–313. https://doi.org/10.1037/pha0000133

Tucker-Seeley, K. R. (2008). *The effects of using Likert vs. visual analogue scale response options on the outcome of a web-based survey of 4th through 12th grade students: Data from a randomized experiment.* Boston College. Retrieved from http://dlib.bc.edu/islandora/object/bc-ir:101314

Vaessen, T., van Nierop, M., Decoster, J., Delespaul, P., Derom, C., de Hert, M., . . . Myin-Germeys, I. (2017). Is sensitivity to daily stress predictive of onset or persistence of psychopathology? *European Psychiatry, 45,* 167–173. https://doi.org/10.1016/j.eurpsy.2017.07.002

van Roekel, E., Goossens, L., Verhagen, M., Wouters, S., Engels, R. C. M. E., & Scholte, R. H. J. (2013). Loneliness, affect, and adolescents' appraisals of company: An Experience Sampling Method (ESM) study. *Journal of Research on Adolescence, 24,* 350–363. https://doi.org/10.1111/jora.12061

Van Roekel, E., Vrijen, C., Heininga, V. E., Masselink, M., Bos, E. H., & Oldehinkel, A. J. (2017). An exploratory randomized controlled trial of personalized lifestyle advice and tandem skydives as a means to reduce

anhedonia. *Behavior Therapy*, *48*(1), 76–96. https://doi. org/10.1016/j.beth.2016.09.009

Voelkle, M. C., Oud, J. H. L., von Oertzen, T., & Lindenberger, U. (2012). Maximum likelihood dynamic factor modeling for arbitrary N and T using SEM. *Structural Equation Modeling: A Multidisciplinary Journal*, *19*, 329–350. https://doi.org/10.1080/10705511.2012.687656

Vogelsmeier, L. V. D. E., Vermunt, J. K., Van Roekel, E., & De Roover, K. (2018). Latent Markov factor analysis for exploring measurement model changes in time-intensive longitudinal studies. Manuscript submitted for publication.

Wallace, M. L., McMakin, D. L., Tan, P. Z., Rosen, D., Forbes, E. E., Ladouceur, C. D., … Silk, J. S. (2017). The role of day-to-day emotions, sleep, and social interactions in pediatric anxiety treatment. *Behaviour Research and Therapy*, *90*(Suppl C), 87–95. https://doi.org/10.1016/j.brat.2016.12.012

Wang, Y., Cham, H., Aladin, M., & Yip, T. (2017). Parental cultural socialization and adolescent private regard: Exploring mediating pathways through daily experiences. *Child Development*. Advance online publication. https://doi.org/10.1111/cdev.12911

Wen, C. K. F., Schneider, S., Stone, A. A., & Spruijt-Metz, D. (2017). Compliance with mobile ecological momentary assessment protocols in children and adolescents: A systematic review and meta-analysis. *Journal of Medical Internet Research*, *19*, e132. https://doi.org/10.2196/jmir.6641

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Characteristics of Studies Using AA in Adolescent Samples Published in 2017.