

A novel self-rating instrument designed for long-term, app-based monitoring of ADHD symptoms: A mixed-methods development and validation study

DIGITAL HEALTH
Volume 10: 1–16
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076241280037
journals.sagepub.com/home/dhj



Amanda Bäcker¹ , David Forsström^{1,2}, Louise Hommerberg³,
Magnus Johansson¹, Ida Hensler¹ and Philip Lindner¹

Abstract

Background: Regular outcome monitoring is essential for effective attention deficit hyperactivity disorder (ADHD) treatment, yet routine care often limits long-term contacts to annual visits. Smartphone apps can complement current practice by offering low-threshold, long-term sustainable monitoring capabilities. However, special considerations apply for such measurement which should be anchored in stakeholder preferences.

Methods: This mixed-methods study engaged 13 experienced clinicians from Region Stockholm in iterative qualitative interviews to inform development of an instrument for app-based ADHD monitoring: the mHealth scale for Continuous ADHD Symptom Self-monitoring (mCASS). A subsequent survey, including the mCASS and addressing app-based monitoring preferences, was administered to 397 individuals with self-reported ADHD. Psychometric properties of the mCASS were explored through exploratory factor analysis and examinations of internal consistency. Concurrent validity was calculated between the mCASS and the Adult ADHD Self-Report Scale-V1.1 (ASRS-V1.1). Additional quantitative analyses included summary statistics and repeated-measures ANOVAs.

Results: Clinicians identified properties influencing willingness to use and adherence including content validity, clinical relevance, respondent burden, tone, wording and preferences for in-app results presentation. The final 12-item mCASS version demonstrated four factors covering everyday tasks, productivity, rest and recovery and interactions with others, explaining 47.4% of variance. Preliminary psychometric assessment indicated satisfactory concurrent validity ($r = .595$) and internal consistency ($\alpha = .826$).

Conclusions: The mCASS, informed by clinician and patient experiences, appears to be valid for app-based assessment of ADHD symptoms. Furthermore, insights are presented regarding important considerations when developing mobile health (mHealth) instruments for ADHD individuals. These can be of value for future, similar endeavours.

Keywords

ADHD, tele-monitoring, smartphone application, mHealth, instrument development, mixed-methods

Submission date: 2 February 2024; Acceptance date: 8 August 2024

¹Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, & Stockholm Health Care Services, Region Stockholm, Stockholm, Sweden

²Department of Psychology, Uppsala University, Uppsala, Sweden

³Takeda Pharma AB, Stockholm, Sweden

Corresponding author:

Amanda Bäcker, Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet & Stockholm Health Care Services, Norra Stationsgatan 69, 113 64 Stockholm, Sweden.
Email: Amanda.Backer@ki.se



Introduction

Attention deficit hyperactivity disorder (ADHD) is a neurodevelopmental disorder characterized by symptoms of inattention, hyperactivity and impulsivity, with population prevalence estimates around 5%.^{1,2} As treatment of ADHD typically involves long-term pharmacotherapy, continuous outcome monitoring is essential to ensure maintained effectiveness, as well as to identify new needs that can arise as patients move through life. Despite this, in routine care for individuals with ADHD, long-term contacts are commonly limited to annual follow-up visits,³ a practice that risks overlooking important periodization in symptoms,⁴ as well as cross-sectional information being affected by recall bias due to the long periods in-between visits.⁵ In addition, there is some evidence from research outside of ADHD suggesting that regular monitoring may promote symptom improvement through facilitation of self-reflection.^{6,7} In light of this, capabilities for regular monitoring outside of periodical visits is of great interest to the management of ADHD and has the potential to complement and improve routine clinical care. In parallel with technological advances and the growing ubiquity of smartphones worldwide, the past decade has seen increased use of mobile health applications, commonly referred to as mobile health (mHealth) apps, dedicated to the management of various mental health disorders. ADHD is no exception to this. A systematic review by Păsărelu et al.⁸ found 109 publicly available apps designed specifically for treating or assessing ADHD. Out of the treatment apps identified, around 12% included functionalities for outcome monitoring. However, concerns have been raised regarding the efficacy of many of the available apps for mental health, and scientific evidence has considerable limitations.⁹ This is also true in the context of ADHD mHealth. In their systematic review, Păsărelu et al.⁸ notably identified that only 16% of included apps presented information about empirical support, such as information about the process of development or about the validity of contents.

In developing an app-based instrument for (ADHD) outcome monitoring that can complement conventional clinical follow-ups, there are some important points to consider. The American Psychiatric Association App evaluation model underlines that to be deemed high-quality, mental health apps should promote user adherence, as well as collect meaningful data.¹⁰ Previous efforts to achieve meaningful monitoring with high adherence has utilized so-called ecological momentary assessment (EMA),¹¹ a method for collecting data, e.g., information on symptoms, behaviours and mood, characterized by measurement in the moment and within the context that experiences of interest naturally occur, thus minimizing the influence of biases.¹² Results from meta-analytical research on EMA have demonstrated compliance rates of around 80% in both adult and adolescent samples,^{13,14} which should arguably be considered

satisfactory. However, important questions remain regarding the feasibility of EMA monitoring approaches when used for longer periods of time: compliance estimates from EMA research primarily comes from high-intensity, short-duration studies, i.e., with many measures collected over a short period of time. In a systematic review by Miguelez-Fernandez et al.,¹⁵ 23 publications describing the use of EMA to evaluate ADHD were identified. Among these, assessment frequencies ranged from every 30 minutes to once per day; importantly, protocol durations ranged from 2 to only 28 days. It was also noted that a majority of the studies included offered some form of monetary compensation to participants. Other approaches will likely be necessary for regular outcome monitoring with the purpose of providing complementary information to long-term clinical contact, especially if adherence is not incentivized financially.

For technology to be useful and engaging, the needs and preferences of the targeted end-users must be considered. Previous research on ADHD patients' attitudes and preferences regarding remote monitoring using digital solutions has shown that, while symptom monitoring is viewed as having potential to enhance regular care received through the sharing of data with clinicians, an initial requirement for this is that patients experience that collected data is relevant to their care.^{16,17} It is however unclear to what degree previous research has been anchored in patients' views and experiences when developing monitoring instruments.¹⁸ To address this, employment of human-centred design principles is highly suitable, and this practice is on the rise within the broader mHealth field.¹⁹ A key issue when involving stakeholders in research is that of representativeness and sampling bias.²⁰ mHealth studies relying only on self-recruited patients risk including only those already interested and well-versed in digital tools and solutions. This may result in neglecting needs within the broader target population. Earlier works, outside the field of ADHD, have tackled this issue by involving clinicians and caregivers as informants,^{21,22} since this group represents both a direct stakeholder, in their role as potential recipients of app collected information, and a knowledgeable source of information about the target population from their clinical experience.

The aim of the current study was to develop a new instrument, the mHealth scale for Continuous ADHD Symptom Self-monitoring (mCASS), with the specific aim of allowing long-term app-based assessment through a publicly released smartphone application.

The study was conducted in two parts, constituting a mixed-methods approach. First, to address the knowledge gaps described above, we set out to systematically collect and synthesize views and opinions from clinicians experienced in treating patients with ADHD, the goal of which was to inform parallel development process. This process resulted in further research questions which guided the second part, wherein we examined the psychometric

properties of the mCASS among individuals with self-reported ADHD, as well as some of the outstanding issues raised by clinicians. In addition to presenting the mCASS, a new instrument suitable for long-term monitoring of ADHD, the aim was also to demonstrate and summarize detailed lessons learned from this development process, to inform similar mHealth initiatives in the future.

Methods

Design and ethics

The current study consists of two parts, a qualitative and a quantitative, conducted in sequence, with the former directly influencing the latter. As the qualitative interviews collected no sensitive personal data (or involved any intervention), this part fell outside the scope of the Swedish Ethical Review Act. However, the Swedish Ethical Review Authority issued an advisory statement that they saw no ethical issues in the study (2022-01690-01). Ethical approval for the quantitative survey study was issued separately (2023-03313-01).

Interview procedure and sample

The first part of the study involved three iterations of semi-structured interviews, individually or in small groups, with clinicians experienced in treating patients with ADHD. Clinician recruitment was initiated in August 2022 by an intranet post within the public healthcare provider of Region Stockholm. Interested clinicians were directed to an online survey serving as an initial screening tool. This survey assessed the clinicians' background, typical clinical duties related to ADHD treatment and their perspectives on the potential of mHealth tools to improve ADHD treatment. The latter was surveyed to ensure that participating clinicians were motivated to assist in the development process. Eligibility criteria included self-reported experience in treating patients with ADHD and current employment within Region Stockholm public healthcare. Purposive stratified sampling was employed to ensure representation across relevant subgroups, considering age, profession and years of experience in treating patients with ADHD. Participation also required approval from the clinic manager, due to interviews being scheduled during regular work hours. Out of 24 interested clinicians, 14 were selected (during a six-month window) and were subsequently provided with, and signed, informed consent. Table 1 shows background information and participation patterns of participating clinicians. After the informed consent was obtained, clinicians were scheduled for interviews. One participant (P13), who

Table 1. Overview of interview participants for all iterations.

Participant	Age	Occupation	Years of Experience	Iteration		
				1	2	3
P1	41-50	Para-medical	>10	D	D	-
P2	41-50	Para-medical	>10	D	D	-
P3	41-50	Para-medical	>10	D	D	D (G)
P4	51-60	Nursing	>10	IP	IP	IP
P5	51-60	Nursing	4-6	D	D	IP (G)
P6	51-60	Para-medical	4-6	D	D	IP (G)
P7	41-50	Nursing	7-9	D	D	D
P8	61 or above	Para-medical	>10	D	D	-
P9	20-30	Nursing	4-6	IP	IP	D (G)
P10	51-60	Nursing	4-6	D	D	D
P11	41-50	Para-medical	7-9	IP	IP	IP (G)
P12	31-40	Medical	1-3	D	-	-
P14	61 or above	Nursing	7-9	-	-	IP (G)

Note. D: digital interview; IP: in-person interview; G: group interview. P13 provided informed consent to partake in the study but subsequently never responded to any contact attempts for interview scheduling. Para-medical professions included psychologists, occupational therapists and counsellors.

initially consented, did not respond to any interview scheduling attempts.

The three interview iterations were conducted during an eight-month period, from September 2022 to May 2023. Each wave of data collection informed subsequent development and revisions of the mCASS. Revised versions were then brought back for further feedback from participating clinicians until a final draft was reached. Figure 1 shows an overview of the iterative development process. During the first round of interviews, clinicians were presented with initial prototype items to facilitate discussions. Clinicians were also presented with more detailed information about the overarching goal of the interview process, i.e., developing an app-based

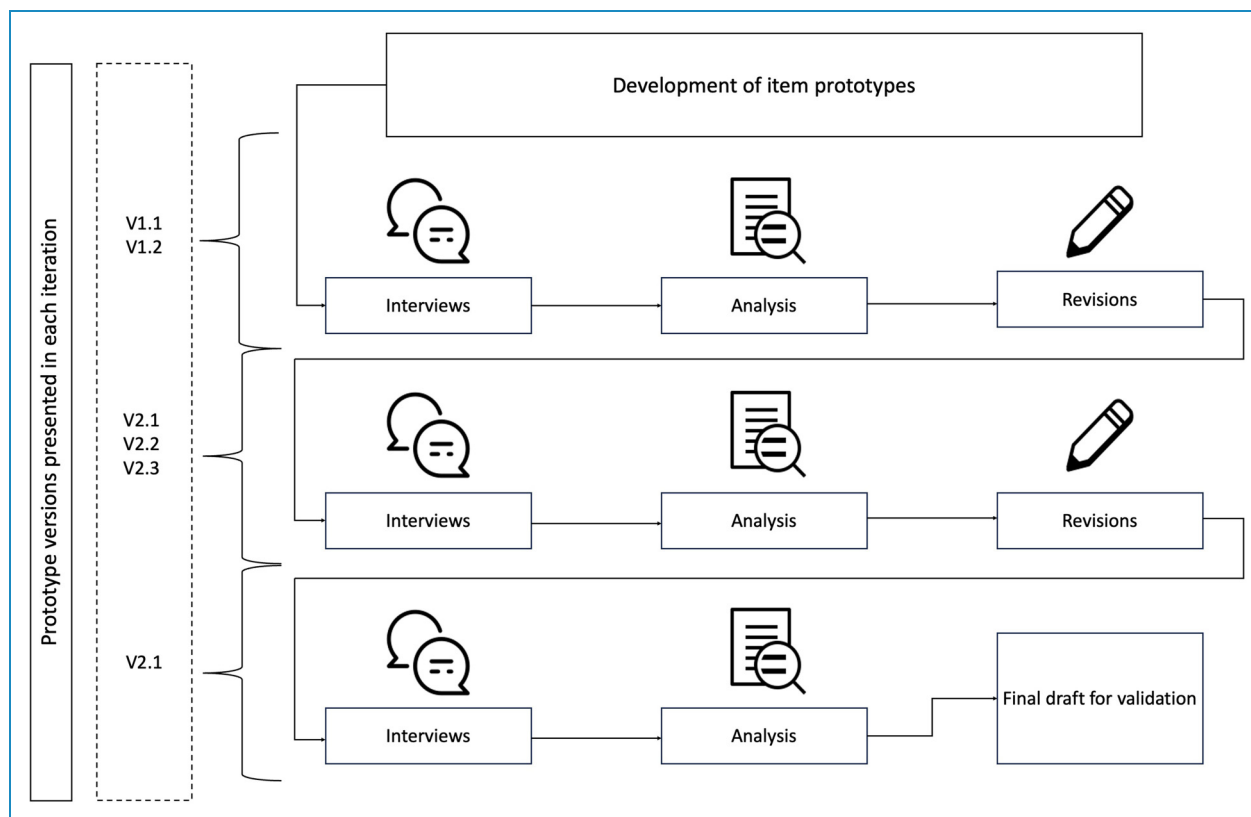


Figure 1. Overview of iterative development process.

instrument for longitudinal monitoring of ADHD symptoms. They were furthermore informed that the intended purpose of the instrument was both to serve as a tool for patients to gain insights about themselves and to enable patients to share relevant information with clinicians overseeing his or her care. Interview protocols for each iteration were designed to centre discussions around the content and design of a smartphone-based instrument for longitudinal monitoring of ADHD symptoms (Table 2).

In the first iteration, 12 participants were interviewed (see Table 1). In the second iteration, one participant was too short on time to participate; thus, 11 participants took part. The third iteration comprised eight initial participants and a newly recruited one, since three participants explicitly declined to take part. Stated reasons included going on parental leave, switching jobs and being short on time. One additional participant did not take part in the third iteration but did so without stating a reason and did not respond to contact attempts.

To reduce the risk of sampling bias and attrition, participants were offered the choice of participating in-person in the workplace setting or digitally (see Table 1). For digital interviews, participants chose their setting themselves. All interviews were audio recorded and transcribed verbatim by either the first author or a professional

Table 2. Interview protocols for all iterations.

Iteration	Interview Protocol
1–2	<p>(Same questions were posed for each prototype)</p> <ul style="list-style-type: none"> • What are your spontaneous impressions? Is there anything in particular you like/dislike? • How can we increase the likelihood that patients will regularly use the instrument? • For you as a clinician, how interesting would it be to have access to the information generated by this instrument? Why? What could you use it for? • Do you have any suggestions for changes to be made? Is anything missing? Is anything redundant? • Which prototype version do you prefer? Why? • If not already covered: what are your opinions on an appropriate recall period and/or monitoring frequency?
3	<ul style="list-style-type: none"> • If results from the app-based instrument could be summarized in a report, what would you as clinicians want to see in such a report? Why? • How should the results be presented? • What information is important? Why?

transcriber. No field notes were used. Interview sessions were approximately one hour long. Transcripts were not shared with participants.

Qualitative analysis

The transcribed material from each iteration was analysed using inductive content analysis, drawing upon the framework proposed by Graneheim and Lundman.²³ By searching for patterns within the codes, categories emerged that guided the decision-making when implementing changes to the prototype item.

Researcher characteristics and reflexivity statement. All interviews and qualitative analyses were conducted by the first author (AB) alone. AB is, at the time of the study, a female PhD student conducting this research as part of her thesis. She is a licensed clinical psychologist, trained in interview techniques, and has been clinically active between the years 2018 and 2022. She had no relationship to participating clinicians apart from also being employed by Region Stockholm public healthcare. While the first author's own clinical experience and preexisting knowledge concerning ADHD may have influenced the information gathered and synthesized throughout this study, continuous efforts have been made to be aware of these preconditions and allow for participants' own views to emerge during interviews and analysis.

Development of prototypes

Initial prototype items for the mCASS were developed following a systematic approach that integrated insights from various sources. A semi-structured literature review was conducted, focusing on items utilized in EMA methodologies for ADHD symptoms, on commonly used clinical instruments and on the diagnostic criteria for ADHD as defined in the *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition (DSM-5). This informed establishment of relevant content and criteria for the prototypes. Subsequently, two sets of prototype instruments were crafted (see Table S1 in Supplementary material 1). This was also done with consideration given to aspects thought to affect an instrument's suitability for app-based long-term monitoring, the aim being to facilitate discussions about different instrument characteristics. One prototype (V1.1) was developed to be shorter and more general, staying closer to measures used in previous literature about EMA for ADHD. The other (V1.2) was developed to be longer and more detailed, measuring symptoms in different areas of life separately. No recall period was set, as this was a dimension that clinicians were to provide input on.

Survey procedure and sample

Once a final draft of the mCASS had been developed, informed by the views expressed in the iterative interview process, a total of 400 individuals were recruited to take part in a quantitative survey study with the purpose of

examining the psychometric properties of the novel instrument. To this end, the final draft was translated from Swedish to English. This was done in collaboration between two of the authors.

Participants were recruited from Prolific.co, an online platform connecting researchers and potential study participants that has seen extensive use in similar research.^{24,25} Sign-up to Prolific requires users to be at least 18 years old. Utilizing Prolific's filtering tool, the survey study was made available only to individuals with self-reported ADHD who lived in English-speaking countries (United Kingdom, United States, Australia, New Zealand and Canada). Furthermore, the sample was balanced for sex of using a feature integrated into Prolific, ensuring equal representation of male and female participants. Participants received compensation for their time, corresponding to an hourly rate of around 100 SEK (approximately 8.50 euros).

The survey posted to Prolific was created and distributed using the REDCap web survey application. To take the survey, participants were first required to sign informed consent. The survey included the translated version of the mCASS. To enable testing of convergent validity, the survey also comprised the Adult ADHD Self-Report Scale-V1.1 (ASRS-V1.1) Screener, a commonly used screening form consisting of six items designed to measure ADHD symptoms.²⁶

Moreover, the purpose of the survey was also to address some additional issues raised and not resolved by clinicians during the interview phase, including issues relating to monitoring frequency and data sensitivity. Participant preferences regarding monitoring frequency were examined using three types of questions. Firstly, after answering the mCASS, the following question was posed in relation to each mCASS item: 'Is once a week an appropriate frequency to ask about this type of experience, despite any variations during the week?'. Answers were recorded on a categorical scale with the following response options: 'No, should ask more frequently', 'Yes' and 'No, should ask less frequently'. Participants were then asked about the importance of being able to individualize the monitoring frequency. Answers were recorded using a visual analogue scale (VAS) ranging from 0 (Not at all important) to 100 (Very important). Following this, participants were asked to indicate their preferred monitoring frequency for utilizing the full monitoring instrument, using a five-point ordinal scale (response options: 'Daily', 'Weekly', 'Monthly', 'Yearly' and 'Other frequency'). Regarding data sensitivity, participants were asked to rate the perceived sensitivity of various types of personal data (ADHD symptoms, pulse, step count and sleep) using a VAS (0 = Not at all sensitive, 100 = Very sensitive). Lastly, the survey also inquired about the face validity of each item in the mCASS ('To what extent do you feel that the question above captures a difficulty that individuals with ADHD face?'). Answers

Table 3. Demographic characteristics of survey participants.

Participant characteristic	<i>n</i> (%)	<i>M</i>	<i>SD</i>
Age	395	34.2	10
Sex	397		
Male	199 (50.1)		
Female	198 (49.9)		
Gender	397		
Male	191 (48.1)		
Female	188 (47.4)		
Other	17 (4.3)		
Prefer not to say	1 (0.3)		
First language	397		
English	356 (89.7)		
Other	41 (10.3)		
Country of residence	397		
United Kingdom	319 (80.4)		
United States	47 (11.8)		
Australia	16 (4.0)		
Canada	9 (2.3)		
New Zealand	6 (1.5)		
Employment status	369		
Full-time	213 (57.7)		
Part-time	69 (18.7)		
Unemployed (and job seeking)	37 (10.0)		
Not in paid work (e.g., homemaker, retired or disabled)	33 (8.9)		
Other	12 (3.3)		
Due to start new job within next month	5 (1.4)		
Student status	368		

(continued)

Table 3. Continued.

Participant characteristic	<i>n</i> (%)	<i>M</i>	<i>SD</i>
No	281 (76.4)		
Yes	87 (23.6)		

Note. *n* = 397. Some missing data for the following variables: age, employment status and student status. This is due to some Prolific user choosing not to report all demographic information during sign-up to Prolific.

were recorded using a VAS (0 = not at all, 100 = completely).

The survey also collected information about participants' age and gender. No medication-focused items were included, as examination of how pharmacological ADHD treatment could affect mCASS ratings was outside of the scope of this first development and validation study. Importantly, the companion app that will include the mCASS will target both medicated and unmedicated ADHD individuals. The survey is presented in its entirety in Supplementary material 2.

In addition to survey data, demographic data as reported in the Prolific screening was exported from Prolific. This contained information about age, sex, country of residence and birth, first language, employment status and student status.

Statistical analyses

Prior to statistical analyses, the data was reviewed to ensure adequate quality. As in other studies,^{27,28} near-zero variation in responses was used to signal potentially low-quality data. In the current study, we calculated the percentage of responses to the mCASS that was equal to the mode. In cases where this percentage was >70, responses to the full survey were examined. This process resulted in three participants being excluded from analyses. Thus, the final sample size totalled *n* = 397. Data was missing on three demographic variables: age (*n* = 395), employment status (*n* = 369) and student status (*n* = 368) (Table 3).

The validity of the mCASS was examined in steps. First, summary statistics were calculated for face validity and preferred monitoring frequency ratings for each item. Next, a repeated measures ANOVA (with Bonferroni-adjusted post hoc tests) was conducted to compare perceived sensitivity between different data types. Finally, to examine the underlying factor structure of the mCASS, exploratory factor analysis (EFA) was conducted using the minimum residuals extraction method. Oblimin rotation was applied to allow factors to be correlated.^{29,30} To determine the number of factors to retain, parallel analysis was

performed.³¹ Internal consistency was calculated using Cronbach's alpha, and concurrent validity was examined by calculating Pearson product moment correlation coefficient between the new instrument and ASRS-V1.1 Screener total score. All analyses were conducted using JAMOVI (version 2.3.28.0), running on the R statistical environment.

Results

Qualitative findings

Results from qualitative analyses are described for each iteration of interviews. For each iteration, emerging categories are presented in descending order based on the number of interviews they occurred in, from most to least frequently discussed. Non-recurrence of categories in subsequent iterations reflects that the issues covered were not brought up by participants again or were only mentioned as resolved.

First iteration of interviews

During the first iteration, the two sets of prototype items were presented to clinicians. In the following sections, we present the categories that emerged from analysis. Subsequently, changes to the prototypes based on these insights are described.

Category 1: content validity. For both prototypes, all clinicians expressed that the items were clearly related to ADHD symptoms and associated difficulties and would provide clinicians with useful information, although in slightly contrasting manners. Regarding V1.1, it was commonly stated that the instrument covered issues that are usually asked about by clinicians during visits. In contrast, V1.2 was described as capturing problems that are usually described by patients. Some clinicians also expressed that V1.2 was 'too detailed' and collected more information than necessary, as well as being 'unbalanced' and covering some areas of life more extensively than others.

Category 2: respondent burden

Subcategory: length. Throughout all interviews in the first iteration, clinicians brought up the issue of instrument length and as an important factor for patients' willingness to use the monitoring instrument. V1.2 in particular raised concerns among clinicians, while the 'conciseness' of V1.1 was viewed as a strength from a monitoring adherence perspective.

Subcategory: cognitive availability. Multiple comments throughout the first iteration also revealed that clinicians viewed cognitive availability to be an important aspect to consider. For example, many liked that the items in V1.2

were grouped based on areas of life addressed. Clinicians expressed that this would limit the number of behaviours and situations to consider when responding and therefore make it easier for patients to use the mCASS. In contrast, several items in V1.1 were seen as 'too abstract' and therefore difficult for patients to answer. Expressions such as 'tasks and activities' were considered as too general and therefore hard to understand and respond accordingly to. Additionally, opinions differed regarding using the same phrase in the beginning of all V1.1 items. Some of the clinicians stated that this was easier to read and would not burden patients' working memory. Working memory impairments and difficulties with maintained attention were generally seen as important patient characteristics to consider, as this could affect adherence to regular monitoring.

Subcategory: monitoring frequency and recall period.

Throughout interviews, clinicians generally viewed once a week to be a suitable monitoring frequency, although there was some hesitation surrounding the feasibility of this. Relating to respondent burden, some clinicians stated that this could be too frequent and stated that once a month might be more realistic. However, when considering the appropriate recall period for the mCASS, a clear majority preferred this to be one week and thus found a monitoring frequency of once a week to be more logical. It was expressed that longer recall periods would make it impossible for patients to remember how they had experienced their symptoms and would result in more challenges for patients. High-frequency monitoring, such as once a day, was considered an 'unthinkable' burden for patients.

Category 3: tone and wording. A common opinion was that the wording in the second prototype set, V1.2, felt more 'alive' and 'personal' and therefore more appealing. It was also expressed that this version had a more 'positive' tone, e.g. by using the phrase 'give others opportunity to speak' rather than only 'interrupt'. Interestingly, this was frequently discussed in proximity with the topic of length, with several clinicians stating that V1.2, because of its tone, might be easier for patients to answer despite its higher number of items. The first prototype set, V1.1, was in contrast described as having a 'negative' or in some cases 'clinical' tone which might be seen as invalidating and therefore demotivate patients to monitor symptoms regularly. The similarity with instruments used in clinical context was seen as increasing the risk that patients would consider it too boring to use.

Specific changes suggested. Clinicians also gave several detailed suggestions for changes to the prototypes. Ideas were expressed to merge some items to shorten the length, especially for V1.2, or to split up questions to make them clearer. There were also multiple suggestions

for content to add to create a more comprehensive instrument, e.g., questions about sleep or problems related to parenting. However, some clinicians suggested that the V1.2 items covering routines for basic needs might be better suited in another app function, expressing that this would both shorten the instrument and that the value of such information might differ too much between patients.

Considerations and revisions after first iteration. Following the first iteration, a decision was made to abandon V1.1 since this was the least preferred version and instead make changes to V1.2 to integrate and accommodate clinicians' views and wishes. The revised prototype, V2.1, was shortened from 23 items to 15 to address the issue of response burden while also utilizing clinicians' input on items to merge, e.g., merging items 1 ('keeping in touch with family, friends or other close relatives') and 5 ('living up to expectations from family, friends and other close relatives') into one item ('maintaining relationships') and suggestions to avoid repetitiveness. While doing so, efforts were made to retain the personal and positive tone described by clinicians. Items pertaining to basic need (e.g., 'eating regularly') were removed since these did not clearly align with the intention to monitor ADHD symptoms and, according to some clinicians, might be more suited for a separate app function. The recall period of the instrument was set to one week. Furthermore, a trade-off was made by removing the grouping of items based on areas of life in favour of shortening the instrument, and merge items that addressed the same symptoms, such as items two ('staying focused while socializing [...]') and nine ('staying focused during meeting/lectures [...]').

In addition, two new sets of prototype items for symptom monitoring were created to address different suggestions from clinicians as well as to facilitate further discussions about respondent burden and content validity. One new prototype (V2.2) consisted of only two items, asking the respondent to rate how frequently they had noticed symptoms of inattention and hyperactivity/impulsivity, respectively. The other new prototype (V2.3) was a combination of the V2.1 symptom frequency checklist and a shorter additional instrument inquiring about symptom burden in five different areas of life. This combination constituted an option where symptom frequency and symptom burden could be measured at different timepoints and with different frequencies. The three prototype sets are presented in Table S2 in Supplementary material 1.

Second iteration of interviews

During the second iteration, clinicians were presented with the three prototype versions. The following categories emerged.

Category 1: content validity and clinical relevance. Both V2.1 and V2.3 were in general viewed as being able to provide

clinicians with relevant information of clinical value. Throughout all interviews, clinicians gave examples of clinical duties for which the instruments could provide information of value, e.g., evaluating treatment effects, assessment of treatment needs and treatment planning. It was also commonly stated that results from both versions could constitute interesting material for discussion during patient visits and thus enhance collaboration between clinicians and patients. V2.2, on the other hand, was generally described as 'imprecise' or 'too broad', and it was stated that the amount of follow-up questions required to achieve a full picture would render this type of monitoring redundant. Some clinicians expressed that V2.2 possibly could be useful for evaluating the effect of a new medication but that it would require more frequent monitoring than once a week to detect any clinically relevant patterns.

Category 2: respondent burden

Subcategory: length. All clinicians viewed it as positive that the previously 23-item instrument version had been shortened to 15 items. Some were however still concerned that this length would be too burdening for patients to keep up monitoring.

Subcategory: patients' understanding of ADHD symptoms.

Throughout the interviews, the topic of patients' own understanding of their symptoms and symptom expressions arose. It was pointed out by multiple clinicians that the demand for such an understanding was high for V2.2 in that patients themselves would need to draw conclusions about what behaviours and experiences that constituted symptom expressions and be able to separate between symptoms representing inattention or hyperactivity/impulsivity. Clinicians worried that this would make it too difficult for patients to know how to answer the questions. Similar comments were made about the symptom burden instrument in V2.3 wherein clinicians frequently mentioned that many patients would find it hard to make inferences about how their symptoms affected each of the different areas of life. In contrast, the items in V2.1 were overall described as concrete and 'relatively easy to understand and answer'.

Specific changes suggested. During discussions about how the issue of instrument length could be addressed, various detailed suggestions were made. As in the previous round of interviews, these included ideas to merge items. There was however no clear consensus or majority for what items to merge. For example, some clinicians felt that three items could be adapted and combined into a single item covering relaxation or recovery, while others viewed them as clearly different and important to keep apart. A couple of clinicians suggested to retain V2.1 in its current form but to change the monitoring frequency from once a week to once a month to alleviate responder burden, commenting that it might not be relevant to answer V2.1

more frequently since results possibly would not vary significantly between each week. For the same reason, some clinicians viewed V2.3 as a suitable compromise, suggesting symptom frequency and symptom burden to be monitored once a month and once a week, respectively.

Considerations after second iteration. After the second iteration, an initial decision was made to abandon V2.2, as this approach was the least preferred. Both V2.1 and V2.3 were considered good options from the perspective of informational value, though there were some concerns from clinicians about patients' ability to understand and answer the additional items in V2.3. The sole reason given for preferring V2.3 over V2.1 was lower respondent burden in terms of length. Because of this, a decision was made to temporarily retain V2.1 in its current form and further address this issue in parallel with examination of the psychometric properties of the mCASS.

Third iteration of interviews

For the third iteration of interviews, clinicians were briefed on the choice to retain V2.1 and given the opportunity to comment. Otherwise, discussions were focused on clinicians' preferences regarding how monitoring results should be presented in the app. Results are presented below.

Category 1: visualization. Throughout interviews, clinicians stated that results should preferably be presented in a visual manner that could clarify potential trends or changes in symptoms over time. Line graphs and bar charts were frequently mentioned as a preferred format. Some clinicians expressed that they often experience being pressed for time during meetings with patients and therefore wished for results to be presented in a way that could allow for an initial quick overview to determine if further examination is needed.

Category 2: level of detail. Another aspect that emerged as important to consider when designing the in-app presentation of monitoring results was the level of detail with which results are presented, though views differed some. A reoccurring position was that a total score for the entire mCASS could be enough, again citing time pressure during patient visits as the reason for this. Others expressed that a total score would be too vague, stating that, if possible, subscales or sub-domains would be preferable to receive more nuanced information about specific patients. One common comment was that needs and preferences for this issue would probably differ between professions as well as what type of care the patient was receiving. However, clinicians were, with few exceptions, not interested in item-level results as this would be too detailed and time-consuming to go through.

Category 3: time horizon. The topic of time horizon also arose during discussions, i.e., how long the period covered by an in-app results view should be. Periods mentioned as interesting varied between the past week and up to one year. Nonetheless, it was frequently mentioned that an app function for this needed to be flexible in this regard, since needs would differ a lot between professions, contact formats and treatments.

Quantitative findings

After conducting the interviews, some questions remained and were subsequently addressed in the survey. Among these were the issue of monitoring frequency which was still not completely resolved as clinicians expressed some doubts surrounding the feasibility of their preferences. Additionally, a deliberate choice was made to incorporate the theme of data sensitivity into the survey, a topic that was not brought up during discussions of instrument contents in the interview phase. Nonetheless, we felt this could constitute an important factor for patients' willingness to use the monitoring function and thus should be addressed to ensure a more comprehensive examination.

Results from survey responses from 397 individuals with self-reported ADHD are described. Characteristics of survey participants are presented in Table 3.

To examine views on monitoring frequency using an in-app instrument examining ADHD symptoms, participants were asked to rate the importance of being able to adjust the said frequency to their own individual preferences. Answers were recorded using a VAS ranging from 0 ('Not at all important') to 100 ('Very important'), and importance was generally rated as high ($M=81.2$; $SD=17.5$).

Examining preferred monitoring frequency specifically using the mCASS, 54.2% out of the 397 participants in the final sample stated that they would prefer to monitor symptoms by answering the outcome instrument once per week, while 32% answered that they would prefer a frequency of once a day, and 9.6% would prefer once a month. Only one person (0.3%) stated that once per year would be preferable. In addition, 16 individuals (4%) chose the 'Other frequency' option.

Furthermore, once per week was revealed to be viewed as the appropriate monitoring frequency in general on the individual item level (Figure 2). It was most commonly viewed as the appropriate cadence for all items separately.

To collect information on face validity of the mCASS, participants were also asked to rate to what degree each item captures a difficulty that individuals with ADHD face. Answers were recorded using a VAS ranging from 0 ('Not at all') to 100 ('Completely'). Mean ratings ranged from 53.3 ($SD=24.5$) to 86.0 ($SD=15.6$). Results are shown in Figure 2.

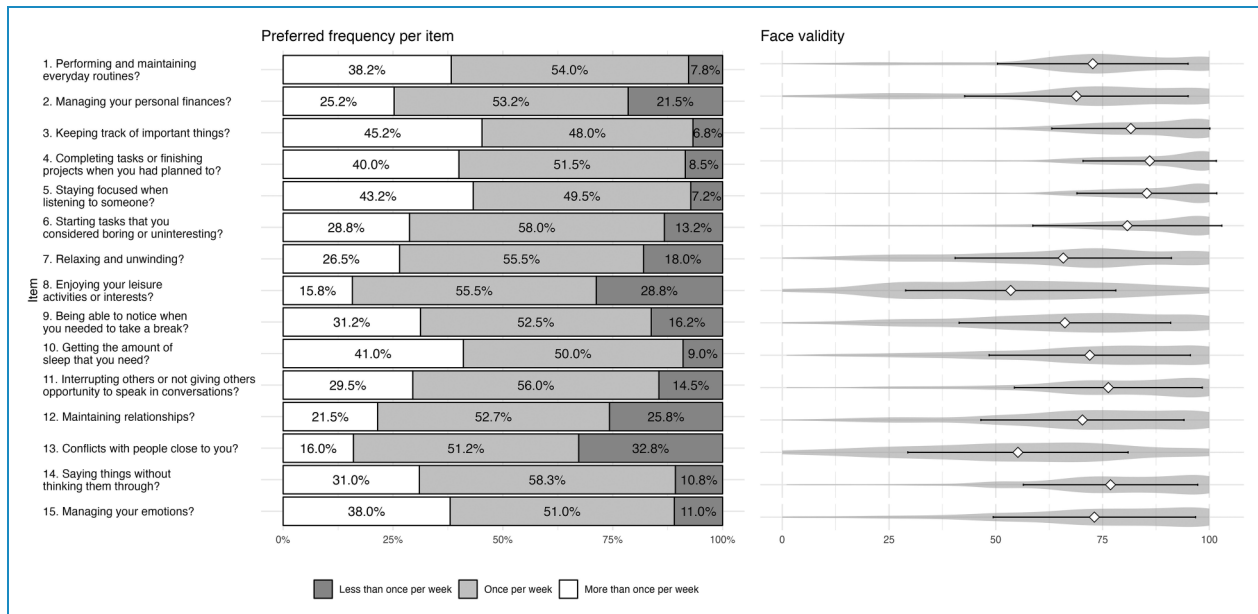


Figure 2. Appropriate monitoring frequency and face validity ratings per item.

Note. $n = 397$. Preferred monitoring frequency per item was explored using the following question: ‘Is once a week an appropriate frequency to ask about this type of experience, despite any variations during the week?’. Response options were ‘No, should ask more frequently’, ‘Yes’ and ‘No, should ask less frequently’. Distribution of answers is displayed in the left-hand grouped bar plot. Face validity ratings were collected using a single question for each item (‘To what extent do you feel that the question above captures a difficulty that individuals with ADHD/ADD face?’). Answers were collected on a VAS ranging from 0 (‘Not at all’) to 100 (‘Completely’). Right-hand violin plots display the results.

Participants were also asked to provide input on to what degree various types of data that could be collected through a smartphone app is perceived as sensitive. Answers were collected on a VAS ranging from 0 (‘Not sensitive at all’) to 100 (‘Very sensitive’). Distributions of ratings for each data type are presented in Figure 3. A repeated measures ANOVA indicated a statistically significant difference between how sensitive the data types were considered to be ($F(396, 3) = 326, p < .001$), with subsequent post hoc tests with Bonferroni adjustments revealing that ADHD symptoms were considered significantly more sensitive than step count, sleep and pulse ($p < .001$).

Psychometric properties of the new instrument. To explore the underlying factor structure of the mCASS, an EFA was performed. The Kaiser–Meyer–Olkin measure of sampling adequacy (.867) as well as the Bartlett test of sphericity (1730 ($df = 105$); $p < .001$) demonstrated that data was suitable for EFA. Parallel analysis suggested a four-factor solution, explaining 44.7% of the variance. However, examination of the factor loadings of the individual item revealed three items with considerable cross-loadings: ‘Getting the amount of sleep that you need?’, ‘Maintaining relationships (e.g. initiating contact with friends and/or family, returning text/calls, participating in social activities)?’ and ‘Managing your emotions?’. These did not fulfil the requirements of the ‘.40-.30-.20’ rule of thumb,

which states that an item is acceptable if loadings onto primary and alternative factors are above 0.40 and less than 0.30, respectively, and if the difference between primary and alternative factors is at least 0.20.³¹ The poor fit of these items remained or worsened when testing other oblique methods for rotation (promax, simplimax). Subsequently, the three items with unsatisfactory factor loadings were removed and the EFA was re-run. Results of the second EFA showed that the four-factor solution remained and now explained 47.4% of the variance. Additionally, model fit measures suggested a better fit (root mean square error of approximation (RMSEA) = .0420; Tucker–Lewis index (TLI) = 0.961) as compared to the 15-item four-factor solution (RMSEA = .0545; TLI = .922).³²

Examination of the factors and included items suggested that the four factors reflected everyday tasks, productivity, rest and recovery and interacting with others, respectively. Next, internal consistency was examined by calculating Cronbach’s alpha. For the full 12-item mCASS, results were satisfactory ($\alpha = .826$). On a subscale level, the rest and recovery factor and the productivity factor had adequate internal consistency ($\alpha = .742$ and $.731$, respectively). The internal consistencies for the everyday tasks and the interacting with others factor were lower ($\alpha = .692$ and $.668$ respectively); however, it should be noted that the equation for Cronbach’s alpha penalizes low number of items and that the commonly used cut-off of >0.7 does not apply to

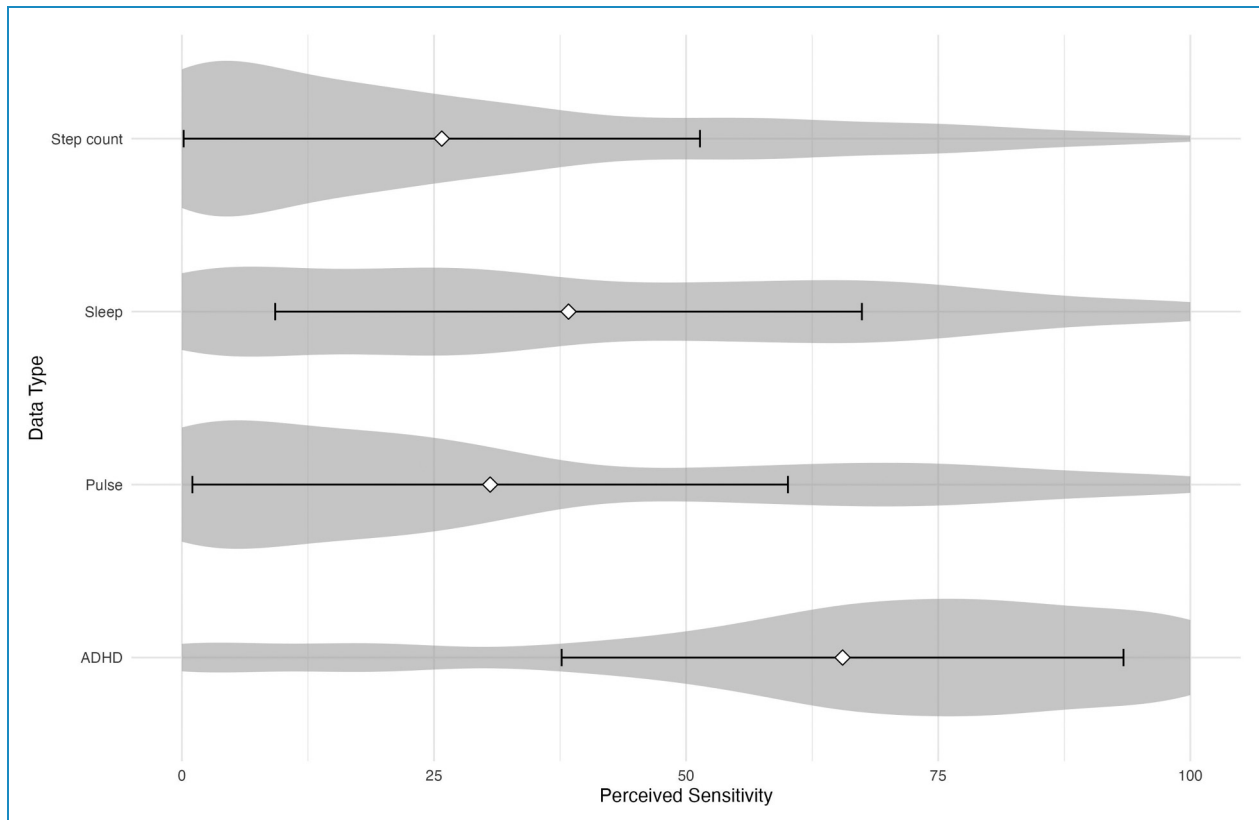


Figure 3. Perceived sensitivity of data types.

Note. Participants rated perceived sensitivity for each data type on a VAS ranging from 0 ('Not at all sensitive') to 100 ('Very sensitive'). Violin plots showing the distributions of ratings are displayed.

subscales. Table 4 shows factor loadings for each item of the mCASS, as well as Cronbach's alpha for each factor.

Concurrent validity of the 12-item mCASS was evaluated by calculating Pearson's r against the ASRS-V1.1 Screener total score. Results from this revealed satisfactory concurrent validity ($r = .595$), suggesting an appropriate overlap in construct measured by the two instruments. The final 12-item version of the mCASS is presented in Supplementary material 3.

Discussion

Principal findings

The current study describes an initiative to scientifically develop and validate the mCASS, a self-reporting instrument for long-term ADHD symptom monitoring, specifically for inclusion in a smartphone app. Insights from the initial qualitative interview phase resulted in a 15-item instrument. It was consequently shortened to 12 items after collection of pilot data from participants with self-reported ADHD ($n = 397$) and subsequent psychometric analyses, which revealed a four-factor solution. Moreover, the mCASS demonstrated satisfactory concurrent validity

and internal consistency, and both qualitative and quantitative results supported face validity for all included items.

In addition to developing and validating a new instrument, the current study also offers insights that may inform similar projects aiming to include app-based symptom measurement for individuals with ADHD. Apart from specific feedback on item content and other instrument properties, qualitative analysis also revealed general aspects relevant for adherence to regular monitoring among ADHD patients and some recommendations for how to address these. A reoccurring topic was that of respondent burden, in which instrument characteristics such as length and monitoring frequency were brought up. Once a week was the frequency most recommended as the appropriate monitoring cadence. However, some clinicians repeatedly worried that this would be too great of a burden for the patient group. Interestingly, this did not seem to be the case when asking the ADHD participants: indeed, survey findings instead confirmed that once a week was a suitable and preferred frequency, with once a day as the second most preferred. These results could suggest a general interest for relatively frequent symptom monitoring among individuals with ADHD, somewhat in contrast with both the current study's findings from interviews with ADHD

Table 4. Exploratory factor analysis of the 12-item mCASS.

Item	Factor Loading			
	1	2	3	4
Factor 1: Everyday tasks				
1. Performing and maintaining everyday routines (e.g., doing laundry, cleaning, showering, brushing your teeth)?	.616	.076	.055	-.049
2. Managing your personal finances (e.g., paying bills, sticking to a budget, saving money)?	.647	.001	.018	.079
3. Keeping track of important things (e.g., phone, keys, other belongings)?	.499	.083	.044	.157
Factor 2: Productivity				
4. Completing tasks or finishing projects when you had planned to (e.g., at home or at work/in your studies)?	.184	.685	.021	-.083
5. Staying focused when listening to someone (e.g., in conversations, meetings or lectures)?	-.079	.596	.033	.234
6. Starting tasks that you considered boring or uninteresting?	-.023	.692	.028	-.027
Factor 3: Rest and recovery				
7. Relaxing and unwinding?	-.020	.044	.635	.116
8. Enjoying your leisure activities or interests?	.030	-.038	.822	-.039
9. Being able to notice when you needed to take a break?	.005	.096	.587	-.032
Factor 4: Interacting with others				
10. Interrupting others or not giving others opportunity to speak in conversations?	-.107	.113	.021	.587
11. Conflicts with people close to you (e.g., partner, family members or friends)?	.006	-.051	.209	.498
12. Saying things without thinking them through?	.149	-.013	-.026	.726
Cronbach's alpha	.692	.731	.742	.668

Note. Minimum residual extraction method was used in combination with an oblimin rotation. Factor loadings above .40 are shown in bold.

clinicians and the standard procedure of yearly follow-ups in routine care. This apparent discrepancy is important to note and raises questions about whether more frequent and regular monitoring of symptoms could positively impact clinical outcomes or health literacy in this patient group, similar to findings in other patient populations.^{6,7} The current literature on this topic is sparse, highlighting a need for future studies to explore these potential benefits. These investigations would be facilitated by the availability of app-based monitoring instruments, such as the mCASS.

Previous studies examining the views of clinicians working in psychiatry regarding implementation of routines or tools for regular monitoring have found that while clinicians also feel positively about this, there are some

identified barriers. Among such is perceived lack of time, which might negatively impact clinicians' willingness or possibilities to follow up on monitoring results.^{33,34} The topic of time pressure was also brought up in the current study, especially when discussing the design of in-app presentation of monitoring results. For this topic, clinicians frequently motivated presentation preferences based on time efficiency. It is not unlikely that clinicians could have assumed that more frequent monitoring would also entail equally frequent clinician review of results. As such, it may be difficult to distinguish between clinicians' knowledge about patient needs and their own preferences affected by their work situation, i.e., viewing more frequent monitoring as inconvenient because of time limitations.

Careful attention to this distinction should be paid in future research.

Moreover, clinician discussions about respondent burden extended beyond instrument length and monitoring frequency to also include adaptations to patient and individual characteristics, such as cognitive function and understanding of ADHD symptoms, as necessary to reduce respondent burden. For example, clinicians recommended considering impairments in working memory and sustained attention when creating instrument items. Suggestions for how to address this included incorporating examples for each item and not using too abstract symptom descriptions. As individuals with ADHD often struggle to follow instructions as well as to remember and complete tasks, especially tasks that are demanding,³⁵ there is cause to believe that this population might face greater barriers to long-term monitoring adherence. For this reason, translation of ADHD symptom expressions into practical design considerations could be of great importance. Some previous research has explored this area and proposed certain guidelines for development of software and assistive technologies^{36,37}; however, the literature is still scarce.

Results from the current study also revealed that wording and tone of the instrument, as well as included items, were believed to potentially affect patients' willingness to use and engage with the monitoring functionality. Using a personal tone and phrasing items to make them feel more alive and closer to reality was viewed as important for this patient group. This was contrasted against what was described as a more clinical tone found in many existing instruments used in routine care. The reasoning behind this advice from clinicians involved a concern that patients would otherwise find the task of self-monitoring far too boring to keep it up over a longer time period. In line with this, there does indeed seem to exist a positive correlation between ADHD symptoms and boredom proneness,³⁸ and previous research on continued use of mHealth apps has also mentioned boredom as a factor causing users to abandon them, though mainly in the context of limited app features.³⁹ Moreover, participating clinicians also indicated that the tone of the instrument could be more important for patient adherence than length. This aligns with previous literature suggesting that instrument length alone is not a sufficient predictor of perceived respondent burden; rather, it is also influenced by perceptions of the instrument, interest and motivation.^{40,41}

Regarding integrity as a potential barrier to using app-based symptom monitoring, previous research has suggested that users of mHealth apps do have concerns about security and privacy⁴² and that such concerns are rated higher when data is considered more sensitive.⁴³ Results from the current study confirm that data on ADHD symptoms is indeed regarded as significantly more sensitive than data on step count, pulse and sleep – three types of health data that are often collected by smartphones and

wearables. Although it was beyond the scope of the current study to examine the reasons for this discrepancy, it is likely driven by the psychological and emotional nature of symptoms as compared to more objective physiological data or by concerns that information on ADHD symptoms could have an economic impact if it reached, e.g. employers or insurance companies. Nonetheless, the results are in line with recent work that has found that sensitivity of information is perceived to be higher for medical diagnoses and mental health as compared to other health information and lifestyle factors.⁴⁴ These concerns must be addressed when developing mHealth apps for symptom monitoring and especially so within the field of mental health. Despite indications that privacy concerns among users may have little effect on actual data-sharing behaviours,^{45,46} there are clear ethical reasons to help users protect data they consider private. In their study, Zhou et al.⁴² suggest several functionalities that users reported would lower their privacy concerns, such as encryption, user authentication and comprehensible privacy policies.

At the outset of the current project, we examined previous literature for examples of instruments suitable for long-term app-based monitoring of ADHD symptoms. We found that studies within the field of EMA have included instruments adapted and used for high-intensity, short-duration monitoring protocols – a stark contrast to the yearly follow-ups conducted in routine care for ADHD. Results from the conducted interviews and survey yielded a new monitoring instrument, the mCASS. This instrument strikes a balance between these two approaches by enabling more frequent assessments than the typical yearly follow-ups while also aiming to accommodate longer monitoring durations than conventional EMA protocols.

Strengths and limitations

To our knowledge, the mixed-methods study described herein is the first to take a comprehensive approach to both develop and validate an instrument specifically intended for app-based long-term monitoring of ADHD symptoms. Strengths of this study include the use of both groups of target users as informants in the development and validation process described, including both experienced ADHD clinicians (a third of which had over 10 years of experience) and a relatively large sample of individuals with self-reported ADHD ($n=397$). Importantly, including both stakeholders, at different development stages, minimizes the risk of sampling bias that could shape design considerations negatively. Importantly, there is some indication in the literature that individuals differ in how frequently they use mHealth apps based on, e.g., demographic variables and individual skills.⁴⁷ mHealth development efforts must therefore strive to consider the broader target group when making design decisions.

An obvious limitation of the current study is that the newly developed and validated mCASS has not yet been deployed in the intended format (a smartphone app); hence, feasibility estimates, including actual adherence rates, are not yet available.

Other limitations include the final interview sample consisting of women only. Only two individuals out of the 24 interested in the study were men, meaning that sex could not be used as a variable for stratified sampling. This likely affected the width of experiences uncovered in the interviews. Likewise, few psychiatrists were recruited to take part in this study; the final sample included only one. Efforts were made to recruit more psychiatrists, but this proved challenging, likely due to the general difficulty psychiatrists face in making time during their workdays. While psychiatrists play a crucial role in the care of ADHD patients, such as prescribing medication, it is important to note that much of the ongoing care is provided by other healthcare professionals represented in this study. This includes follow-ups on medication effects and potential side effects, as well as non-pharmacological therapeutic interventions. As to transferability of interview findings, the clinicians were recruited from Swedish psychiatric care and some interview findings could therefore be specific to this context. Furthermore, qualitative analysis was conducted only by the first author, which might affect the credibility of the results. It should however be noted that the iterative nature of the interviews constituted an internal process of validation since analysis results were brought back for clinicians to react to and correct if necessary. For the quantitative survey phase, data collection entailed a low degree of control regarding the conscientiousness with which participants provided their answers. These issues can occur in most survey-based procedures, and previous research suggests that data collected through the Prolific.co platform is of good quality.⁴⁸ Relatedly, there was no way to confirm if participants were actually diagnosed with ADHD. However, since the final app, in which the mCASS will be included, will be released to the public and not require any verification of diagnosis, such an inclusion criterion could have rendered the findings from this preliminary validation less representative of the larger, intended user group. Moreover, even though results indicated satisfactory internal consistency and concurrent validity of the mCASS, it is important to note that since an initial translation was required already at the development stage for data collection reasons, the standard procedure for instrument translation⁴⁹ was not followed in this case. This limitation is also connected to the risk of losing aspects specifically related to the tone and wording of items, one of the points raised by clinicians during the development phase. While the authors involved in the translation process were aware of this risk and made every effort to bring all recommendations by clinicians into the English version, no detailed conclusions can be drawn regarding

how tone and wording of items are perceived by individuals with ADHD, regardless of translation used. Importantly, the preliminary psychometric findings presented herein suggest that items are interpreted and function as intended.

Conclusions and future directions

The current study describes the development and validation of the mCASS, a novel self-rating instrument for long-term app-based monitoring of ADHD symptoms. Findings revealed that the instrument, informed by experiences of both clinicians and individuals with ADHD, appears to be valid for assessing ADHD symptoms in the intended format. Future studies should reassess the psychometric properties of the mCASS using a clinical sample of individuals with verified ADHD diagnosis, as well as examine measurement invariance of the mCASS across variables such as sex or pharmacological treatment status.

Moreover, this study gathered important insights into what to consider when designing mHealth instruments for individuals with ADHD. Clinicians provided information about factors pertaining both to the clinical care context and to patient characteristics, which were all viewed as important from the perspective of willingness to use and adherence to an mHealth symptom monitoring app. These insights can be valuable for future, similar, mHealth initiatives. Future studies concerned with increasing use of mHealth apps should consider all parts of the app, including any instruments it contains. Specifically, regarding the instrument described herein, future steps should involve generation of real-world evidence on feasibility and adherence to an app utilizing the mCASS for long-term monitoring of ADHD symptoms.

Acknowledgements: The authors would like to thank all participating clinicians for taking from their time to contribute to this study during interviews. They would also like to thank the other members of the Digital Psychiatry Lab at the Centre for Psychiatry Research, Karolinska Institutet, for their valuable input throughout the process of conducting this study.

Contributorship: AB: methodology, validation, formal analysis, data curation, writing – original draft, visualization, project administration; DF: methodology, writing – review and editing, supervision; LH: conceptualization, project administration, funding acquisition; MJ: writing – review and editing, supervision; IH: methodology, writing – review and editing; PL: conceptualization, methodology, formal analysis, resources, writing – original draft, supervision, funding acquisition. All authors reviewed and edited the manuscript and approved the final version of the manuscript.

Declaration of conflicting interests: The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: LH is an

employee of Takeda Pharmaceutical, which funded the study. The other authors have no conflicts of interest to declare.

Ethical approval: As the qualitative interviews collected no sensitive personal data (or involved any intervention), this part fell outside the scope of the Swedish Ethical Review Act. However, the Swedish Ethical Review Authority issued an advisory statement that they saw no ethical issues in the study (REC number: 2022-01690-01). Ethical approval for the quantitative survey study was issued separately (REC number: 2023-03313-01).

Funding: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by Takeda Pharmaceutical and governed by a signed industry–academia collaboration that guaranteed full academic freedom (registration number: SLSO 2022-0517). The funder had no role in the choice of study design; the collection, analysis, or interpretation of the data; writing the manuscript; or the decision to submit the paper for publication.

Guarantor: PL.

ORCID iD: Amanda Bäcker  <https://orcid.org/0000-0002-9328-9805>

Supplemental material: Supplemental material for this article is available online.

References

1. Sonuga-Barke EJS, Becker SP, Bölte S, et al. Annual research review: perspectives on progress in ADHD science – from characterization to cause. *J Child Psychol Psychiatry* 2023; 64: 506–532.
2. Ahnemark E, Schiena D, Fredman M, et al. Health-related quality of life and burden of illness in adults with newly diagnosed attention-deficit/hyperactivity disorder in Sweden. *BMC Psychiatry* 2018; 18: 1–11.
3. Asherson P, Leaver L, Adamou M, et al. Mainstreaming adult ADHD into primary care in the UK: guidance, practice, and best practice recommendations. *BMC Psychiatry* 2022; 22: 640.
4. Pedersen SL, Kennedy TM, Joseph HM, et al. Real-world changes in adolescents' ADHD symptoms throughout the day and across school and non-school days. *J Abnorm Child Psychol* 2020; 48: 1543.
5. Schmier J and Halpern MT. Patient recall and recall bias of health state and health status. *Expert Rev Pharmacoecon Outcomes Res* 2004; 4: 159–163.
6. Kauer SD, Reid SC, Croke AHD, et al. Self-monitoring using mobile phones in the early stages of adolescent depression: randomized controlled trial. *J Med Internet Res* 2012; 14: e67.
7. Wichers M, Simons CJP, Kramer IMA, et al. Momentary assessment technology as a tool to help patients with depression help themselves. *Acta Psychiatr Scand* 2011; 124: 262–272.
8. Păsărelu CR, Andersson G and Dobrea A. Attention-deficit/hyperactivity disorder mobile apps: a systematic review. *Int J Med Inform* 2020; 138: 104133.
9. Grundy Q and Bloomberg LS. A review of the quality and impact of mobile health apps. *Annu Rev Public Health* 2022; 43: 117–134.
10. Torous J, Bucci S, Bell IH, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. *World Psychiatry* 2021; 20: 318–335.
11. Stinson L, Liu Y and Dallery J. Ecological momentary assessment: a systematic review of validity research. *Perspect Behav Sci* 2022; 45: 469–493.
12. Shiffman S, Stone AA and Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol* 2008; 4: 1–32.
13. Wen CKF, Schneider S, Stone AA, et al. Compliance with mobile ecological momentary assessment protocols in children and adolescents: a systematic review and meta-analysis. *J Med Internet Res* 2017; 19: e132.
14. Williams MT, Lewthwaite H, Fraysse F, et al. Compliance with mobile ecological momentary assessment of self-reported health-related behaviors and psychological constructs in adults. *Systematic Review and Meta-Analysis. J Med Internet Res* 2021; 23: e17023.
15. Miguelez-Fernandez C, de Leon SJ, Baltasar-Tello I, et al. Evaluating attention-deficit/hyperactivity disorder using ecological momentary assessment: a systematic review. *ADHD Attention Deficit and Hyperactivity Disorders* 2018; 10: 247–265.
16. Simons L, Valentine AZ, Falconer CJ, et al. Developing mHealth remote monitoring technology for attention deficit hyperactivity disorder: a qualitative study eliciting user priorities and needs. *JMIR Mhealth Uhealth* 2016; 4: e31.
17. Denyer H, Deng Q, Adanijo A, et al. Barriers to and facilitators of using remote measurement technology in the long-term monitoring of individuals with ADHD: interview study. *JMIR Form Res* 2023; 7: e44126.
18. Solstad SM, Castonguay LG and Moltu C. Patients' experiences with routine outcome monitoring and clinical feedback systems: a systematic review and synthesis of qualitative empirical literature. *Psychother Res* 2019; 29: 157–170.
19. An Q, Kelley MM, Hanners A, et al. Sustainable development for mobile health apps using the human-centered design process. *JMIR Form Res* 2023; 7: e45694.
20. Robinson OC. Sampling in interview-based qualitative research: a theoretical and practical guide. *Qual Res Psychol* 2014; 11: 25–41.
21. Ferguson C, Inglis SC, Breen PP, et al. Clinician perspectives on the design and application of wearable cardiac technologies for older adults: qualitative study. *JMIR Aging* 2020; 3: e17299.
22. Geryk LL, Roberts CA, Sage AJ, et al. Parent and clinician preferences for an asthma app to promote adolescent self-management: a formative study. *JMIR Res Protoc* 2016; 5: e229.
23. Graneheim UH and Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures

- to achieve trustworthiness. *Nurse Educ Today* 2004; 24: 105–112.
24. Palan S and Schitter C. Prolific.ac – A subject pool for online experiments. *J Behav Exp Finance* 2018; 17: 22–27.
 25. Stanton K, Carpenter RW, Nance M, et al. A multisample demonstration of using the Prolific platform for repeated assessment and psychometric substance use research. *Exp Clin Psychopharmacol* 2022; 30: 432–443.
 26. Kessler RC, Ames L, Ames M, et al. The World Health Organization adult ADHD self-report scale (ASRS): a short screening scale for use in the general population. *Psychol Med* 2005; 35: 245–256.
 27. Curran PG. Methods for the detection of carelessly invalid responses in survey data ★. *J Exp Soc Psychol* 2016; 66: 4–19.
 28. Dunn AM, Heggestad ED, Shanock LR, et al. Intra-individual response variability as an indicator of insufficient effort responding: comparison to other indicators and relationships with individual differences. *J Bus Psychol* 2018; 33: 105–121.
 29. Costello AB and Osborne J. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Res Eval Pract Assess Res Eval* 2005; 10: 7.
 30. Gaskin CJ and Happell B. On exploratory factor analysis: a review of recent evidence, an assessment of current practice, and recommendations for future use. *Int J Nurs Stud* 2014; 51: 511–521.
 31. Howard MC. A review of exploratory factor analysis decisions and overview of current practices: what we are doing and how can we improve? *Int J Hum Comput Interact* 2016; 32: 51–62.
 32. Montoya AK and Edwards MC. The poor fit of model fit for selecting number of factors in exploratory factor analysis for scale evaluation. *Educ Psychol Meas* 2021; 81: 413–440.
 33. Gelkopf M, Mazor Y and Roe D. A systematic review of patient-reported outcome measurement (PROM) and provider assessment in mental health: goals, implementation, setting, measurement characteristics and barriers. *Int J Qual Health Care* 2022; 34: ii13–ii27.
 34. Van Wert MJ, Malik M, Memel B, et al. Provider perceived barriers and facilitators to integrating routine outcome monitoring into practice in an urban community psychiatry clinic: a mixed-methods quality improvement project. *J Eval Clin Pract* 2021; 27: 767–775.
 35. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. 5th ed. Washington, DC: American Psychiatric Association, 2013.
 36. Sonne T, Marshall P, Obel C, et al. An assistive technology design framework for ADHD. In: *Proceedings of the 28th Australian Conference on Computer-Human Interaction - OzCHI '16*, 29 November–2 December 2016, pp.60–71. New York, New York, USA: ACM Press.
 37. McKnight L. Designing for ADHD in search of guidelines. *IDC 2010 Digital Technologies and Marginalized Youth Workshop* 2010; 30.
 38. Malkovsky E, Merrifield C, Goldberg Y, et al. Exploring the relationship between boredom and sustained attention. *Exp Brain Res* 2012; 221: 59–67.
 39. Woldeyohannes HO and Ngwenyama OK. Factors influencing acceptance and continued use of mHealth apps. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 2017; 10293 LNCS: 239–256.
 40. Rolstad S, Adler J and Rydén A. PATIENT-REPORTED OUTCOMES response burden and questionnaire length: is shorter better? A review and meta-analysis. *Value Health* 2011; 14: 1101–1108.
 41. Yan T, Fricker S and Tsai S. Response burden: What is it and what predicts it?. In: *Advances in questionnaire design, development, evaluation and testing*. John Wiley & Sons, 2019, pp.193–212.
 42. Zhou L, Bao J, Watzlaf V, et al. Barriers to and facilitators of the use of mobile health apps from a security perspective: mixed-methods study. *JMIR Mhealth Uhealth* 2019; 7: e11223.
 43. Bansal G, Zahedi FM and Gefen D. The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. *Decis Support Syst* 2010; 49: 138–150.
 44. Belen-Saglam R, Nurse JRC and Hodges D. An investigation into the sensitivity of personal information and implications for disclosure: a UK perspective. *Front Comput Sci* 2022; 4: 908245.
 45. Barth S and de Jong MDT. The privacy paradox – investigating discrepancies between expressed privacy concerns and actual online behavior – A systematic literature review. *Telemat Inform* 2017; 34: 1038–1058.
 46. Norberg PA, Horne DR and Horne DA. The privacy paradox: personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs* 2007; 41: 100–126.
 47. Bol N, Helberger N and Weert JCM. Differences in mobile health app use: a source of new digital inequalities? *Inf Soc* 2018; 34: 183–193.
 48. Peer E, Rothschild D, Gordon A, et al. Data quality of platforms and panels for online behavioral research. *Behav Res Methods* 2022; 54: 1643–1662.
 49. Sousa VD and Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. *J Eval Clin Pract* 2011; 17: 268–274.
-