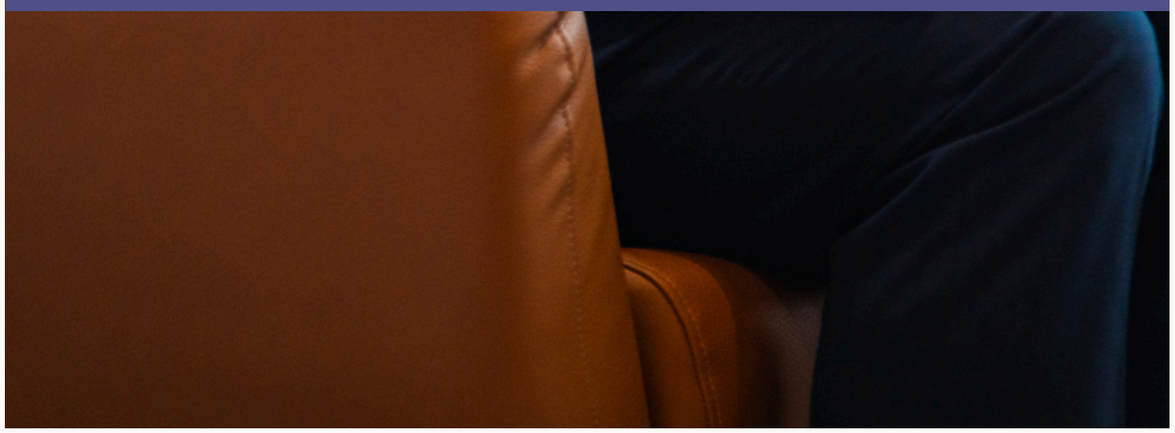
brain in hand

Re-imagining a SCED design for evaluating the impact of BiH in neurodivergent people

November 2025

Background



In 2023 the research team at Brain in Hand (BiH) worked with colleagues at the University of Exeter on developing and implementing a Single Case Experimental Design (SCED) to understand the impact of BiH for a small but diverse cohort of autistic people. The hope was to gather robust and quantitative data on how BiH was having an impact on the lives of people using it.

The aim of a SCED is to determine whether a cause-effect relationship exists, for example, whether there is a functional relationship between using BiH and positive impact measures such as reduced overwhelm. These experimental designs (often referred to as N-of-1 trials in medical settings) test effects using a small number of people using repeated measurements in the absence and presence of an intervention (Epstein and Dallery, 2022). These repeated measures allow participants to serve as their own controls. SCEDs have been used for many years in the field of education and psychology and have recently seen a resurgence given the availability of quality assessment tools and reporting guidelines, and methods of data analysis suitable for SCED data (Krasny-Pacini and Evans, 2018). Given the expense, ethical concerns of withholding support and unpredictability of the utility of randomised controlled trials (RCTs), we wanted to explore a more innovative approach to understanding a cause and effect relationship. Other researchers are also exploring its potential in relation to digital health (eg. Barteles et al., 2022) or for neurodivergent people (Birdsey and Walz, 2021), though we are not aware of any previous attempts to explore a SCED for neurodivergent people seeking digital support.

Authors:
Helen Guyatt and Abi Jones

Citation: Guyatt, H and Jones, A. (2025). Re-imagining a SCED design for evaluating the impact of BiH in neurodivergent people. Brain in Hand Ltd., November 2025

Why do we want to try something different?

"For some time, randomised control trials (RCTs) have been considered the gold standard of research. Although they are undeniably powerful when used correctly, recent thinking suggests that there is perhaps an assumption that any RCT is inherently useful by virtue of its format, when in fact inadequate planning and reporting of RCT studies are contributing to avoidable wasted research." (Guyatt, 2023)

The design followed standard SCED methodologies and was designed and implemented by the researchers at the University of Exeter, highly experienced in this type of study (see Box 1). As a research team at BiH we were really excited at the potential of this study design as we had reported previously about the issues with RCTs (Guyatt, 2023) and had shared in a recent interview with BATA about the SCED study that we were about to explore (BATA)

The research proposal underwent a robust ethics review (University of Exeter Ethics Application ID: 3480639) and received approval to start from October 2023 to January 2024. One of the main concerns in a SCED is often participation rates but both retention and participation with the study was high with 17 of the 19 people who started the study completing it, with 12 people having a full dataset. Instead, an unexpected issue arose in the selection of the metrics whereby baseline scores were already so high that little change was possible. This ceiling effect was due to both the limitation of a 5 point scale and metrics chosen (see Box 2). The problems with the metrics meant that outcomes were patchy and the data difficult to interpret to such an extent that the SCED design utility was deemed invalid. Although these findings were disappointing given the time and resources invested by both the participants and the research teams, we still believed in the value of a SCED and wanted to understand how we could redesign the approach to collect outcomes in a way that would be more appropriate to both our type of support and our users.

This paper reports on what we discovered in this re-imagining of a SCED design as a tool to better understand the impact of BiH for our users.

Box 1. What did our SCED look like – the basics

A randomised, multiple baseline, single-case experimental design (SCED) was adopted over a 12-week period. Participants completed repeated outcome measures during a baseline phase (phase A, minimum of six days) and an intervention phase (phase B, minimum of 50 days). The transition phase was five days. Participants were randomised to one of five tracks which determined the length of their baseline and intervention phases (see Figure 1). The repeated daily outcome measures were collected five days a week.

Participants were asked to respond to three statements, five times a week via a text message. The time of day that participants received their texts was personalised to be between 4pm and 9pm depending on individual's schedules. The first two questions were "Today I worked on the things that are important to me", and "I was content with the way I coped with challenges today". Participants responded to these using a five-point scale. The third statement asked participants to rank on a five-point scale whether their day was overwhelming (one) to relaxing (five).

Annex 1 illustrates what the questionnaires looked like for the participants on their mobile devices.

	Week 1					Week 2					Week 3					Weeks 4-12								
Track	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	21	26	31	36	41	46	51	56
1	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
2	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
3	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
4	A	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
5	A	A	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B	B	B	B	B
						TRANSITION WEEK																		

A - Baseline period
B - Intervention period
A - Baseline period during transition phase
B - Intervention period during transition phase

Figure 1. SCED study design

Initial Analytic Approach

The analysis of this outcome data comprises a visual analysis with graphs showing the individual data points, the median scores during intervention in comparison to baseline and in the case of Brain in Hand timings of the specialist (coaching) sessions. Visual analysis was carried out using the SCDA Shiny application. Central tendency (involving median scores for each phase) and trend (involving the split middle technique) graphs were created for each participant. An example is shown below for participant BiH03 for Overwhelm: “Today was an overwhelming day (1) to relaxed (5) (see Figure 2). Each circle is a data point and the horizontal line represents the median at baseline (here up to 10 days) and then with intervention (10 days through to 60 days). The medians for BiH03 were two in the baseline phase and three in the intervention phase.

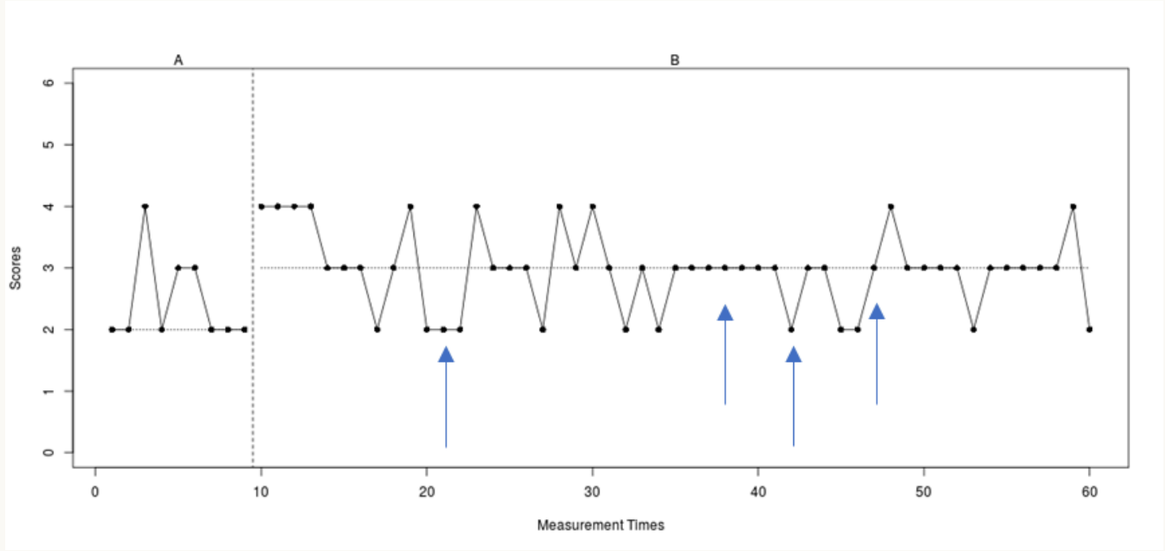


Figure 2. Responses for BiH03 to Overwhelm: “Today was an overwhelming day (1) to relaxed (5)

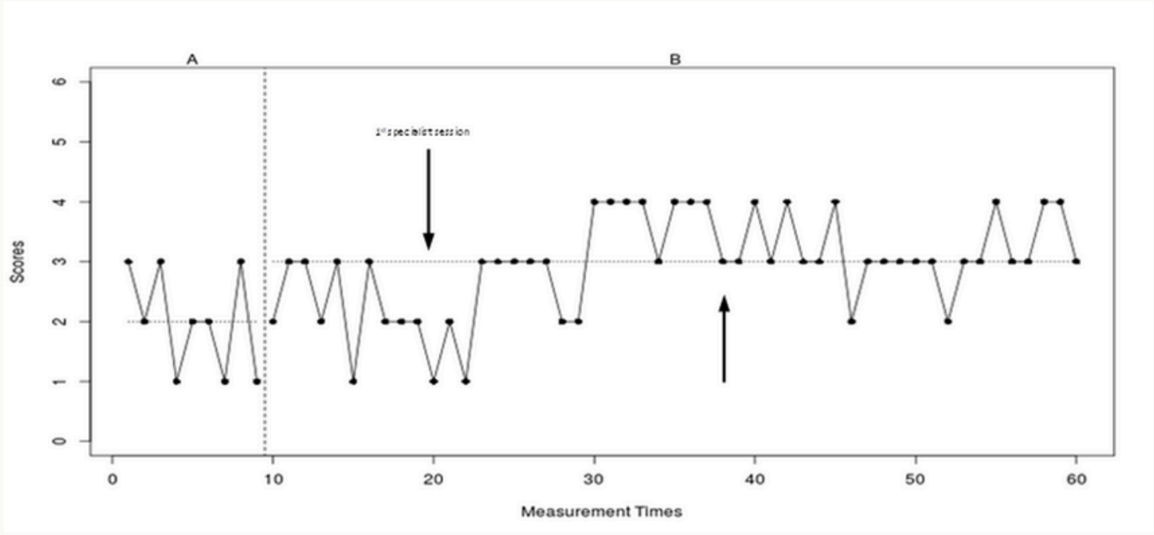
The visual analysis was supplemented with statistical analyses. Non-overlap of pairs (NAP) were calculated for each individual using an online calculator (for example - https://singlecaseresearch.org/calculators/nap/). NAP refers to the proportion of data points during the intervention phase that do not overlap with those collected during the baseline phase. The statistical software R (v4.2.2; R Core Team 2022) was used to perform randomisation tests. This involved Monte Carlo simulation with 1000 iterations to assess significant differences in daily outcome measures between phases for each of the 12 participants. The p-value threshold to determine a significant difference was set at .20. The threshold was calculated by dividing 1 by the number of days in the transition phase (i.e., 1/5 = .20; Bulté and Onghena, 2008). Using the same example as above – BiH03 and overwhelm – the mean score (and SD) during the 50 days intervention (2.96 (0.67)) was significantly different from the 10 days at baseline (2.57 (0.79) according to the assumptions above (NAP 0.6531 and p=0.188).

Box 2. The ceiling effect

Most of the 12 participants reported outcome baseline scores in the higher tier of the 1 to 5 scale. A deep dive into the data found that only one person (BiH05) did not have scores of 4 or 5 at baseline for the goals and coping metrics, and three people for the overwhelm metric (BiH07, BiH11 and BiH15).

The example of BH05 in the Coping - *"I was content with the way I coped with challenges today"* on a scale of 1 to 5 is shown below (significant change and moderate effect size; Figure 3 (a)). This contrasts markedly with an example for BiH03 (Figure 3 (b)) where there were high scores at baseline demonstrating a ceiling effect (small effect size).

a)



b)

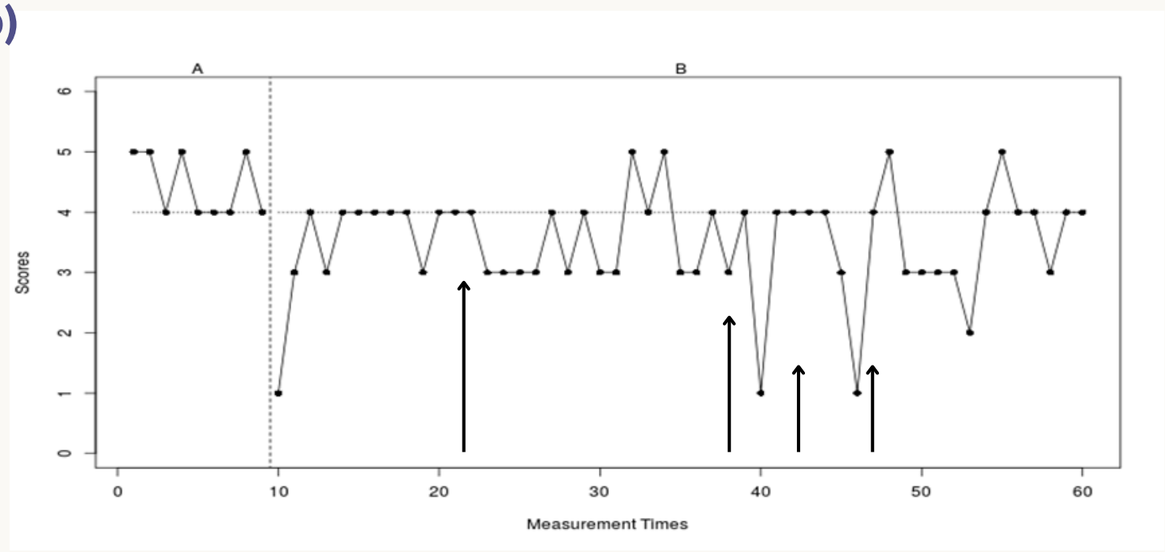


Figure 3. Responses to Coping – "I was content with the way I coped with challenges today for (a) BiH05 and (b) BiH03

Approach

Our main approach was to go back to the participants to ask them some questions about how we could have designed the SCED differently. The questions we asked were directed by three practical ways suggested by the researchers on how the tools or data collection could be changed to enable quality data: the scale of the daily measures, the measures themselves and the duration of the study.

The main details of the data collection tool are presented in Box 3.

The first SCED design issue was the scale for daily measures which had just five different ordinal options.

> *"One of the key limitations was the use of daily outcome measures that were restricted to 5 data points (i.e., a scale of 1 to 5). The scale was implemented based on feedback from the experts by experience who did not want a longer scale, however, this is likely to have impacted the SCED effect sizes... The use of a more detailed scale e.g., a less restrictive sliding scale or 10-point Likert scale would increase the likelihood of identifying any changes taking place."*
>
> Exeter University researchers

In order to explore alternative scales, we presented the participants with six different scaling options (see Table 1b, Box 3): the 5 point scale with wording, a 10 point scale with wording and one with numbers, a rating scale and a sliding scale. For each of these we also asked if there is anything that could be changed to make the scale "perfect". We also asked them to rank the six scales in terms of preference and asked for their preference on agreeing or disagreeing with a statement, or finishing the start of a sentence.

Approach (continued)

The second SCED design issue was the daily outcome measures. In the original design, 3 outcomes were employed, but as noted by the researchers most of these started off with high scores suggesting that these were probably not the most meaningful outcomes to capture change for participants.

In order to explore other more appropriate outcomes we went back to the participants and asked them what they thought about the original outcomes, their ideas for any others and for their reactions to a possible list of outcomes.

We provided these based on a range of outcomes related to executive functioning (for example, 'Today I felt organised' or 'Today I was able to problem solve/make decisions') and wellbeing (for example, 'Today I was able to manage anxiety' or 'Today I looked after myself') (see Table 1a, Box 3).

The third SCED design issue related to the duration of the study as the researchers noted that many participants were still just getting set up by the end of the Phase B (intervention phase).

> *"Many participants started with high scores on the daily outcome scales and therefore this likely contributed to a ceiling effect (particularly for the goals daily outcome). ... It would also be helpful for future studies to consider the daily outcome measures used. For example, overwhelm was reported by all participants and this appeared to be a meaningful concept for them which also showed change in most participants. In contrast the goals question showed a ceiling effect for many participants and perhaps did not adequately tap into the aspects of everyday life that benefit from the use of BiH."*
>
> Exeter University researchers

> *"It would be of value to undertake a longer evaluation, beyond the set-up phase of BiH... The start of the intervention phase of the SCED was marked by the users being given access to BiH. However, due to various external factors that could not be controlled for there was variability in terms of how long users waited for their specialist session and how much they interacted with BiH before this session, making it more difficult to evaluate the impact of individuals using BiH. It is also apparent that users are still in the set-up phase within the period of this study, and many users were still meeting the specialist in the latter weeks of this study."*
>
> Exeter University researchers

In order to explore the possibility of extending the study duration we asked participants how often they would be willing to submit outcome data for 6 different study durations (4 months, 5 months, 6 months, 7 months, 8 months and 9 months) and whether they thought each of these time durations would be "long enough to see change" (see Table 1c, Box 3)).

In addition, as part of the original SCED study, the participants had already fed back insights on the study design during their exit interviews with the researchers where they were explicitly asked about the outcome measures. Some of their feedback is shared in the results, but it was also noteworthy that many noted how they appreciated that they could select both what time of the day they received the text messages on outcomes, but also which 5 days of the week worked best for them. This was an important innovative design by the researchers which really worked for our user group.

> *"It was handy that we could choose what time as well what time we could have it because then I had it. I think mine were like 7:00 or 8:00 o'clock when I've like, I'm on my own I've sat down for the evening."*

> *"The fact that they sort of were chosen to come sort of in the evening knowing that my daughter had sort of be in bed and my day was sort of finished and that kind of thing. It was, yeah, it was a nice check-in."*

We also had follow-up conversations with one of the researchers around the three design issues and these are presented in the results alongside the findings from the survey with participants.

Box 3. The data collection tool

The data collection tool was sent out via survey monkey. The research team invested time in developing a platform that was easy to navigate and understand to facilitate quality data collection and a good participant experience. This included visuals instead of text where possible.

Table 1a, 1b, 1c document the questions asked, and Annex 2 shows screenshots on the format and illustrations for the main areas of interest (a) Section 2. Daily Outcome measures. (b) Section 3. The scaling preferences and (c) Section 4. The duration of the study.

Context before question	Question	Answer type
The following outcome measures were sent to you daily via a link in a text for you to rate: 1. Today I worked on the things that are important to me 2. Today I used strategies that are in my Brain in Hand even if I didn't look at it 3. I was content with the way I coped with challenges today 4. Today was a (1) overwhelming to (5) relaxing day	Do you think these were the best things to be measuring?	Open text
	Would there be more relevant daily outcome measures, from your perspective?	Open text
	Now that you have had some time using Brain in Hand, do you think any of the following would be useful to track in the daily outcome measures?	Closed - multiple choice • Today I felt organised • Today I felt that I was able to be flexible in my tasks and any demands • Today I was able to manage overwhelm • Today I looked after myself • Today I was able to problem solve/make decisions • Today I was able to manage stress • Today I was able to manage anxiety • Other (please specify) • None of the above

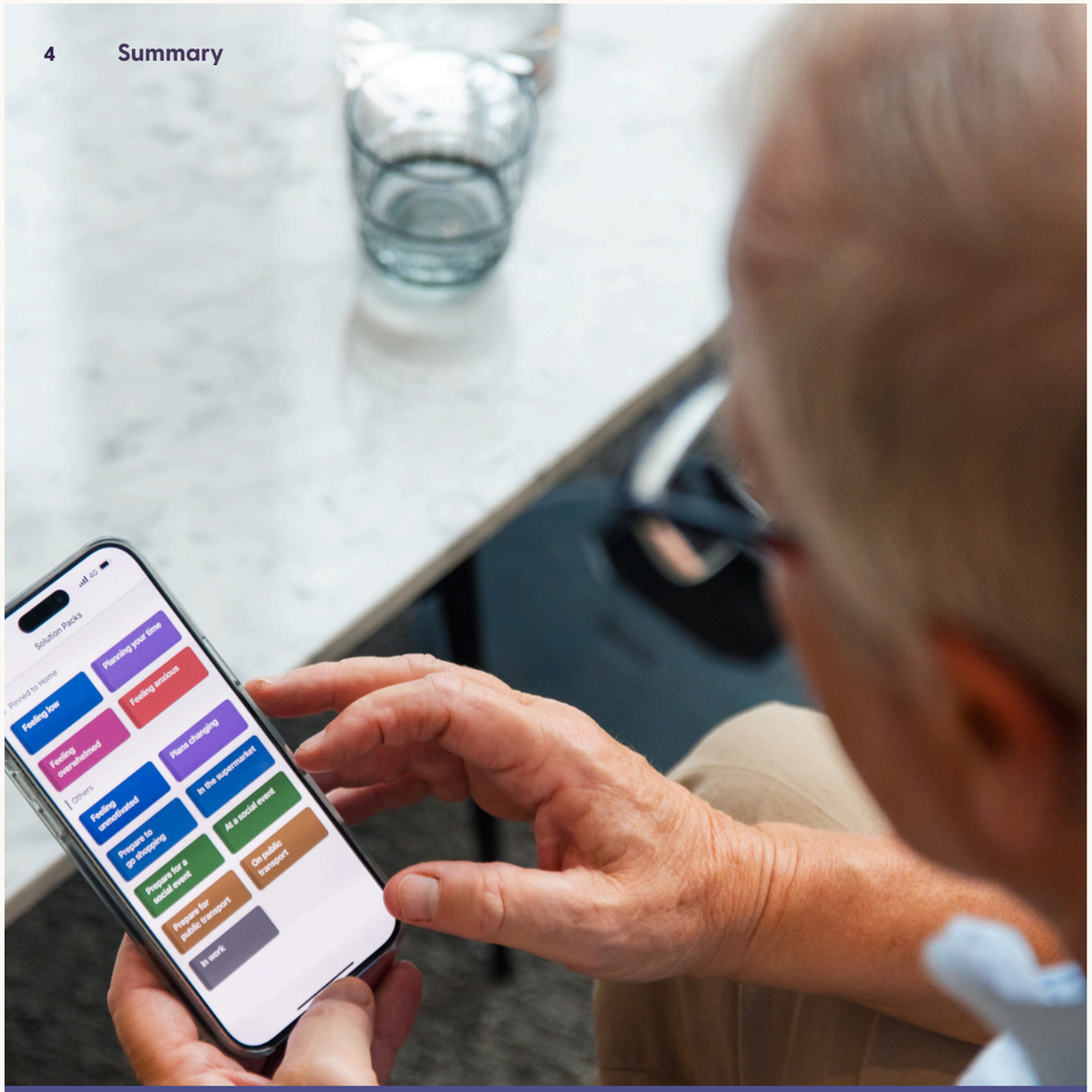
Table 1a: Survey questions asked to participants: Daily outcome measures

Context before question	Question	Answer type
The research showed that the 5-point scale used in research was not sensitive enough to see change. We wanted to see what you felt would be the best scale to represent your answers. Below you will see 5 different scaling questions. At the end of this section we will ask you to rank which scales you like the best	1. A smaller scale with wording 2. A larger scale with wording 3. A larger scale with numbers 4. A rating scale 5. A sliding scale	Open-text Is there anything you think 'this question would be perfect if it changed this part?
	Based on the pictures above, how would you rank the scales, when considering that you want to be able to give the most accurate representation of your day?	Ranking question
	Is there a different scale that could be used to measure outcomes that you would prefer?	Open text
The daily outcome measures could be asked in two ways • A statement to agree/disagree with • The start of a sentence that you can complete by indicating which word you felt better described your day	Do you feel like it is easier to answer questions in format of question one (agree/disagree), or question 2 (start of a sentence)	Open text

Table 1b: Survey questions asked to participants: Scaling preferences

Context before question	Question	Answer type
The design had a 12 week data collection period, submitting responses 5 days each week. You got a text reminder for this and submitted responses via a link. It has been supported that given it takes time to set up and see changes from Brain in Hand, this window of data collection should be longer.	What are your thoughts on how long this could be and what could be done to make it easier to complete?	Open text
	For each of the following options for the study duration, can you select firstly how often you would be prepared to provide responses **(5 times a week, 3 times a week, once a week)** and then whether you think the length is long enough to see change, and if the length would be ok for you to take part in the study.	Dropdown How many times I would be happy to submit the daily questions per week Comment on the length of study • 4-month study • 5-month study • 6-month study • 7-month study • 8-month study • 9-month study
	Please let us know which of these would be your preferences (month stuidy, submitting responses, number of times per week. Or provide an alternative preference	Open text
Based on the preferred choice you made above for the length of the study and frequency of answering questions:	Which of the following would you prefer?	Single choice • Have a break in the middle of that time, with 1 month of not doing the daily outcome measures at all • Submit responses for weeks out of the month and have one week off, then repeat • Do the same amount of times per week for the duration of the study • Other (please specify)
There have been suggestions that regularly using an online diary or other opportunity for more detailed feedback would be helping in interpreting the results	Would you have participated in this? What would make it work or not work for you?	Open text

Table 1c: Survey questions asked to participants: Duration of the study



Findings

The scales and how they are presented

The survey was sent out to all 17 people who were originally in the study, and four people completed it.

The overall finding was that participants would have been happy to use a 10 point scale instead of the 5 point scale and that they would have actually preferred it as long as such a scale utilised words not numbers. There could be potential to use a sliding scale but it wasn't liked by everyone. The main findings related to the type of scale (see Table 1b, Box 3)) are visualised in Table 2 where the preference of each participant is colour coded (darker blue being more preferred than a lighter blue).

The original SCED design employed a 5 point scale with wording, and when asked how this could be improved, one participant mentioned *"colour coding'* [SC-1] and another *"While I think this is good it feels a little too condensed especially for the more complex days"* [SC-3].

The 10 point scale with wording, which would have helped with the ceiling effect, had mixed feedback with one participant noting *"this would be overwhelming"* and another *"I personally prefer this as it gives more options and gives a broader spectrum allowing feedback to be more specific"* [SC-3].

Similar mixed feedback was given for the rating scale with one participant noting *"I like this but with colour"* [SC-1] and another finding this "too unclear" [SC-2] and *"I prefer the list format and all the thumbs up make it feel too confusing"* [SC-3]. For the sliding scale there were also divergent views : *"I like that there aren't too many options and you can slide it to where it feels best"* [SC-3].

Where there seemed to be most consensus was the dislike of the 10 point scale with numbers with feedback that *"I don't know what this means id just always push 5!"* [SC-1] and *"I'm personally not a fan of using numbers as it doesn't work for me"* [SC-3].

When asked to rank the 5 different scale options, the 10 point scale with wording was preferred by all but one participant, most did not place the 10 point numbering scale high on the list and there was a very diverse response to the sliding scale with one person ranking it their favourite and one their least favourite. This is interesting when placed alongside the observation by the senior researcher that a sliding scale would be useful. *"The reason I suggested a sliding scale is that it tips it more towards interval data so we would be much more likely to show a meaningful difference. I think it was this ceiling effect that was one of our biggest challenges on this project in terms of showing meaningful change"* [Senior Researcher].

The scales and how they are presented

The participants were also asked about their preference on how the daily measures could be asked, either a statement to agree to disagree with [option 1] or the start of a sentence that can be completed by indicating the word that best described the day [option 2]. When asked to choose which format is easier to respond to three of the participants chose option 2 noting that *"that's the way my brain works"* [SC-1] and *"because a day wasn't relaxing doesn't mean it was overwhelming and I like the idea of being able to slide to my answer"* [SC-3].

One participant noted that they felt similarly about all of them and that generally *"I think I find it hard to summarise my own qualitative experiences of a day into a quantitative scale. Especially as within a day I can go from being very happy and engaged in an interest, to being at a hospital appointment in the afternoon and being so overwhelmed I cannot speak. To then have to summarise that into a scale is very difficult. Autistic people often experience extremes of emotions, especially when subject to uncontrollable environmental factors, which means asking how a day was overall can be very difficult as we like to be honest and explain things but the design of the research doesn't allow that complexity to be captured."* [SC-4].

Participant	5 point word	10 point	10 point	Rating scale	Sliding scale
SC-1					
SC-2					
SC-3					
SC-4					
Coding	Favourite [1]	[2]	[3]	[4]	Least favourite [5]

Table 2. How would you rank the scales, when considering that you want to be able to give the most accurate representation of your day?

The outcome measures

There was mixed feedback on the outcome measures used but a general feeling that short and focused is good. However, some reported that they were maybe too vague and could therefore be mis-interrupted, with implications for validity of the measures if people were indeed just guessing. The results related to the outcome measures were captured in Section 2 of the tool (see Table 1a, Box 3) where participants were asked both about the previous outcome measures and their thought s on a list of other possible measures that could have been used.

On the original questions, two of the participants (SC-2 and SC-3) felt the questions being asked each day were the right things to be measuring, "they were short and focused on what the app was aiming to assist with" [SC-3]. The other two of the participants (SC-1 and SC-4) did not like the question "Today was a (1) overwhelming to (5) relaxing day", because *"the day could be both calm and relaxed at points but totally overwhelming at other points"* [SC-1] or *"you're trying to quantify a very qualitative measure of feelings, experiences etc..."* [SC-4]. One participant also found the question "Today I worked on the things that are important to me" confusing as they did not feel they were able to judge what is important to them *"going to the doctors is important to my physical health but isn't something that is important to my mental health or autism as it aggravates them. Examples such as that mean I didn't know how to answer the question"* [SC-4]. The only question that didn't have any negative feedback related to "I was content with the way I coped with challenges today"

Box 4. Extract from one of the exit interviews from the original SCED study regarding the outcome measure “Today was a (1) overwhelming to (5) relaxing day”

Participant: I found the last question of the texts confusing. It was like how overwhelming was the day? I think was that...I can't remember the exact words. And actually I found that really difficult to answer because some points of my day were incredibly overwhelming and distressing, and other parts of my day were really good and not overwhelming. I don't know, it was like, how do I turn that into a constructive answer when actually some parts are really good and some parts were like hell. Like I found that really hard to put a number on because I didn't know what to part it referred to.

Interviewer: What did you generally lean towards?

Participant: The middle.... so in the middle kind of thing like some days, were obviously extremely overwhelming, you know, put that. But a lot of the time, if it had been a mixture, I would put it in the middle. But I thought also that didn't reflect my experience because that wasn't reflecting the amount of overwhelm that I was feeling at points. So I don't know if there's a way that can be change in future, but I found that a really hard to sum up a whole day.

When the participants were presented with seven alternative measures all starting “Today I felt...”, three of the four participants thought some of these could be useful, with “Today I felt that I was able to be flexible in my tasks and any demands”, “Today I was able to manage overwhelm”, and “Today I looked after myself” selected by them all. One participant noted that “none of the above” were relevant because *“The questions are too vague and broad to accurately capture valid information”* [SC-4]. Instead, they recommended *“More specific, less broad questions unless you're allowing for a qualitative response. Examples of more specific things could include: 'today I feel I managed a difficulty better than I would have previously due to the support from BiH'.* The same participant later reflected in another comment that if the question is not specific, *“people guess and just randomly select something”*. This is nicely illustrated in an extract from one of the exit interviews in the original SCED study as well (see Box 4).

The senior researcher in the study reflected that *“In terms of maximum outcomes I think it is ideal to have one primary outcome and 2 or 3 secondary outcomes. I still think goals is an important outcome for executive function as one of the key definitions is that "executive function is a group of skills that govern goal directed behaviour" but a second one would be about being able to adapt to manage novel or complex situations (...this speaks very much to struggling with change and unexpected events), a third one would be making a realistic plan and being able to follow-through with it (but this could be trickier as many people don't explicitly think in this way).”*

They also suggested an alternative approach to answering standalone daily outcome metrics by collecting frequency data instead which could be built into the app itself. *“In terms of what is most commonly used in SCED I would say behaviour frequencies are the most common outcome ... if you could think of a frequency outcome measure I think that would be really good - and could perhaps even be recorded directly on BiH e.g. press the button every time you achieve x).”*

The frequency and duration of data collection

The general feeling was that 3 months was not long enough to have seen impact and that most would have been prepared to provide data for 6 months but at a lower frequency of 1-3 times a week.

The participants were reminded about the duration of the study design and an open-ended question was proposed to gather their thoughts on the duration of the study: "The design had a 12 week data collection period, submitting responses 5 days each week. You got a text reminder for this and submitted responses via a link. It has been suggested that given it takes time to set up and see changes from Brain in Hand, this window of data collection should be longer. What are your thoughts on how long this could be and what could be done to make it easier to complete?". Two of the participants [SC1 and Sc-2] noted that they hadn't really started using Brain in Hand by the time data collection stopped, though another [SC-3] thought 3 months would be sufficient *"as it takes 66 days to make a habit"*. This of course assumes that you are set up and using it by then and this wasn't the case for everyone - *"I wasn't even set up in that time frame. I suggest the study should include the entire year"* [SC-2]. As noted by one of the participants *"The data collection needs to be timed for when you have had enough sessions to understand and use BiH"* [SC-4]. The issue of duration being inadequate was also raised in the exit interviews in the original SCED study with one participant saying they had volunteered to keep providing data for longer as they thought the study duration was too short *"I offered [the] Researcher I said you know I'll carry on if you want me to. I have no problem with it because I just feel like we don't have a lot of data to go with."*

To explore what this duration should be we had a follow up question with multiple choice answers for 6 different study durations (4 months through to 9 months). The exact question was: "For each of the following options for the study duration can you select firstly how often you would be prepared to provide responses (5 times a week, 3 times a week, and once a week) and then whether you think the length is long enough to see change, and if the length would be ok for you to take part in the survey".

The findings with respect to the frequency of responses they would be prepared to complete and their perceptions on whether these durations were also sufficient to capture the impact of Brain in Hand are visualised in Tables 3 and 4 respectively. The response of the participants is colour coded as defined in each Table.

Participant	4 months	5 months	6 months	7 months	8 months	9 months
SC-1						
SC-2						Too long
SC-3				Too long	Too long	Too long
SC-4						
Coding	5/week		3/week		1/week	

Table 3. How often would you be prepared to provide responses if the study duration was increased

Participant	4 months	5 months	6 months	7 months	8 months	9 months
SC-1						
SC-3						
SC-4						
Coding	Not long enough		Not sure		Long enough for impact	

Table 4. Do you think the length is long enough to see change (SC-2 missing data)

The frequency and duration of data collection

Although most of the participants were prepared to provide responses for up to 8 months, though only at once a week, when asked directly on what combination would best work for them, two participants suggested twice a week for either a year [SC-2] or 8 to 9 months [SC-4]. One participant did not respond to this question and the other one suggested “six months but not 5 days as a week as I often missed them or just wrote whatever I felt like due to tiredness so honesty was reduced 3 times would reduce fatigue and allow more detailed and honest answers” [SC-3]. This last comment reinforces the importance of weighing up data quantity with quality.

When asked about the number of data points needed to enable an effective SCED design our senior researcher responded that “… we continue to guide our students that we aim for 35 data points with a minimum baseline of 5 days, minimum of 5 phase changes (i.e. randomised start time of intervention phase) - this allows for some missing data when including a minimum of 4 participants. Obviously, the more participants and data points the greater the power of the study.”

The researcher also suggested the option of having a break in the middle of the study:

“I think breaks in data collection may work best - I have talked with colleagues about how we could provide the intervention for a period after baseline and our intervention phase is basically a couple of months later so we see the longer-term effects (so the design would be the same but there would be a gap between baseline and intervention)”.

When we asked our participants that given their response on the duration and frequency and that “It is important for us to see impact through lots of data points across a time period longer than three months” which of three options they would prefer to go alongside that no one picked “Have a break in the middle of that time, with 1 month not doing the daily outcome measures at all “as suggested by the researcher – with two choosing “ Do the same amount of times per week for the duration of the study “ [SC-2 and SC-4] noting they had both suggested two times a week previously and one participant who had previously noted the challenges with fatigue chose “Submit responses for 3 weeks out of the month and have one week off, then repeat.” [SC-3]

An online diary or qualitative feedback

In the initial SCED design we discussed having a diary approach in addition to daily metrics but decided that the ask of the participants was already high and we did not want to add on another task that may affect the quality of the daily outcome data.

Interestingly when we now asked if they would have participated in “regularly using an online diary or other opportunity for more detailed feedback would be helpful in interpreting the results” only 1 person said no [SC-2] with one saying that “yes, this seems vital” [SC-4]. Although SC-3 did note that “I would but I know I would not be willing to make detailed answers every time as this is something I sometimes struggle with”, the general sense was that this would have been a good idea and something to consider in the future even if it is as one participant suggested “ Having a text box to explain your answer “ [SC-4]

Discussion

Why was it important to do this?

Many researchers have studies that do not go to plan, especially when attempting something new and innovative. We were faced with two options when we examined the findings of our SCED study, shelve it or explore it further. We went for the latter for a number of reasons. Firstly, because we felt accountable to all the stakeholders and participants in the study to do something with this. Secondly, we have a strong ethos around learning and see "challenges" or "failures" as opportunities. And thirdly, we really believed that a SCED design was right for our user group and the type of support we are providing.

As noted by the researchers, SCEDs has an intrinsic value in understanding individual differences, which is so important to our user group

"The SCED allowed investigation of individual response differences, which are to be expected given the individualisation of the support offered and showed that some participants reported significant improvements in coping with challenges, while others reported significant reductions in feelings of overwhelm. Such individual differences in responses would be lost in a study design focusing on group outcomes (e.g., randomised controlled trial designs)... This study suggests that as a relatively new avenue of research, SCED is likely to continue to permit more detailed examination of who BiH may be most useful for and what effects different users may experience, helping to answer the important broader of question "what works for whom?". [Exeter University researchers]

What have we been able to achieve with these new insights?

Reviewing the data from the four people who fed back on the study as well as one of the lead researchers we feel that this exercise has been extremely worthwhile. As well as providing insights into how and what types of questions we could have asked instead, they encouraged us to explore both co-design and a diary approach to data collection. Taking with us the feedback from the study participants – avoid numbers and go with words, be specific, and ask about things that may not change during the day – and the themes suggested by the researchers (goal directed behaviour, managing or adapting to novel or complex situations) – we established a codesignworking group, that includes one of the SCED participants, who came up with a set of revised outcomes (see Box 5). It is hoped that these clearer metrics will lead to better quality data as people will not find them confusing and respond reliably and honestly. Furthermore, we have also been able to build on the suggestion to enable narratives to go alongside the metrics in a diary approach that was initiated at the end of October 2024. We hope that all these changes will make for a better experience for the participants.

Box 5. Exploring co-design in developing SCED outcomes and a diary approach with people with lived experience

In September 2024 we created a co-design working group to help with the design of a diary study. The group consists of one of the SCED participants, and five Brain in Hand staff. The Brain in Hand staff are all Autistic or have ADHD and bring their lived experience to the group.

The initial task was to define the outcome measures that would be asked to participants building on the feedback received from four of the SCED participants and the themes suggested by the researchers. The group discussed how the outcome measures needed to be clear and concise. The language used was also discussed at length to ensure ambiguity was minimised.

The result of the discussions was to establish four outcome measures:

1.Using the Brain in Hand traffic lights, please select which feels most relevant (Image of the Brain in Hand traffic lights with option to select Red, Amber, or Green).
2.Thinking back on the past 24 hours, I feel I managed a situation that used to be more difficult for me
3.Thinking back on the past 24 hours, I feel I did something good for myself
4.Thinking back on the past 24 hours, I feel I engaged with a goal I have set myself

We included the first question on the traffic lights to understand the mood of the user when they are completing the questions. This question was included following consultations with staff on how mood can affect the responses given. We felt this was an opportunity to understand the relationship between how the participant registers their mood and their responses.

Box 5. (continued)

In questions 2 to 4 we replaced “Today” used in the original SCED study with “Thinking back on the past 24 hours” Thoughts from the group were that users may not have anything to respond if the question asked about ‘today’, so asking about the past 24 hours would mean the user could think about morning, afternoon, evening, and during the night. We also inserted “I feel” before the specific outcome as this was suggested by one of the SCED participants in the initial feedback.Question 2 aimed to capture the issue of “managing or adapting to novel or complex situations” suggested by the researchers and similar to the original question ““I was content with the way I coped with challenges today” but modified to be really specific about “situations that used to be more difficult”; and removing the word “content” as and we also removed “content” as the group found “content” as the group felt that content might be a bit leading (toxic positivity), it is important to acknowledge that things can be difficult, and it is important to remain neutral. Question 4 builds on the theme of “goal directed behaviour” and the original SCED question of “Today I worked on the things that are important to me” but being more specific that this was “a goal they had set themselves” as feedback from the codesign group was that there are lots of things that are important to them ranging from what may be considered as a short term goal such as getting up that day and brushing your teeth, to a long-term goal such as getting a new job. The SCED participants and the codesign group did not like the original question in the SCED to rank on a five-point scale whether their day was overwhelming (one) to relaxing (five). The major reason was that the day could be both overwhelming and relaxing at different times during the day. Instead the group worked on a positive outcome related to wellbeing. The wording of ‘did something good for myself’, was used to incorporate both self-care activities and activities that would improve the user’s quality of life. Thoughts from the group were that this could range from doing the washing up, to meeting up with friends.

Box 5. (continued)

Outcome measures will be asked on a 10-point scale. The 10-point scale was used as it gives a wider scale than the initial 5-point adopted in the SCED study. We have opted to remove a mid-point option as feedback from Brain in Hand staff was that the respondent will need to select the level to which they agree or disagree to the statement. Further feedback was that when a participant in a survey is unsure of their answer, or would like to answer quickly, they usually opt for the mid-option. Additionally, a mid-point option of ‘neither agree nor disagree/neutral’ does not provide data as useful as selecting an agree/disagree option.

The options provided are: (1) Strongly Disagree, (2) Disagree, (3) Mostly Disagree, (4) Moderately Disagree, (5) Slightly Disagree, (6) Slightly Agree, (7) Moderately Agree, (8) Mostly Agree, (9) Agree, (10) Strongly Agree.

These options were selected in consultation with the codesign group. The group found these options to be different enough for participants to respond to.

Participants will be able to select an option of a dropdown with words, or a sliding scale out of 10 with words on each end. These two options are due to limitations on the platform. Participants are given the option at the start of the study to select how they would like to view the questions.

Participants will be asked to complete these outcome measures at least 2 times a week and the study aims to run for 6 months. This should ensure that at least 35 days of outcome data is available for each of the participants as suggested by the researchers and also aligns with the feedback from the SCED participants on how long and how often they would complete data.

Box 5. (continued)

A key feedback from the SCED participants was the importance of adding context to their outcome responses through narrative. The diary aspect of the study will be able to build on that by providing not only the opportunity to provide text but also photos and videos with a caption, and audio recordings. The diary study has been set up in a way to understand what went well that day, what may have not gone so well during the day, and any other comments. The codesign group has been working on the prompts that will be asked for each section. For example, in the challenges section, prompts include:

- What could you do differently if something similar happened again?
- Was there anything you tried that didn't work the way you wanted it to?
- Have any situated occurred that you think you might discuss with your coach?

Where are we taking this learning next?

The research team at BiH run a number of surveys and research projects and we plan to take the learning from this followup study into these going forward, specifically around five issues. The first is how we phrase our questions, ensuring that we are really clear and specific (so there is no ambiguity). The second is to think really carefully about how we provide options for people to respond to quantitative metrics, making it as easy as possible and looking at ways that we can provide choice in how they respond. The third is to provide choice on when they would like to respond, this is often implicit when a survey is sent out or an indepth interview is put into place, but when there are prompts being sent out or in-app questions, can we work to have these sent at times that best work for individuals. The fourth is to ensure that there is always the option to provide qualitative feedback in addition to quantitative, so that participants can provide more contextual and nuanced feedback on how and why they may have scored something one way or another. And finally, the value and insights from codesign. Despite people coming in with distinct individual preferences and ideas we have seen with the workshops that given time and space people are able to work together, express themselves and listen to other peoples views and then decide and agree on how to approach things.

Conclusion

In reimaging our SCED study we went back to the participants to discuss the findings and gather insights into what we could do differently. We learnt there is real value in doing this and it gave us the confidence to explore codesign in our research in an authentic way. We still have a great deal to learn but our excited to see how our reimagined SCED pans out within our new diary study and what is achievable by involving people with lived experience and our potential BiH users in the design of our research.

References

Bartels, S. L., Johnsson, S. I., Boersma, K., Flink, I., McCracken, L. M., Petersson, S., & Wicksell, R. K. (2022). Development, evaluation and implementation of a digital behavioural health treatment for chronic pain: study protocol of the multiphase DAHLIA project. Bmj Open, 12(4), e059152

Birdsey, N., & Walz, L. (2021). Exploring the concurrent use of standardised and idiographic measures to assess cognitive behavioural therapy in a university student with autistic spectrum condition–a single case experimental design. The Cognitive Behaviour Therapist, 14, e18.

Bulté, I., & Onghena, P. (2008). An R package for single-case randomization tests. Behavior Research Methods, 40(2), 467-478. https://doi.org/10.3758/BRM.40.2.467

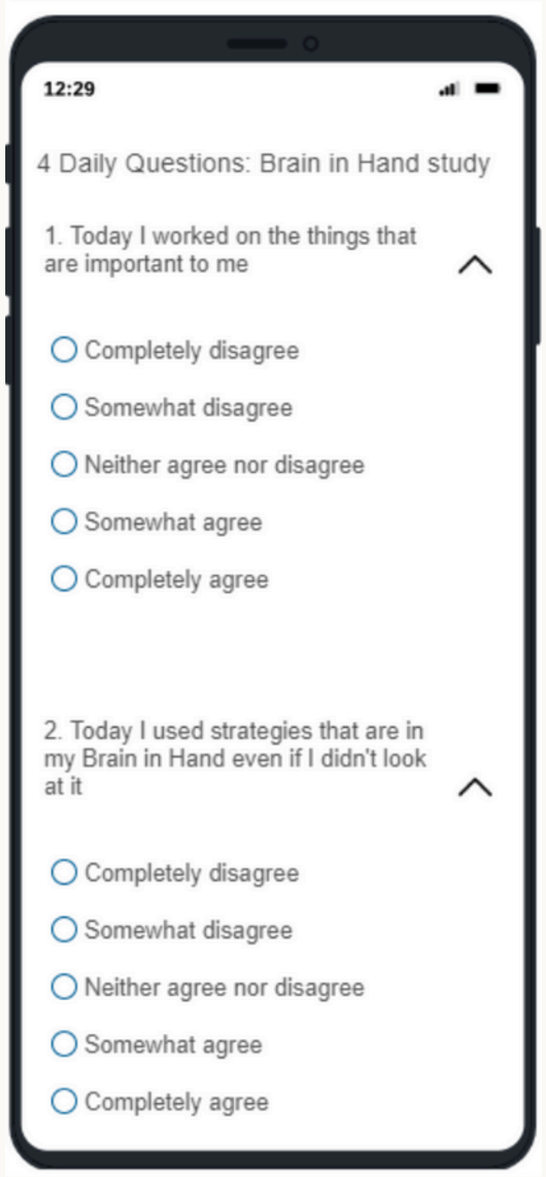
Epstein, L. H., & Dallery, J. (2022). The family of single-case experimental designs. Harvard data science review, 4(SI3).

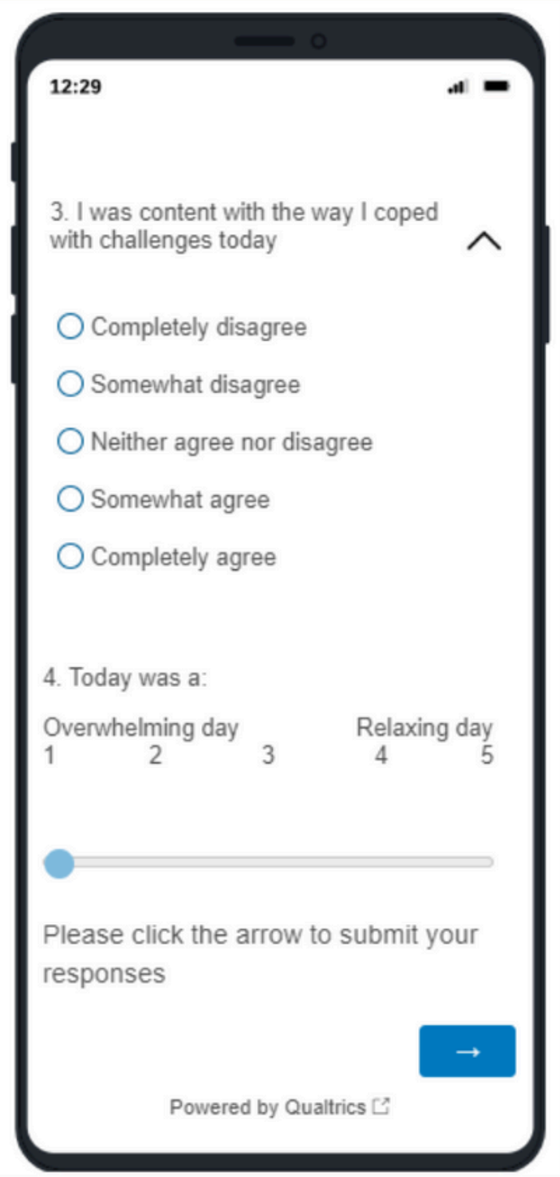
Guyatt, HL (2023). Three essential changes to evidence that will drive adoption of digital innovation. Open Access Government - https://www.openaccessgovernment.org/evidence-drive-adoption-digital-innovation/157477/

Krasny-Pacini, A., & Evans, J. (2018). Single-case experimental designs to assess intervention effectiveness in rehabilitation: A practical guide. Annals of physical and rehabilitation medicine, 61(3), 164-179.

Annex

Annex 1: Screenshots of the daily outcomes





Annex

Annex 2: Screenshots of the questionnaire

Section 2: Daily Outcome Measures

1. The following outcome measures were sent to you daily via a link in a text for you to rate:

1. Today I worked on the things that are important to me
2. Today I used strategies that are in my Brain in Hand even if I didn't look at it
3. I was content with the way I coped with challenges today
4. Today was a (1) overwhelming to (5) relaxing day

Do you think these were the best things to be measuring?

Please explain why

2. **Would there be more relevant daily outcome measures, from your perspective?**

Noting that you would be asked about these on a frequent basis

* 3. **Now that you have had some time using Brain in Hand, do you think any of the following would be useful to track in the daily outcome measures?**

Select all relevant options

- ☐ Today I felt organised
- ☐ Today I felt that I was able to be flexible in my tasks and any demands
- ☐ Today I was able to manage overwhelm
- ☐ Today I looked after myself
- ☐ Today I was able to problem solve/make decisions
- ☐ Today I was able to manage stress
- ☐ Today I was able to manage anxiety
- ☐ Other (Please add in comment box below)
- ☐ None of the above

Please could you explain why you have chosen these options?

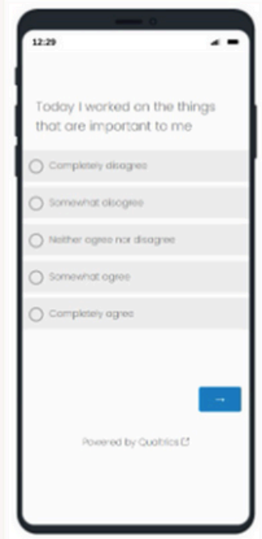
Prev Next

Section 3: Scaling preferences

The research showed that the 5 point scale used in the research was not sensitive enough to see change. We wanted to see what you felt would be the best scale to represent your answers.

Below you will see examples of 5 different scaling questions. At the end of this section we will ask you to rank which scales you like best

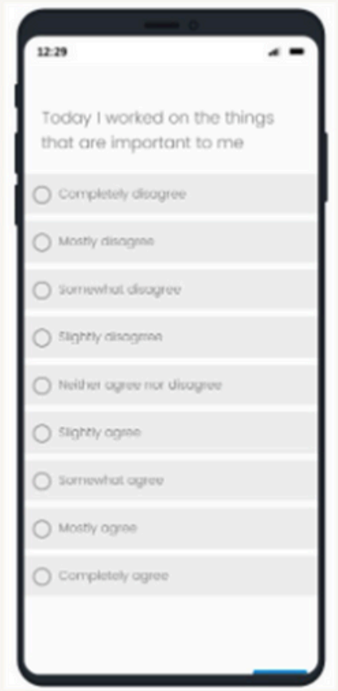
1. **A smaller scale with wording:** *This is the example that was used in the research*



Is there anything you think 'this question would be perfect if it changed this part'?

Please share what could make this scale better

2. **A larger scale with wording:**



Is there anything you think 'this question would be perfect if it changed this part'?

Please share what could make this scale better

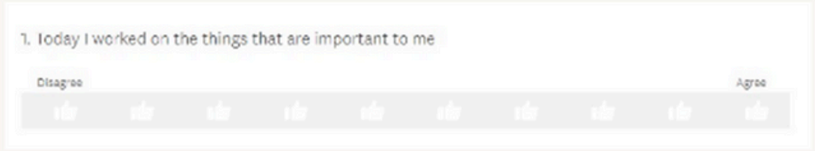
3. **A larger scale with numbers:**

Is there anything you think 'this question would be perfect if it changed this part'?

Please share what would make this rating scale better

Annex

4. **A rating scale:**

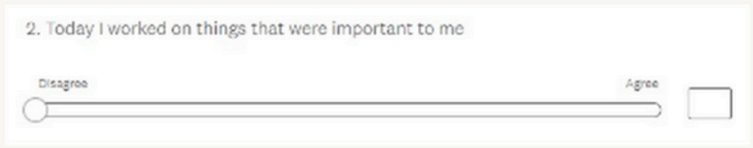


Is there anything you think 'this question would be perfect if it changed this part'?

Please share what you think would make this rating scale better

5. **A sliding scale:**

Today I worked on things that are important to me

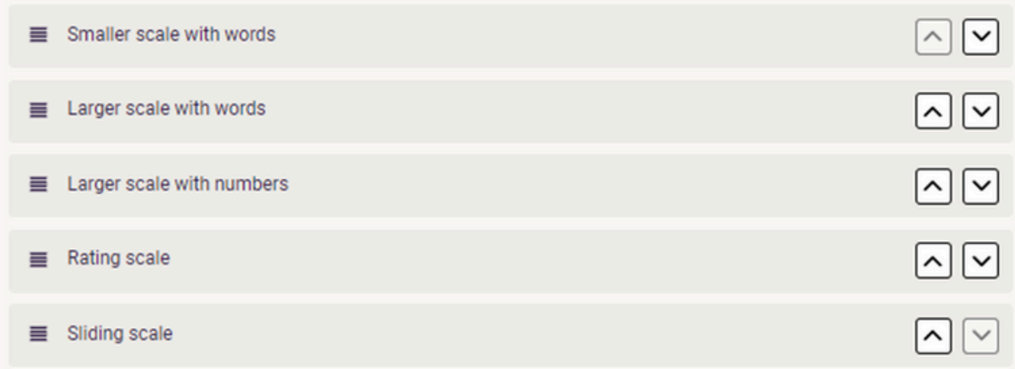


Is there anything you think 'this question would be perfect if it changed this part'?

Please share what could make this sliding scale question better

* 6. **Based on the pictures above, how would you rank the scales, when considering that you want to be able to give the most accurate representation of your day?**

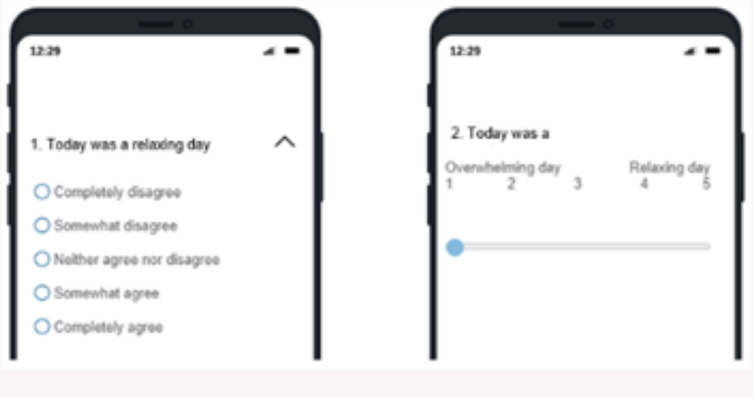
You can move the toggle or arrows to change the order of the scale. Number one on the list will be your favourite option, the fifth on the list is your least favourite.



7. Is there a different scale that could be used to measure outcomes that you would prefer?

* 8. **The daily outcome measures could be asked in two ways:**
- A statement to agree/disagree with
- The start of a sentence that you can complete by indicating which word you felt better described your day.

This shows an example of one statement written in the two ways.



Do you feel like it is easier to answer questions in the format of Question 1 (photo on the left), or Question 2 (photo on the right)? Can you explain why?

Prev Next

Section 4: Duration of the study

1. The design had a 12 week data collection period, submitting responses 5 days each week. You got a text reminder for this and submitted responses via a link.

It has been suggested that given it takes time to set up and see changes from Brain in Hand, this window of data collection should be longer.

What are your thoughts on how long this could be and what could be done to make it easier to complete?

2. **For each of the following options for the study duration can you select firstly how often you would be prepared to provide responses (5 times a week, 3 times a week, and once a week) and then whether you think the length is long enough to see change, and if the length would be ok for you to take part in the survey.**

	How many times I would be happy to submit the daily questions per week	Comment on the length of study
4 month study		
5 month study		
6 month study		
7 month study		
8 month study		
9 month study		

Please also let us know:
Which of these would be your preference if you had to choose one (_ month study, submitting responses _ times a week)
Or if we have missed a preference you would choose?

3. **Based on the preferred choice you made above for the length of study and frequency of answering questions:**

Which of the following would you prefer to go alongside that?

It is important for us to see impact through lots of data points across a time period longer than three months.

- ◯ Have a break in the middle of that time, with 1 month not doing the daily outcome measures at all
- ◯ Submit responses for 3 weeks out of the month and have one week off, then repeat.
- ◯ Do the same amount of times per week for the duration of the study

Other (please specify)

4. **There have been suggestions that regularly using an online diary or other opportunity for more detailed feedback would be helpful in interpreting the results.**

Would you have participated in this?
What would make it work or not work for you?

Prev Next

For more information visit: www.braininhand.co.uk
Email: info@braininhand.co.uk
Telephone: +44 (0)1392 247909

brain in hand