A Quick-Start Guide to Designing, Running, and Analysing an Experience Sampling Study

Version Date: 14/06/16

Jessie Sun
jesssun@icloud.com

Personality Processes Lab

THE UNIVERSITY OF MELBOURNE
MELBOURNE SCHOOL OF PSYCHOLOGICAL SCIENCES

Table of Contents

Introduction 3

Designing and Running an ESM Study 5

Design Considerations 5

Participant Motivation, Compliance, and Remuneration 10

Preparing ESM Data 12

Data Cleaning 12

Computing and Centring Variables 27

Dates and Timecourse 28

Mplus Data File 34

Bonus: Creating Lagged Variables 35

Multilevel Modelling Using Mplus 37

Multilevel Modelling of Personality Processes 37

Data Screening 39

Calculating Scale Reliabilities for Repeated State Measures 40

Within-Person Mediation Models 43

Example Write-up 51

Troubleshooting and Further Questions 54

References 55

Appendix: Example Wellbeing Report 58

These sticky notes will describe my **top recommended must-reads** (or must-cites…) for that section, and why they're so helpful and important.

You'll also see example methods and results write-ups in these boxes. To see how these "bits" would fit in the context of a full paper, see bit.ly/1Kjycdc.

Introduction

How do personality processes unfold in everyday life? Experience sampling methods (ESM) allow us to examine this question by measuring the affects, behaviours, and cognitions of individuals in their *natural settings*, in *real time* (or close to real time; e.g., "How lively do you feel *right now*"; "How talkative were you *in the past hour*"), and on *repeated occasions* (Mehl & Conner, 2012, p. xix). This manual will specifically focus on study designs in which individuals respond to a questionnaire on a smartphone app, several times each day, typically across 1-2 weeks[1].

Why might this be a good idea? Laboratory experiments have their place, as they enable a high degree of experimental control, and therefore stronger causal inferences. In contrast, the type of ESM design discussed in this manual involves tracking people's experiences as they naturally unfold, across a range of situations they may encounter in everyday life. As this type of data is still correlational, they are causally ambiguous. However, ESM provides increased ecological validity and generalisability across contexts, in contrast to many laboratory-based studies, where participants are observed in a limited range of (often artificial) contexts.

ESM also captures a different type of information, compared to traditional questionnaire methods. This is because the way we describe our experiences *in the moment* (the "experiencing self") can be very different to the way we describe our experiences *in general* (the "believing self") or over a period of time (e.g., the past week; the "remembering self"). ESM reports are less susceptible to memory biases, relative to traditional questionnaire methods. However, this doesn't necessarily mean that ESM methods are *better* than traditional self-report measures (e.g., "biased" global or retrospective assessments can be good predictors of future choice); rather, they provide different types of information (see Conner & Barrett, 2012).

Finally, in the context of personality research, a major trend in the past decade has been an increased focus on *within-person processes*, and not just *between-person differences* (e.g., Fleeson, 2001). Whereas *nomothetic* methods aim to identify associations between variables across a population of individuals (e.g., What is the relation between trait extraversion and trait positive affect [PA]?), *idiographic* methods aim to do this across a population of experiences or situations for a given individual, and do not assume that the relation between variables is the same for every individual (e.g., What is the average relation between extraverted behaviours and PA states for Luke?). Multilevel modelling (MLM) then allows us to combine both idiographic and nomothetic approaches to answer questions about people in general (e.g., What is the average

---

[1] ESM belongs in a larger family of methods for studying daily life (see Mehl & Conner, 2012). For example, *daily diary* designs involve less intensive observation, as participants might be asked to complete just one questionnaire each day. Whereas ESM and daily diary studies both rely on self-report, another recent method, the *Electronically Activated Recorder* (EAR; Mehl & Robbins, 2012), involves passive behavioural observation by recording snippets of sounds in participants' daily lives (e.g., Mehl, Vazire, Holleran, & Clark, 2010).

relation between extraverted behaviours and PA states? Is this relation greater for extraverts?), while starting at the level of the individual (Conner, Tennen, Fleeson, & Barrett, 2009).

The purpose of this manual is to provide a practical guide to designing, running, and analysing an ESM study, based on what I learned through the course of an Honours project in the Personality Processes Lab. Part I briefly outlines some key design considerations, including power, participant compliance, and organising participant data. Part II provides a user-friendly guide to data cleaning using SPSS syntax (and Excel). Finally, Part III describes how to conduct and interpret MLM analyses using MPlus.

For Parts II and III, I have provided simplified, deidentified datasets to allow you to follow along with the manual and try this yourself. For background on the example data and research questions, see bit.ly/1Kjycdc. These datasets will be uploaded onto GitHub (https://github.com/jesssun/esm-manual), where I plan to migrate this manual to facilitate usage and edits/improvements (potentially including extensions how to do all of this using R!). Therefore, please let me know if you notice any errors or have suggestions for improvements.

-Jessie Sun

Reis, H. T. (2012). Why researchers should think "real-world": A conceptual rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 3–21). New York: Guilford Press.

Discusses the kind of information that daily life methods provide, highlights how they complement more traidtional methods, reviews conceptual bases (ecological validity, the value of field research, the need to take context seriously), and discusses the value of descriptive data.

Conner, T. S., & Barrett, L. F. (2012). Trends in ambulatory self-report. *Psychosomatic Medicine*, *74*(4), 327–337. http://dx.doi.org/10.1097/PSY.0b013e3182546f18

An excellent discussion of the "experiencing self" vs. the "remembering" and "believing" selves".

Conner, T. S., Tennen, H., Fleeson, W., & Barrett, L. F. (2009). Experience sampling methods: A modern idiographic approach to personality research. *Social and Personality Psychology Compass*, *3*(3), 292–313. http://dx.doi.org/10.1111/j.1751-9004.2009.00170.x

Explains the idea of idiographic vs. nomothetic approaches.

Designing and Running an ESM Study

Design Considerations

As with any other study, the first step is to determine your key research questions and target variables, and the population of interest. Here, I'll assume that you have some idea of what questions you're interested in asking (or, see Conner & Lehman, 2012), and will instead focus on these issues: sufficient sample sizes, sampling protocol, technology, participant compensation and motivation. This table summarises these considerations, along with some recommendations.

Decision	General Recommendations
Number of participants and reports	
How many participants will I recruit?	The more the merrier—always! Bare minimum ~60, but if you're interested in Level 2 relationships, you should consider recruiting > 100 participants.
How many reports should I ask participants to complete?	Typical studies sample 4-6 reports per day across 1-2 weeks (28-72 total; *M* = 20-50 assuming 70% completion rate). This is generally sufficient, but you may need more reports if you want to look at lagged effects and situational moderators.
How long will the study be?	Generally 1-2 weeks, depending on your resources and research questions.
How many observations each participant, and how many reports per day?	Most studies use 4-6 reports per day, but you could set more reports if you have a shorter survey. Within reasonable limits of participant burden, it's better to have more observations than not enough, especially if you want to look at situational moderators.
ESM survey details	
What kind of sampling protocol?	For regularly-occurring phenomena, signal-contingent (variable schedule) is best. For irregular events, use event-contingent sampling.
How long will participants have to complete the report?	Within 15 or 30 minutes of the signal. It's possible to have a "snooze" function if your software allows it.
What timeframe will questions refer to?	e.g., "In the [*past hour*/*past 30 minutes*]…"; "*Right now*…"
How many questions per report?	Aim to keep reports short (< 2 mins), especially if you're sampling more intensively.
Fixed or random presentation of items?	Random is ideal, but depends on the limitations of your software.
What scale will you use?	0-10 seems pretty intuitive; you can also ask participants to indicate time periods on a scale from (for e.g.) 0-60 minutes.
Participant motivation & compliance	
How will participants be compensated?	Baseline compensation + compliance-contingent payment (e.g., minimum % or $/report; lottery draw; feedback).
How will you encourage a sense of responsibility to the research(er)?	Emphasise the importance of the research; stay in touch with participants during the study (text messages/call them); tell them you'll be in touch at the end of the week for a "debrief".

Sufficient sample sizes. How many participants do you need to (1) detect an effect, and (2) accurately estimate your parameters? Statistical power refers to the likelihood that a study will detect an effect when there is a true effect (see Fraley & Vazire, 2014, for a review of why this is important). If you don't have enough power (typically .80 is the target), then you may invest heaps of resources into running an ESM study, only to not actually be able to detect an effect.

In an ESM study, power is a function of several things: sample size ("Level 2 clusters"), number of observations per person (Level 1), and expected effect sizes. Ideally, if you have some idea of the expected effect size, then you could do a power analysis using something called a Monte Carlo simulation. But otherwise, here's a handy reference:

> to obtain sufficient power to detect cross-level interactions at least 30 groups[2], and 30 observations within each group, are needed. It is also observed that 60 groups, with 25 observations per group (total n = 1500), will produce sufficiently high power. With fewer groups, for instance 30, many more observations per group are needed to obtain a power of 0.90. When many groups are present, for example 150, five observations per group will suffice to obtain a power of 0.90, bringing the total number of observations to 750. Using fewer observations (either groups or individuals) leads to a rapid decline of power for the detection of cross-level interactions. (Kreft & de Leeuw, 1998, p. 125)

When it comes to adequate sample size, statistical power isn't the only consideration. In another simulation study, Maas and Hox (2005) were interested in the question of accurate estimation of regression coefficients, variance components, and standard errors. They found that:

> only a small sample size at level two (meaning a sample of 50 or less) leads to biased estimates of the second-level standard errors. In all of the other simulated conditions the estimates of the regression coefficients, the variance components, and the standard errors are unbiased and accurate. (Maas & Hox, 2005, p. 86)

Snijders (2005) emphasises that it's usually desirable to have as many clusters as possible at the top level, and that this is even more important than the average size of the clusters:

> for testing the effect of a level-one variable, the level-one sample size (in the example, 3,300) is of main importance; for testing the effect of a level-two variable it is the level-two sample size (150 in the example); etc. The average cluster sizes (in the example, 22 at level two…) are not very important for the power of such tests. This implies that the sample size at the highest level is the main limiting characteristic of the design. Almost always, it will be more informative to have a sample of 60 schools with 3,300 pupils than one of 30 schools also with 3,300 pupils. A sample of 600 schools with a total of 3,300 pupils would even be a lot better with respect to power, in spite of the low average number of students (5.5) sampled per school… (Snijders, 2005)

[2] Here, the Level 2 cluster was a group, comprised of individuals at Level 1. In an ESM study, the Level 2 cluster is the individual, with multiple reports per individual at Level 1.

Snijders (2005) also notes that small cluster sizes are not a problem for testing fixed regression coefficients, but will limit your ability to test random slope variances at the higher level (e.g., in ESM studies, the between-person variances of Level 1 effects).

Here are some of my key take-aways from these recommendations:

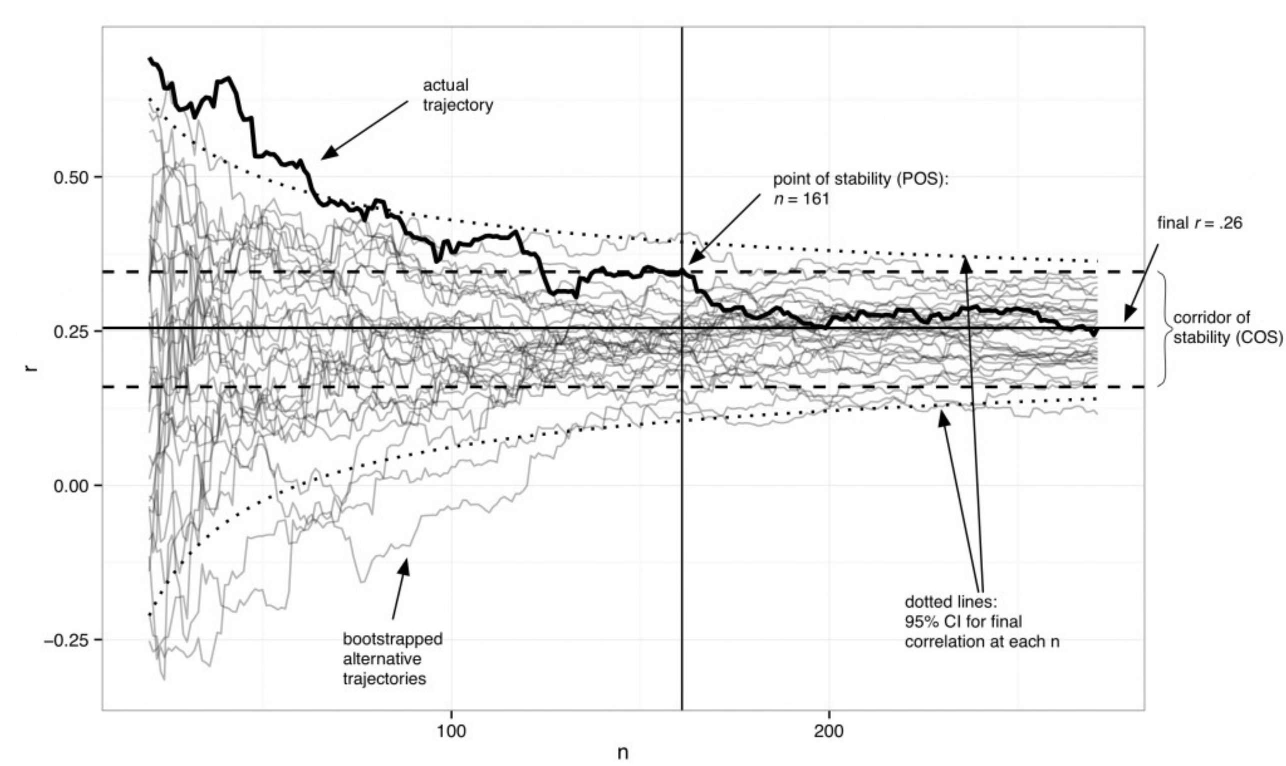
1. You need at least 50-60 participants for adequate power to detect cross-level interactions and accurately estimate Level 2 standard errors
2. The sample size that matters most is the number of observations/clusters at the level the effect is measured
3. The number of clusters at the highest level (in an ESM study, that's the number of participants) is typically the limiting factor in a design

But, is 50-60 participants *really* enough? If you're just interested in simple Level 1 effects (e.g., do extraverted behaviours predict increased PA states?), then sure, you would have plenty of power to detect this within-person relation if you have even 5–10 observations per person.

However, a lot of the time, Level 2 effects are also of interest (especially for personality research). For example, a very basic question that I asked in my ESM study was: Do trait extraverts report greater average levels of extraverted behaviours? This didn't involve multilevel modelling; rather, for each of 62 participants, I calculated the mean level of extraverted behaviours that they reported across up to 42 observations ("aggregate state extraversion"), and simply correlated this with their trait levels of extraversion that we measured at baseline.

Based on a meta-analysis of 15 studies (Fleeson & Gallagher, 2009), I expected a correlation around $r = .42$, which is a fairly substantial effect. G*Power says that we only need 42 participants to detect this effect size with .80 power. But let's ignore this and assume that the expected effect size was actually half that size: .21 (which is actually the average effect size for published personality research; Richard, Bond, & Stokes-Zoota, 2003). G*Power tells me that we need at least **175 (!)** participants to detect a correlation of this size with .80 power. And to detect a correlation of .30 with .80 power, you need about 84 participants.

In fact, I found no relation between trait extraversion and aggregate extraverted states (r = -.001!). There were some differences in the methods, but this unexpected null finding made me look more closely at the individual studies in the meta-analysis, which revealed that there was massive variability in the correlations (*r*s = -.21–.70). These studies had *N*s ranging from 12–63. To me, this suggests that we need larger sample sizes to get more accurate estimates of these Level 2 correlations. I think this diagram on the evolution of the correlation (Schönbrodt & Perugini, 2013) illustrates this issue best:



This shows that correlations typically stay in the $\pm .1$ corridor of stability after around *N* = 161, but only stablise when *N* approaches 250 (in typical scenarios).

So, all things considered, **bigger is clearly better when it comes to Level 2 sample size.** Depending on the resources (time, $$) you have on hand, 161 participants may not always be feasible. But let's say that **100**[3] is a bare minimum (see Allen & DeYoung, 2015) if you're interested in medium-sized Level 2 effects. It's still not ideal, but it's better than 50-60! Otherwise, (1) constrain your scope to Level 1 questions, or (2) accept that your Level 2 analyses may be underpowered and draw appropriate conclusions based on the strength of your evidence.

Sufficient number of observations. Typical ESM studies sample about 4–6 responses per day across 1-2 weeks (28–72 potential observations). Assuming a completion rate ~70%, that gives you about 20–50 reports per participant. Is that enough? Again, it depends on your research questions. As mentioned above, if you're interested in relatively simple effects (e.g., the within-person relation between extraverted behaviour and state PA), then even 5–10 observations per participant is probably enough to detect this effect.

But, here are a couple of situations when you might want more observations per person. First, if you want to investigate lagged effects (e.g., does extraverted behaviour at the previous report predict PA at the next report?), you will need to exclude several reports because of the timeframe of the effect. For example, how long do you expect the effects of extraverted behaviour to last for? We don't really expect extraverted behaviour the night before to predict PA in the morning. Nor to we expect extraverted behaviour to have much of an effect on PA 6 hours

[3] Taking into account potential exclusions (e.g., drop-outs, technological issues, people who don't complete enough reports), this realistically means oversampling and aiming for about 120 participants.

later (e.g., if participants missed a report in between). So, if you are interested in lagged effects, you need to anticipate such exclusions and set a longer sampling period or more reports/day.

Second, you might also be interested in situational moderators (e.g., do effects depend on who the participant is, where they are, and what they're doing?). If so, you would need more reports to capture the full range of these situations. This is especially important for categorical moderators (e.g., is there a stronger relation between extraverted behaviour and PA states in social situations than when participants are alone?) as you would want a reasonable number of reports for each participant in each situation (it's kinda like adding another level to the structure of the data).

That said, ESM does ask a lot of participants, so we need to always consider participant burden. This means taking seriously the tradeoff between number of reports and the number of questions you're asking in each report, or upping incentives to maintain motivation.

ESM survey details. Once you've decided on the length of the study and number of reports per day, you need to decide on how these reports will be distributed across the day, how long participants will have to respond to each report, the timeframe of the questions, and whether the questions will be randomised.

Sampling protocols. There are three key types of sampling protocols. *Fixed timing schedules* (i.e., interval-contingent sampling) ask participants to complete reports at set times every day. In contrast, for *variable timing schedules* (i.e., signal-contingent sampling), participants complete reports when they are signalled by the app, at unpredictable times. Finally, *event-contigent sampling*, where participants complete reports following a pre-defined event, is good for examining the experiences around specific (and usually infrequent) events (e.g., conflict, lying, smoking, drinking). Conner and Lehman (2012) discuss the advantages and disadvantages of fixed vs. variable schedules in great detail, but a key advantage of a variable timing schedule is that you would capture a greater sample of times during the day (e.g., not just 9am, 12pm, 3pm, 6pm, and 9pm, but also the times in between).

Survey completion window. You also need to decide how strict you're going to be with when you allow participants to complete reports. For example, you might set a report to expire so that participants are no longer allowed to complete that report if they don't complete it within 5, 15, or 30 minutes. To an extent, giving participants too generous a window of time might bias their responses as they may choose to delay the response until they "feel like" completing it. However, making the window too tight may increase the chance that the participant simply does not see the notification in time to complete the report, or cannot complete it because they are driving, in the shower, or in an otherwise inconvenient time. So, I would err on the side of generosity (e.g., a 30 minute window), unless a high degree of randomness is essential for your research questions. If possible, it's helpful to program automatic reminders (e.g., halfway through the expiry time, and 5 min before the report expires).

If your software allows it, one option is to include a "snooze button" functionality, which allows participants to delay the report for 10–15 mins. Note that for either of these options—a generous survey completion window or a snooze button—you should ask participants to complete the survey about their experiences at the time when they complete it (instead of retrospecting). For example, if the initial notification is at 10am, but participants complete the report at 10.20am, they should report on their experiences at 10.20am (or the hour preceding 10.20am, for questions in the form "*In the past hour…*").

Timeframe of questions. One purpose of running an ESM study is to reduce biases associated with retrospective reports, and instead, capture momentary experiences. However, a momentary report (i.e., *right now*) may not be appropriate for all experiences. Sometimes you might want to know about how someone was acting or what they were doing in the past hour, 30 mins, or 15 mins. So, some questions might be framed, "*In the past hour*, how *talkative* were you?", whereas others might be framed, "*Right now*, how *lively* do you feel?". If you have a mix of timeframes, it makes sense to first ask about immediate experiences before asking about experiences "in the past hour" (which require some retrospection).

Number of questions. As above—consider participant burden and how intensively and for how long you're sampling for. Pilot test this and time how long it takes to complete each report. If you can't put up with it, can you reasonably expect your participants to? :)

Nitty-gritty details. Do you want questions to appear in the same order, in a random order, or block-randomised by construct (e.g., always extraversion items followed by PA items, but individual questions appear in a random order)? What kind of scale will you use (e.g., 1–7, 0–10, 0–100)? Some of these decisions may be limited by software constraints, and I'm not sure if there's been much research on how much item order matters in this kind of study. But these are nevertheless decisions that need to be made.

Participant Motivation, Compliance, and Remuneration

Unlike a typical cross-sectional survey or one-off laboratory study, ESM studies ask participants to undergo minor interruptions to their everyday lives, several times each day, for several days. It is therefore important to offer sufficient incentives for participants to respond to as many surveys as *accurately* as possible.

Monetary incentives. An obvious incentive is to pay participants in cash or vouchers. Typically, this involves a flat payment at baseline, plus extra compensation depending on how many reports participants complete (i.e., compliance-contingent compensation). For example, you might offer participants an extra voucher or entry into a lottery draw for a substantial prize if they complete a minimum number of reports, or offer them extra cash or lottery entries for each report they complete. Of course, one concern is that this kind of incentivisation encourages participants to complete as many reports as possible, without necessarily guaranteeing high-quality responses. This is where non-monetary incentives and human elements come in.

Non-monetary incentives. People want to find out more about themselves. You can capitalise on this natural curiosity by offering participants additional feedback on themselves as a bonus for completing a sufficient number of reports. For example, I offered participants a wellbeing report based on their responses in the baseline survey (see Appendix). Another Honours student this year has generated a more comprehensive report which gave participants feedback on how they compared to norms for the Big Five and wellbeing measures, as well as the average emotions they reported that week. The beauty of including data from the actual ESM component is that it encourages participants to provide higher-quality responses, so that they can received more accurate information about themselves.

Staying in touch. Another strategy to maintain high compliance levels is to stay in regular contact with your participants and provide them with feedback on how many reports they've completed. In my study, I texted participants twice (usually around Days 3 & 5) with an update on the proportion reports they'd completed to date, along with an encouraging message (e.g., "Great job—keep it up!! 😀"). I also know of others who call their participants if they notice that their response rates drop. There are ways to automate the texts (or email updates), but otherwise, this is probably the most hands-on component of running an ESM study.

Valuing participants. Above all, be nice to your participants! Participants will provide more and better-quality responses if they feel like valued partners in research (and not just guinea pigs). So it's really important to maintain a positive, appreciative attitude in any interactions with participants—remember that they are doing you a favour by being willing to share information about their lives. Tell them how much you value their participation, and emphasise how important it is for the science and how much it will help you, the individual researcher, if they complete reports accurately and honestly.

Conner, T. S., & Lehman, B. (2012). Getting started: Launching a study in daily life. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 89–107). New York: Guilford Press.

Read closely. This is the most up-to-date practical guide that outlines each of the steps involved in designing a study of everyday life.

Green, A. S., Rafaeli, E., Bolger, N., Shrout, P. E., & Reis, H. T. (2006). Paper or plastic? Data equivalence in paper and electronic diaries. Psychological Methods, 11(1), 87–105. http://dx.doi.org/10.1037/1082-989X.11.1.87

Considers issues relating to partipant compliance and how to encourage this.

Preparing ESM Data

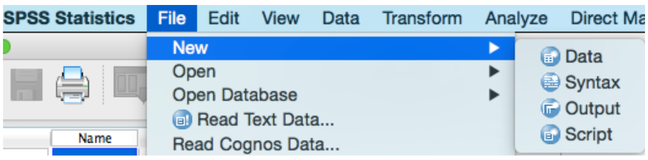
Data Cleaning

So you have your data—now what? If you have a baseline survey, ESM data, and a final survey (e.g., at the end of the week), you'll have three files to clean up. But to keep things simple for this tutorial, I'll just assume you have two files: baseline and ESM data.

First, a quick note on the joys of syntax for those unfamiliar with it. Syntax is far superior to the point-and-click interface in two respects:

1. Efficiency: Saves a lot of time when running repeated commands with multiple variables
2. Accurate documentation: Describes all preprocessing for anyone who might want to know (e.g., your supervisor, reviewers), and makes it easy to trace any errors back to the syntax

To open up a syntax file, just click File > New > Syntax.



All data files and syntax are available in the accompanying data supplement.

We'll be saving iterated versions of the dataset as we go, so the first thing you need to do is to specify the folder you want these data files to be saved to. Let's say you're working from a folder named "Folder" on your desktop. Here are the commands in PC and Mac forms, respectively:

```
cd 'C:\Desktop\Folder'.
cd '/Users/Jessie/Desktop/Folder'.
```

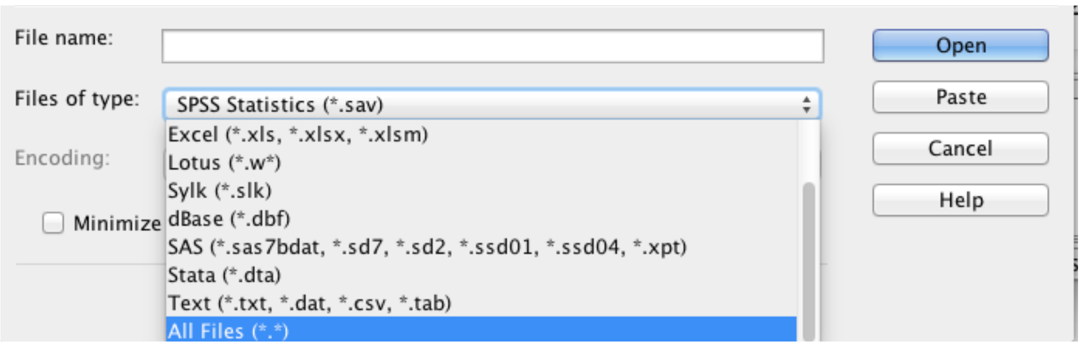
Now, we need to fix the datasets by identifying any **mistyped UserIDs** and **non-participants**. We can use the Match Files command to help us with this.

This is what your baseline file (EPA_ESM_1_Baseline.sav) should look like (stripped down, with the trait extraversion scale computed to keep things simple):

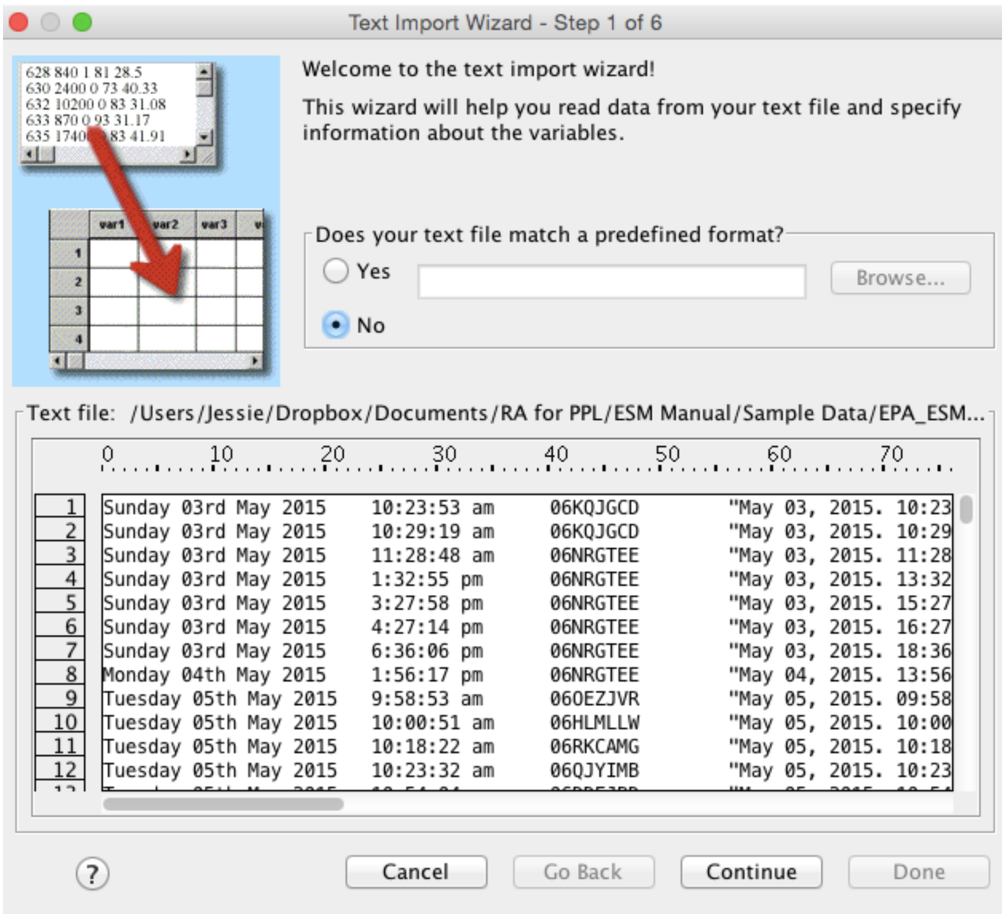
	surveystart	surveyend	id	extrav	Age	Gender
1	4-May-201...	4-May-2015 09:...	06XYJXJG	3.90	19.00	2
2	4-May-201...	4-May-2015 10:...	06VKQXVS	3.60	19.00	2
3	4-May-201...	4-May-2015 11:...	06HLMLLW	2.85	18.00	2
4	4-May-201...	4-May-2015 12:...	06OEZJVR	3.90	19.00	2
5	4-May-201...	4-May-2015 14:...	06QJYIMB	3.65	18.00	1
6	4-May-201...	4-May-2015 15:...	06TZZBDD	3.90	18.00	2
7	4-May-201...	4-May-2015 16:...	06FRPJSJ	3.05	19.00	1
8	4-May-201...	4-May-2015 17:...	06DDEJBD	3.45	21.00	2

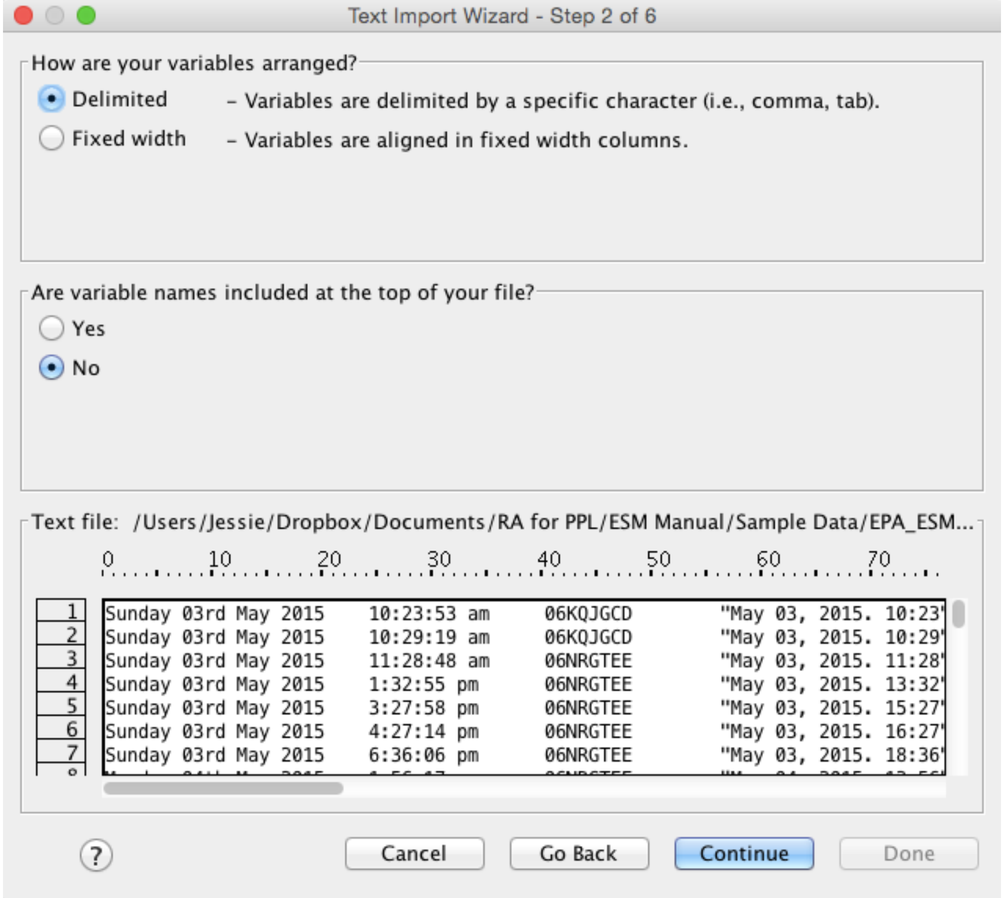
The UserID (`id`) is a unique participation code from the smartphone app. If you carefully supervised the participants while they entered their UserIDs into the survey in the lab session, and checked that they typed it correctly, there should be minimal typos in the baseline survey. However, if something goes wrong (e.g., the participant is having trouble downloading the app onto their phones), you could give participants a placeholder ID so that they can complete the survey (and download the app when they get home)—just make a note to replace the placeholder with the correct UserID (and ask them to send you their UserID from the app once it's on their phones).

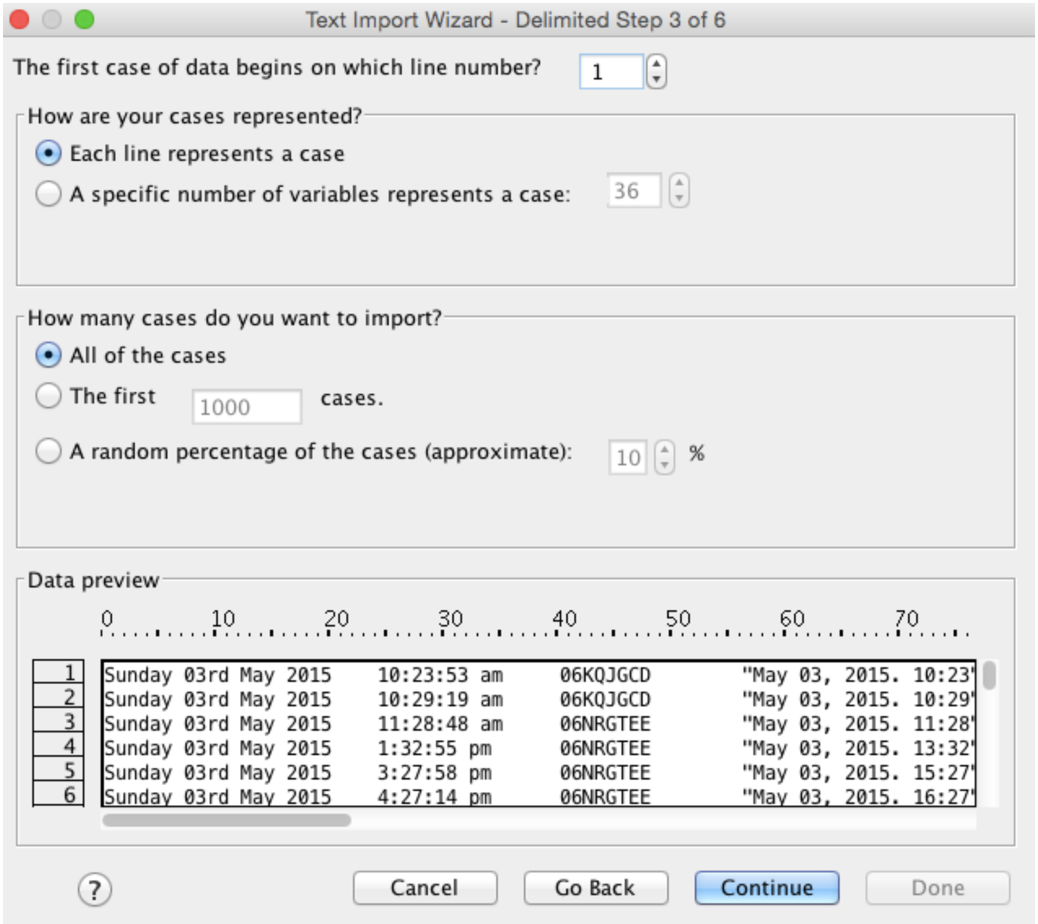
Your ESM data might come as a .txt file. In your example dataset (EPA_ESM_2_AppData.sav), I've imported it into SPSS format and relabelled the variables already. So, feel free to skip this next section for now, but in case you need it, this is how you import a .txt file: File > Open > Data…In this dialog box, change Files of type to "All Files". Now you'll be able to see and open the .txt file.

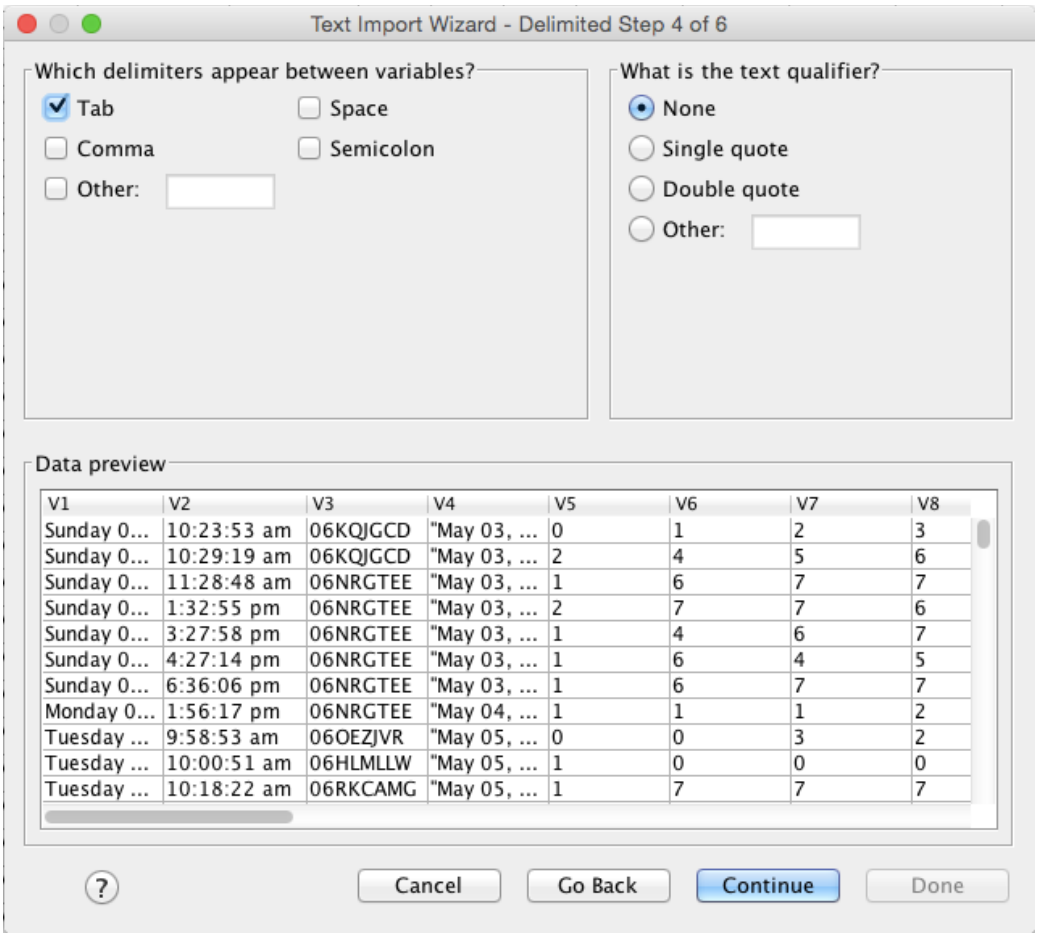


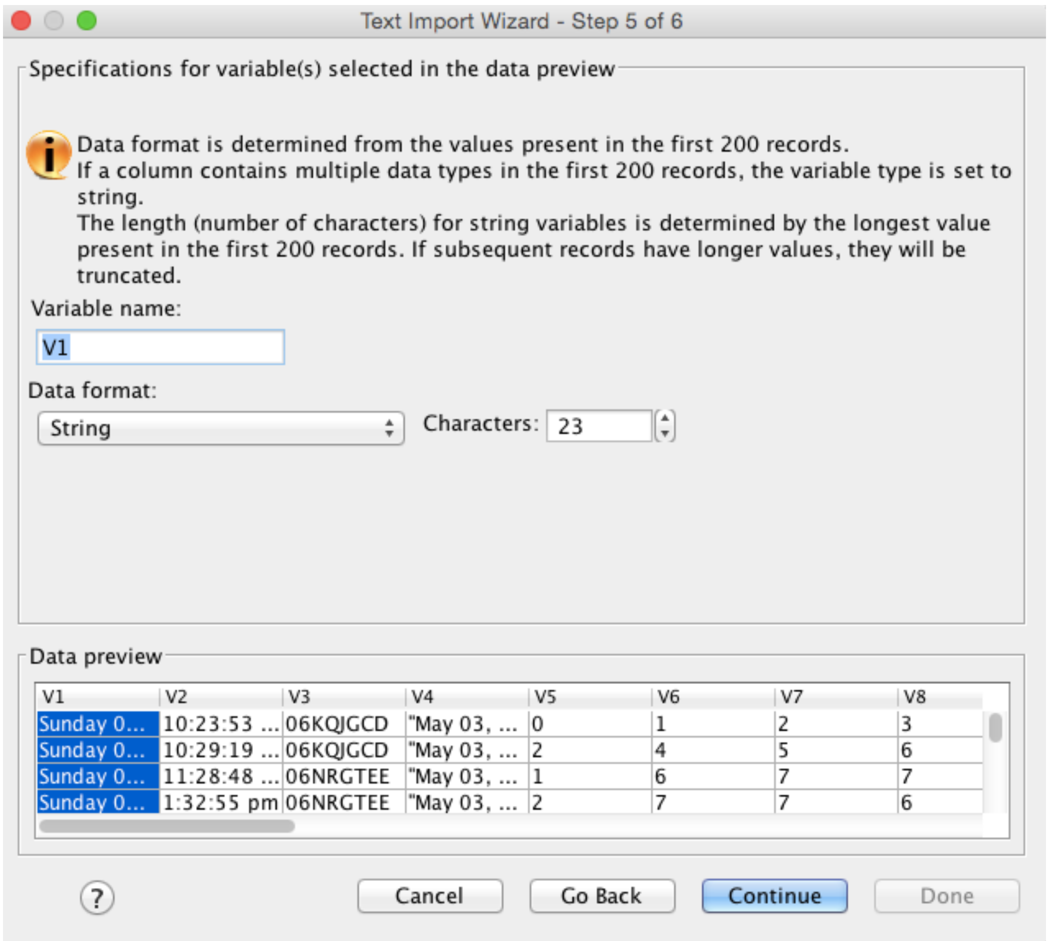
The text import wizard will help you get the data into SPSS format:

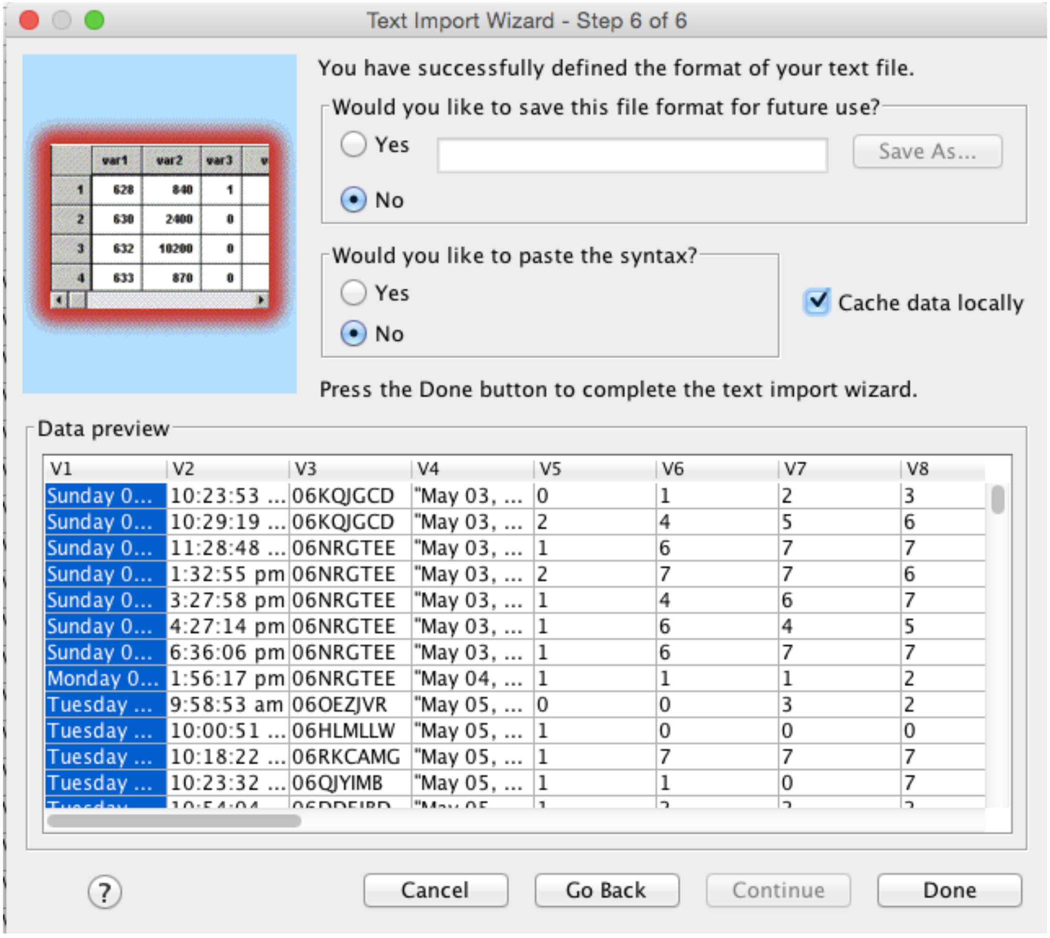



Text Import Wizard - Step 2 of 6
How are your variables arranged?
Delimited – Variables are delimited by a specific character (i.e., comma, tab).
Fixed width – Variables are aligned in fixed width columns.
Are variable names included at the top of your file?
Yes
No
Text file: /Users/Jessie/Dropbox/Documents/RA for PPL/ESM Manual/Sample Data/EPA_ESM...
0 10 20 30 40 50 60 70
1 Sunday 03rd May 2015 10:23:53 am 06KQJGCD "May 03, 2015. 10:23
2 Sunday 03rd May 2015 10:29:19 am 06KQJGCD "May 03, 2015. 10:29
3 Sunday 03rd May 2015 11:28:48 am 06NRGTEE "May 03, 2015. 11:28
4 Sunday 03rd May 2015 1:32:55 pm 06NRGTEE "May 03, 2015. 13:32
5 Sunday 03rd May 2015 3:27:58 pm 06NRGTEE "May 03, 2015. 15:27
6 Sunday 03rd May 2015 4:27:14 pm 06NRGTEE "May 03, 2015. 16:27
7 Sunday 03rd May 2015 6:36:06 pm 06NRGTEE "May 03, 2015. 18:36
?
Cancel
Go Back
Continue
Done


Text Import Wizard - Delimited Step 3 of 6
The first case of data begins on which line number? 1
How are your cases represented?
Each line represents a case
A specific number of variables represents a case: 36
How many cases do you want to import?
All of the cases
The first 1000 cases.
A random percentage of the cases (approximate): 10 %
Data preview
0 10 20 30 40 50 60 70
1 Sunday 03rd May 2015 10:23:53 am 06KQJGCD "May 03, 2015. 10:23
2 Sunday 03rd May 2015 10:29:19 am 06KQJGCD "May 03, 2015. 10:29
3 Sunday 03rd May 2015 11:28:48 am 06NRGTEE "May 03, 2015. 11:28
4 Sunday 03rd May 2015 1:32:55 pm 06NRGTEE "May 03, 2015. 13:32
5 Sunday 03rd May 2015 3:27:58 pm 06NRGTEE "May 03, 2015. 15:27
6 Sunday 03rd May 2015 4:27:14 pm 06NRGTEE "May 03, 2015. 16:27
?
Cancel
Go Back
Continue
Done


Text Import Wizard - Delimited Step 4 of 6
Which delimiters appear between variables?
Tab
Space
Comma
Semicolon
Other:
What is the text qualifier?
None
Single quote
Double quote
Other:
Data preview
V1 | V2 | V3 | V4 | V5 | V6 | V7 | V8
Sunday 0... | 10:23:53 am | 06KQJGCD | "May 03, ... | 0 | 1 | 2 | 3
Sunday 0... | 10:29:19 am | 06KQJGCD | "May 03, ... | 2 | 4 | 5 | 6
Sunday 0... | 11:28:48 am | 06NRGTEE | "May 03, ... | 1 | 6 | 7 | 7
Sunday 0... | 1:32:55 pm | 06NRGTEE | "May 03, ... | 2 | 7 | 7 | 6
Sunday 0... | 3:27:58 pm | 06NRGTEE | "May 03, ... | 1 | 4 | 6 | 7
Sunday 0... | 4:27:14 pm | 06NRGTEE | "May 03, ... | 1 | 6 | 4 | 5
Sunday 0... | 6:36:06 pm | 06NRGTEE | "May 03, ... | 1 | 6 | 7 | 7
Monday 0... | 1:56:17 pm | 06NRGTEE | "May 04, ... | 1 | 1 | 1 | 2
Tuesday ... | 9:58:53 am | 06OEZJVR | "May 05, ... | 0 | 0 | 3 | 2
Tuesday ... | 10:00:51 am | 06HLMLLW | "May 05, ... | 1 | 0 | 0 | 0
Tuesday ... | 10:18:22 am | 06RKCAMG | "May 05, ... | 1 | 7 | 7 | 7
?
Cancel
Go Back
Continue
Done


Text Import Wizard - Step 5 of 6
Specifications for variable(s) selected in the data preview
Data format is determined from the values present in the first 200 records.
If a column contains multiple data types in the first 200 records, the variable type is set to string.
The length (number of characters) for string variables is determined by the longest value present in the first 200 records. If subsequent records have longer values, they will be truncated.
Variable name:
V1
Data format:
String
Characters: 23
Data preview
V1 | V2 | V3 | V4 | V5 | V6 | V7 | V8
Sunday 0... | 10:23:53 ... | 06KQJGCD | "May 03, ... | 0 | 1 | 2 | 3
Sunday 0... | 10:29:19 ... | 06KQJGCD | "May 03, ... | 2 | 4 | 5 | 6
Sunday 0... | 11:28:48 ... | 06NRGTEE | "May 03, ... | 1 | 6 | 7 | 7
Sunday 0... | 1:32:55 pm | 06NRGTEE | "May 03, ... | 2 | 7 | 7 | 6
?
Cancel
Go Back
Continue
Done



When the data are imported, they come unlabelled—so you should have a separate spreadsheet where you note what order the variables are presented in the questionnaires.

V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V15
Sunday 03rd May 2015	10:23:53 am	06KQJGCD	"May 03, 2015. 10:23"	0	1	2	3	4	5	6	7	10
Sunday 03rd May 2015	10:29:19 am	06KQJGCD	"May 03, 2015. 10:29"	2	4	5	6	7	8	9	10	7
Sunday 03rd May 2015	11:28:48 am	06NRGTEE	"May 03, 2015. 11:28"	1	6	7	7	7	4	4	7	7
Sunday 03rd May 2015	1:32:55 pm	06NRGTEE	"May 03, 2015. 13:32"	2	7	7	6	3	8	2	7	8
Sunday 03rd May 2015	3:27:58 pm	06NRGTEE	"May 03, 2015. 15:27"	1	4	6	7	3	6	3	7	6
Sunday 03rd May 2015	4:27:14 pm	06NRGTEE	"May 03, 2015. 16:27"	1	6	4	5	8	4	6	3	3
Sunday 03rd May 2015	6:36:06 pm	06NRGTEE	"May 03, 2015. 18:36"	1	6	7	7	3	4	4	4	7
Monday 04th May 2015	1:56:17 pm	06NRGTEE	"May 04, 2015. 13:56"	1	1	1	2	10	2	7	0	1
Tuesday 05th May 2015	9:58:53 am	06OEZJVR	"May 05, 2015. 09:58"	0	0	3	2	10	1	9	5	3
Tuesday 05th May 2015	10:00:51 am	06HLMLLW	"May 05, 2015. 10:00"	1	0	0	0	10	0	0	6	7
Tuesday 05th May 2015	10:18:22 am	06RKCAMG	"May 05, 2015. 10:18"	1	7	7	7	2	8	3	8	8
Tuesday 05th May 2015	10:23:32 am	06QJYIMB	"May 05, 2015. 10:23"	1	1	0	7	3	2	8	2	3
Tuesday 05th May 2015	10:54:04 am	06DDEJBD	"May 05, 2015. 10:54"	1	2	2	2	7	2	7	4	2

The current ESM file (EPA_ESM_2_AppData.sav) is already in the right structure for Mplus: variables in columns, and **one row for each report** (the id variable denotes which participant the report is from, and will be used as the clustering variable in Mplus). This means that in this study, **every participant has up to 42 rows** (depending on how many reports they completed).

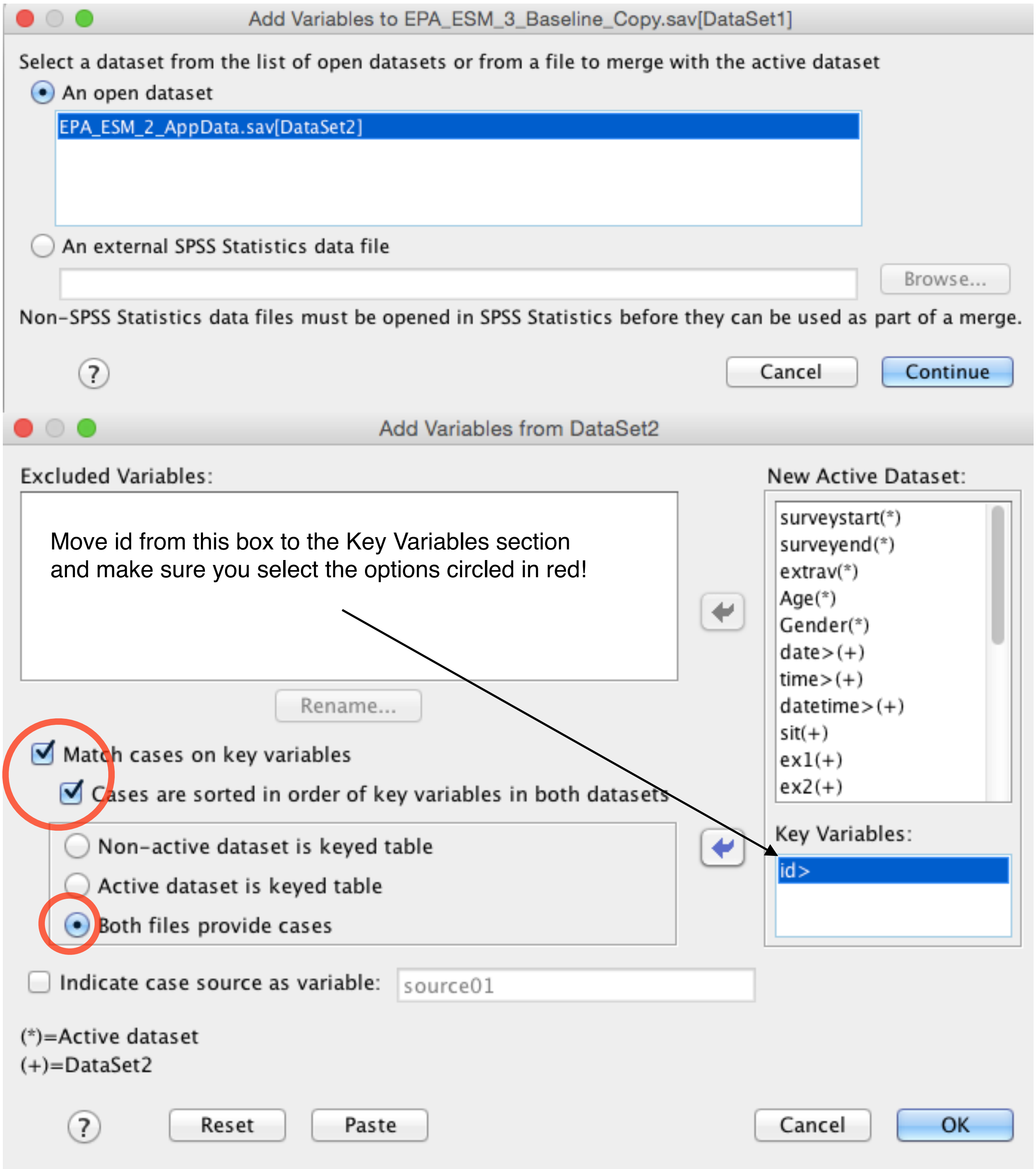
Let's now merge the two files to identify mistyped UserIDs. First, **make a copy of the baseline** file:

```
SAVE OUTFILE= 'EPA_ESM_3_Baseline_Copy.sav'
```

Have the baseline copy and ESM files both open. Now, sort both datasets by id (ascending), by running this syntax on both files:

```
SORT CASES BY id(a).
```

Now, working from the **baseline** file copy, go to Data > Merge Files > Add variables…



Now, sort one of the ESM variables ascending, e.g.:

```
SORT CASES BY sit(a).
```

You should get something like this:

	surveystart	surveyend	id	extrav	Age	Gender	date	time	datetime	sit	ex1	ex2	ex3
1	25-May-20...	25-May-2015 1...	111111	3.45	19.00	1				.	.	.	.
2	18-May-20...	18-May-2015 1...	123456	3.00	18.00	2				.	.	.	.
3	18-May-20...	18-May-2015 1...	D22Y7M	3.00	20.00	2				.	.	.	.
4	18-May-20...	18-May-2015 1...	H9PZ8F	4.00	18.00	1				.	.	.	.
5	11-May-20...	11-May-2015 1...	HTRD0S	3.00	29.00	2				.	.	.	.
6	18-May-20...	18-May-2015 1...	JKW6JO	3.90	18.00	2				.	.	.	.
7	18-May-20...	18-May-2015 1...	M0ET4I	3.30	20.00	2				.	.	.	.
8	3-Aug-201...	3-Aug-2015 10:...	anon	3.65	20.00	1				.	.	.	.
9	11-May-20...	11-May-2015 1...	X838RV	3.85	25.00	2				.	.	.	.
10	.	.	06AAYCZY	.	.	.	Wednesday 27th May 2...	4:13:03 pm	"May 27, 2015. 16:13"	0	3	4	4
11	.	.	06AAYCZY	.	.	.	Friday 29th May 2015	10:22:03 am	"May 29, 2015. 10:22"	0	3	3	3
12	.	.	06AAYCZY	.	.	.	Friday 29th May 2015	3:23:19 pm	"May 29, 2015. 15:23"	0	2	3	3
13	.	.	06AAYCZY	.	.	.	Saturday 30th May 2015	2:01:28 pm	"May 30, 2015. 11:33"	0	2	2	2

At the top, you can clearly see the baseline responses are not matched up with any ESM data. Note that the rest of the file isn't matched up properly either, but it's not supposed to be at this stage, so don't worry about it.

Remember that the correct UserID is the one that comes from the app. So, the baseline data that are unmatched with ESM data either did not complete any ESM surveys (perhaps due to technical issues; you should have kept a record of this during the study), or mistyped their UserIDs in the baseline survey. In our dataset, some of the unmatched baseline IDs are typos, a few are placeholders ("111111", "123456", "anon"), and some did not complete any reports due to technical difficulties (if you monitored participants closely during the study, you should already have a list of these UserIDs, ready for exclusion). Copy these UserIDs into a separate spreadsheet, for easy reference.

Now, go back to the ESM data to find the *correct* UserIDs. You want to look through the ID column to see if there are any close matches with your new list of incorrect baseline UserIDs. Run a Ctrl + F search using the first few characters of each incorrect UserID (e.g., "D22Y"), one at a time. When you find the correct UserID, add it to a second ("correct") column in your spreadsheet.

If the typo is serious (e.g., the participant enters their student number), then you can use additional pieces of information (e.g., date of survey completion, email address for compensation) to triangulate on who it is. To make this process easier, you should keep a record of UserIDs in a separate spreadsheet, along with the date of the participant's lab session. This will let you go back and only check the UserIDs for the relevant dates, if you have a lot of participants. And of course, for the placeholders, you should have already made a note in your participant management spreadsheet.

Once you've found the correct UserIDs, you need to fix errors in the **original** baseline file. You can open up the original file with this syntax:

```
GET
  FILE='EPA_ESM_1_Baseline.sav'.
DATASET NAME DataSet1 WINDOW=FRONT.
```

And then fix the incorrect UserIDs using a series of IF commands:

```
IF id = '111111' id = '06AAYCZY'.
IF id = '123456' id = '06OBABZQ'.
IF id = 'anon' id = 'PMQNPX'.
IF id = 'D8U0L8' id = '06YDZRFO'.
EXECUTE.
```

This is also a good point to mark participants who dropped out for technical or other reasons. You should keep a list of these UserIDs as you're running the study, so it's easy to just flag them when it comes to cleaning the data. In this study, these participants did not complete the study:

```
COMPUTE droppedbug = 0.
If id = 'HTRD0S' or id = 'X838RV' or id = 'M0ET4I' or id= 'D22Y7M'
or id = 'JKW6JO' or id = 'H9PZ8F' or id = 'MY4HSC' or id = '6RD4Q8'
droppedbug = 1.
EXECUTE.
```

You might also have other reasons to exclude participants (e.g., if they tell you that they didn't complete reports accurately):

```
COMPUTE badparticipant = 0.
If id = '210J95' badparticipant = 1.
EXECUTE.
```

Now compute a variable that marks reports as valid, and set reports from non-participants , dropouts or bad participants as invalid:

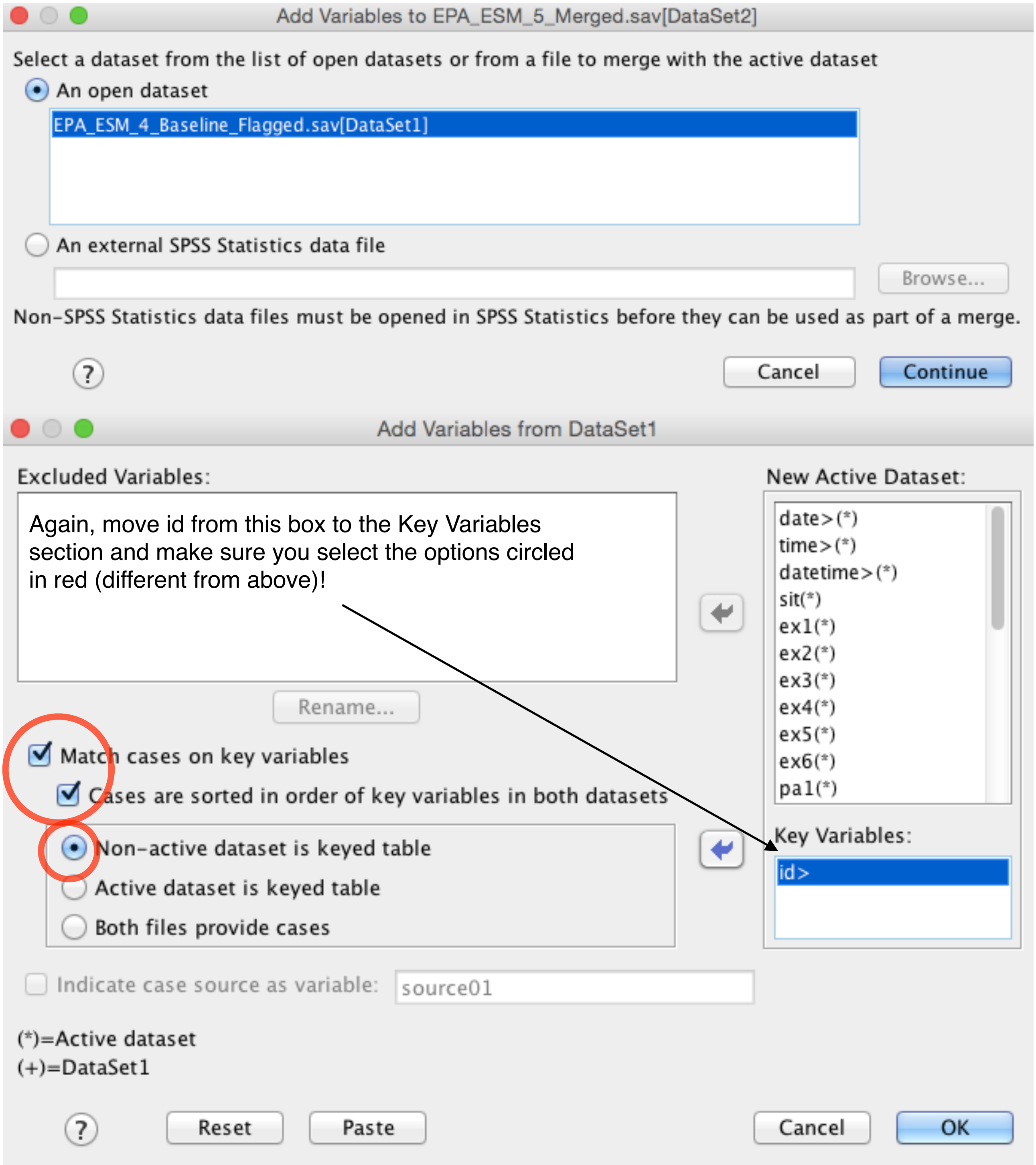
```
COMPUTE valid = 1.
If droppedbug = 1 or badparticipant = 1 valid = 0.
EXECUTE.
```

We've now fixed the UserIDs and flagged invalid participants in the baseline file. Save it as a corrected file (e.g., EPA_ESM_4_Baseline_Flagged.sav). You no longer need the incorrect merged file you just created (EPA_ESM_3_Baseline_Copy.sav). If you have a final survey file, you'll need to repeat each of these steps with that file too, to correct any UserID typos in that file (there will tend to be more typos compared to the baseline file, if participants are allowed to complete that survey outside of the lab). But I will leave the final survey data out of these examples, to keep things simple.

Now we need to identify any **non-participants** in the ESM data file. You might have extra responses from piloting the app, or from randoms who downloaded the app for fun. This time, **make a copy** of the ESM file (EPA_ESM_5_Merged.sav). This time, we'll be merging the copy of the ESM file with the **corrected** baseline file. Make sure both files are sorted in the order of UserID (ascending):

```
SORT CASES BY id(a).
```

Working from the **ESM** file copy, Data > Merge files > Add variables…



Now the ESM file will also contain the variables from the Baseline survey. I've moved the Baseline variables closer to the top of the file, so you can see what's happening. Sort one of the baseline variables by ascending:

```
SORT CASES BY age(a).
```

From this, you can easily see the 171 ESM responses that are not matched with any baseline data (and therefore not part of your study):
Get your spreadsheet out again and make a note of all of these unmatched ESM UserIDs. You can now flag them in the file, using syntax:

```
COMPUTE nobaseline = 0.
IF id = '03YAKC' or id = '06BGXCCF' or id = '06DZEFYM' or id =
'06FFYUPA' or id = '06FTOFZY' or id = '06IOTKBG'
or id = '06IUYGPU' or id = '06KVFOUK' or id = '06MSYWUJ' or id =
'06NRGTEE' or id = '06NUSVXU' or id = '06PUSSCL'
or id = '06RKCAMG' or id = '06SRKBPO' or id = '06YTACLY' or id =
'19XB39' or id = '738TQV' or id = 'FRIYWJ'
or id = 'INZ09U' or id = 'QHUX90' or id = 'SPCCG0' or id = 'YL18D7'
or id = '06KQJGCD' nobaseline = 1.
EXECUTE.
```

Flag the "nobaseline" responses as invalid, on the "valid" variable we created earlier:

```
IF nobaseline = 1 valid = 0.
EXECUTE.
```

At this stage, we have a dataset where nonparticipants, dropouts, and bad participants have been marked, and we have some descriptives on the total number of reports from participants who completed the study with no suspicions. Save this file, and then save it as a new copy:

```
SAVE OUTFILE='EPA_ESM_5_Merged.sav'.
SAVE OUTFILE='EPA_ESM_6_Excluded.sav'.
```

Working from this new file (EPA_ESM_6_Excluded.sav), let's get rid of the invalid reports to date. The SELECT IF command permanently deletes cases that don't meet certain conditions:

```
SELECT IF valid = 1.
EXECUTE.
```

At this stage, you'll want to compute some descriptives to summarise the total number of participants who completed the study, and the total number of reports they completed. First, filter out invalid responses, and sort the file by id (ascending):

```
FILTER BY valid.
SORT CASES BY id(a).
```

id	surveystart	surveyend	extrav	Age	Gender	date	time	datetime	sit
03YAKC	.	.	.	.	.	Monday 18th May 2015	3:04:33 am	"May 17, 2015. 12:04"	2
06BGXCCF	.	.	.	.	.	Friday 07th August 2015	3:38:12 am	"Aug 06, 2015. 13:38"	2
06BGXCCF	.	.	.	.	.	Friday 07th August 2015	5:13:58 am	"Aug 06, 2015. 15:13"	2
06BGXCCF	.	.	.	.	.	Friday 07th August 2015	7:14:06 am	"Aug 06, 2015. 17:14"	2
06BGXCCF	.	.	.	.	.	Saturday 08th August 20	4:36:06 am	"Aug 07, 2015. 14:36"	2
06DZEFYM	.	.	.	.	.	Friday 08th May 2015	9:51:43 am	"May 08, 2015. 09:51"	1
06DZEFYM	.	.	.	.	.	Friday 08th May 2015	11:53:16 am	"May 08, 2015. 11:53"	2
06DZEFYM	.	.	.	.	.	Friday 08th May 2015	5:18:45 pm	"May 08, 2015. 17:18"	2
06DZEFYM	.	.	.	.	.	Saturday 09th May 2015	2:14:13 pm	"May 09, 2015. 14:14"	2
06DZEFYM	.	.	.	.	.	Sunday 10th May 2015	11:34:04 am	"May 10, 2015. 11:33"	1
06DZEFYM	.	.	.	.	.	Sunday 10th May 2015	12:33:06 pm	"May 10, 2015. 12:33"	1
06DZEFYM	.	.	.	.	.	Sunday 10th May 2015	2:32:40 pm	"May 10, 2015. 14:32"	2
06DZEFYM	.	.	.	.	.	Sunday 10th May 2015	5:03:03 pm	"May 10, 2015. 17:03"	2
06DZEFYM	.	.	.	.	.	Monday 11th May 2015	6:05:06 pm	"May 11, 2015. 18:05"	2
06FFYUPA	.	.	.	.	.	Friday 17th July 2015	8:34:24 pm	"Jul 17, 2015. 18:26"	2
06FFYUPA	.	.	.	.	.	Friday 17th July 2015	8:36:52 pm	"Jul 17, 2015. 20:36"	1
06FFYUPA	.	.	.	.	.	Saturday 18th July 2015	9:04:50 am	"Jul 18, 2015. 09:04"	0
06FFYUPA	.	.	.	.	.	Saturday 18th July 2015	2:37:35 pm	"Jul 18, 2015. 14:37"	2
06FFYUPA	.	.	.	.	.	Saturday 18th July 2015	6:19:27 pm	"Jul 18, 2015. 14:37"	2
06FFYUPA	.	.	.	.	.	Saturday 18th July 2015	7:05:03 pm	"Jul 18, 2015. 19:05"	2
06FFYUPA	.	.	.	.	.	Sunday 19th July 2015	10:12:33 pm	"Jul 19, 2015. 15:55"	2
06FFYUPA	.	.	.	.	.	Monday 20th July 2015	2:46:04 pm	"Jul 20, 2015. 14:46"	0
06FFYUPA	.	.	.	.	.	Tuesday 21st July 2015	10:00:28 am	"Jul 21, 2015. 10:00"	0
06FFYUPA	.	.	.	.	.	Tuesday 21st July 2015	11:53:14 am	"Jul 21, 2015. 11:53"	1
06FFYUPA	.	.	.	.	.	Tuesday 21st July 2015	3:52:04 pm	"Jul 21, 2015. 15:52"	2
06FFYUPA	.	.	.	.	.	Wednesday 22nd July 201	9:35:03 am	"Jul 21, 2015. 19:23"	2
06FFYUPA	.	.	.	.	.	Wednesday 22nd July 201	9:36:34 am	"Jul 22, 2015. 09:36"	2

Now, use this command to identify duplicate cases of the id variable (line 3). The total frequency corresponds to the total valid response count, whereas the number of primary cases tells you how many valid participants there are.

```
MATCH FILES
  /FILE=*
  /BY id
  /LAST=PrimaryLast.
VARIABLE LABELS  PrimaryLast 'Indicator of each last matching case
as Primary'.
VALUE LABELS  PrimaryLast 0 'Duplicate Case' 1 'Primary Case'.
VARIABLE LEVEL  PrimaryLast (ORDINAL).
FREQUENCIES VARIABLES=PrimaryLast.
EXECUTE.
```

Indicator of each last matching case as Primary

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Duplicate Case	1867	96.5	96.5	96.5
	Primary Case	67	3.5	3.5	100.0
	Total	1934	100.0	100.0	

This tells us that we have 67 participants who completed a total of 1934 reports.

Now, we need to screen individual reports for problems. Before you look at the data, it's good practice to define your exclusion criteria, *a priori*, given recent concerns about research integrity and researcher degrees of freedom (see Simmons, Nelson, & Simonsohn, 2012). Note that I explored the effect of making the criteria more stringent or loose in my own data, and found that it doesn't change the results substantially. Even so, it's good scientific practice to reduce noise and ensure that your final data is as valid as possible, based on criteria that you can justify.

One sign of non-valid responding is if reports consist of too many identical responses. For example, if a report consists of all "10"s (or any other value), that is suspicious. You should remove reports that consist of 80-90% (you can set the threshold) identical responses. For this dataset, we have 14 items (not including the situational variable), and if we set a threshold of 85%, we plan to exclude reports with 12 or more identical responses.

This syntax counts the number of 0's, 1's, 2's, etc., in each report:

```
COUNT NumZeros=ex1 to cont3 (0).
COUNT NumOnes=ex1 to cont3 (1).
COUNT NumTwos=ex1 to cont3 (2).
COUNT NumThrees=ex1 to cont3 (3).
COUNT NumFours=ex1 to cont3 (4).
COUNT NumFives=ex1 to cont3 (5).
COUNT NumSixes=ex1 to cont3 (6).
COUNT NumSevens=ex1 to cont3 (7).
COUNT NumEights=ex1 to cont3 (8).
COUNT NumNines=ex1 to cont3 (9).
COUNT NumTens=ex1 to cont3 (10).
EXECUTE.
```

This syntax flags reports that have 12 or more (i.e., > 85%) identical responses (change this number depending on the number of items in your study):

```
COMPUTE tooidentical=0.
IF NumZeros > 11
OR NumOnes > 11
OR NumTwos > 11
```

```
OR NumThrees > 11
OR NumFours > 11
OR NumFives > 11
OR NumSixes > 11
OR NumSevens > 11
OR NumEights > 11
OR NumNines > 11
OR NumTens > 11 tooidentical=1.
EXECUTE.
```

Now flag tooidentical reports as invalid, count how many there are (there are 31), and filter them out:

```
If tooidentical = 1 valid = 0.
EXECUTE.

FREQUENCIES tooidentical.

FILTER BY valid.
```

You may also set other criteria for excluding individual questions and individual reports if they are answered too quickly. For example, you might choose to remove individual items and entire reports where more than 50% of items are answered too quickly (< 500 ms), or remove reports that are completed too quickly. I will not discuss this further as the InstantSurvey app we used did not capture timing data, but see McCabe et al. (2012) for details.

After marking invalid reports, you need to remove participants with an insufficient number of valid reports. This criterion is a bit more subjective; we want enough reports to be able to detect correlations between pairs of variables within each participant, but don't want to set it so high that we lose too many participants. I chose to exclude participants with fewer than 15 valid reports.

This command creates a new variable with each participant's final valid response count (the valid filter should already be on at this stage, but if not, run it again):

```
AGGREGATE
  /OUTFILE=* MODE=ADDVARIABLES
  /BREAK=id
  /responsecount=N(id).
```

Now, create another variable that marks participants who have too few valid responses:

```
Compute toofewtotal=0.
If responsecount < 15 toofewtotal=1.
EXECUTE.
```

Finally, we'll mark these participants as invalid:

```
If toofewtotal = 1 valid = 0.
EXECUTE.
```

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
responsecount	62	15	42	29.69	5.933
Valid N (listwise)	62				

At this stage, save your file as a backup with exclusions flagged but not deleted, and then save a new version of that same file:

```
USE ALL.
SAVE OUTFILE='EPA_ESM_6_Excluded.sav'.
SAVE OUTFILE='EPA_ESM_7_Valid.sav'.
```

And use the Select If command to remove invalid cases:

```
SELECT IF valid = 1.
EXECUTE.
```

Once you've removed invalid individual reports and participants with insufficient reports, compute your updated descriptives using the Filter and Duplicate Case commands:

```
DELETE VARIABLES PrimaryLast.
FILTER BY valid.
SORT CASES BY id(a).

MATCH FILES
  /FILE=*
  /BY id
  /LAST=PrimaryLast.
VARIABLE LABELS  PrimaryLast 'Indicator of each last matching case
as Primary'.
VALUE LABELS  PrimaryLast 0 'Duplicate Case' 1 'Primary Case'.
VARIABLE LEVEL  PrimaryLast (ORDINAL).
FREQUENCIES VARIABLES=PrimaryLast.
EXECUTE.
```

So, the final data comprise 1841 valid reports from 62 participants.

Indicator of each last matching case as Primary

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Duplicate Case	1779	96.6	96.6	96.6
	Primary Case	62	3.4	3.4	100.0
	Total	1841	100.0	100.0	

To get the average number of reports completed, and the standard deviation, filter by PrimaryLast (the variable created by the Match Files command just used, to determine the number of participants and reports), then run descriptives on the responsecount variable:

```
FILTER BY primarylast.
DESCRIPTIVES responsecount.
```

And get descriptives on the demographics of your final analysed sample:

```
DESCRIPTIVES age.
FREQUENCIES gender.
```

Finally, save the current file and then save a copy.

```
USE ALL.
SAVE OUTFILE='EPA_ESM_7_Valid.sav'.
SAVE OUTFILE='EPA_ESM_8_Final.sav'.
```

Here's an example of how you might report the data cleaning process (in the Method section):

Participants

Participants (*N* = 76) were recruited via flyers posted around the University of Melbourne … Of the initial sample of 76 participants, 14 were excluded due to technical and data quality issues described below. The final analysed sample (*N* = 62) comprised 39 (62.9%) females and 23 (37.1%) males, aged between 18 and 33 years (M_{age} = 21.40, SD_{age} = 3.55).

Procedures and Design

… Due to software incompatibility with some smartphones, eight participants were unable to complete the ESM component. One additional participant revealed that he consistently responded inaccurately as his smartphone froze during reports. After excluding data from these participants, the remaining 67 participants completed 1934 out of 2814 (68.73%) possible reports; a typical response rate for ESM studies (Fleeson & Gallagher, 2009). Of these reports, 31 reports that contained more than 85% identical responses (e.g., responding “10” for nearly all questions) were excluded. Finally, 5 participants who completed fewer than 15 valid reports were excluded, eliminating a further 62 reports. The final data retained for analysis comprised 62 participants who completed 1841 out of 2562 (72%) possible reports (M_{reports} = 29.69, SD_{reports} = 5.93).

McCabe, K. O., Mack, L., & Fleeson, W. (2012). A guide for data cleaning in experience sampling studies. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 321–338). New York: Guilford Press.

The data cleaning method described above was adapted from this chapter. Check it out for more data cleaning tips, e.g., reports that are completed too quickly, if you have that data.

Computing and Centring Variables

Now that we've cleaned up issues in the data, it's time to compute scales and centre variables (see Enders & Tofighi, 2007 for a discussion of when to centre), working from the file you just saved (EPA_ESM_8_Final.sav).

If you have reverse-scored items, you need to recode them:

```
RECODE
ex4 ex6 (0=10) (1=9) (2=8) (3=7) (4=6) (5=5) (6=4) (7=3) (8=2)
(9=1) (10=0) INTO ex4r ex6r.
EXECUTE.
```

Then compute the scales:

```
COMPUTE e = mean (ex1, ex2, ex3, ex4r, ex5, ex6r).
COMPUTE pa = mean (pa1, pa2, pa3).
COMPUTE cont = mean (cont1, cont2, cont3).
COMPUTE pow = mean (pow1, pow2).
EXECUTE.
```

Now we need to centre the state variables around each participant's mean, to remove trait influences.

The following syntax creates aggregate variables (eav and paav), which is each participant's average level of extraverted and PA across all reports.

```
SORT CASES BY id(A).
AGGREGATE
  /OUTFILE=* MODE=ADDVARIABLES OVERWRITE=YES
  /BREAK=id
  /eav=MEAN(e)
  /paav=MEAN(pa)
  /contav=MEAN(cont)
  /powav=MEAN(pow).
```

Now that we have aggregate variables, we can compute state variables, centred around each participant's mean:

```
COMPUTE est=e-eav.
COMPUTE past=pa-paav.
COMPUTE contst=cont-contav.
COMPUTE powst=pow-powav.
EXECUTE.
```

In other words, `est` represents the momentary deviation from the participant's usual levels of extraversion. If it is positive, then the participant is acting relatively extraverted, compared to the average of levels of extraverted behaviour they reported across the study.

We also need to compute the effect coded variables for the situation variable, which we'll be including as a covariate. Effect coding is like dummy coding, but the reference category is coded as -1 rather than 0 (for more info, see http://www.ats.ucla.edu/stat/mult_pkg/faq/general/effect.htm). The situation variable was coded 0 = alone, 1 = semi-social, and 2 = social. We'll create variables for social and semi-social, using alone as the reference category.

```
Compute social=0.
if sit=2 social=1.
if sit=0 social=-1.
execute.

Compute semi=0.
if sit=1 semi=1.
if sit=0 semi=-1.
execute.
```

Dates and Timecourse

As the date and time information is currently in string format, we need to find a way to convert that into a numerical format. We also want to create a variable that gives us information about how far along the participant is in the study; in this case, the amount of time elapsed since the first report. For these two steps, Excel seems to do the job effectively, so bear with the slight clunkiness for the next few pages. I've included the spreadsheet with the relevant formulae (DateTime) to make things easier.

First, copy over the column of data containing the "datetime" variable into a new column in an Excel spreadsheet. Label the column for your own convenience (I've called it "stringdt" in the spreadsheet"):

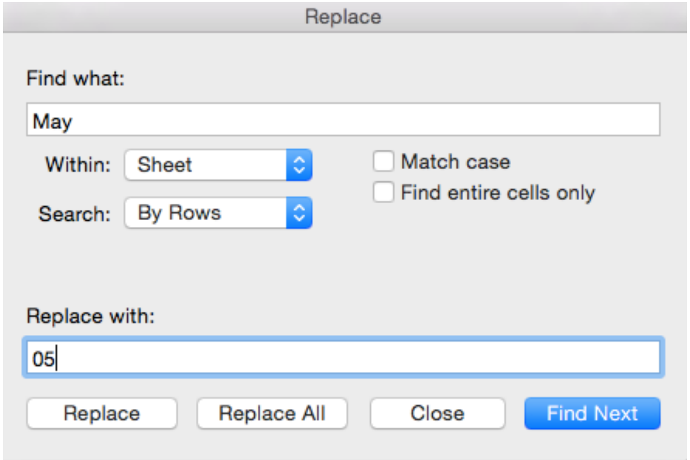
datetime
"May 26, 2015. 11:02"
"May 26, 2015. 13:02"
"May 26, 2015. 19:04"
"May 26, 2015. 19:04"
"May 27, 2015. 16:13"
"May 28, 2015. 10:43"
"May 28, 2015. 15:41"
"May 28, 2015. 17:40"
"May 28, 2015. 21:44"
"May 29, 2015. 10:22"
"May 29, 2015. 15:23"
"May 29, 2015. 20:22"
"May 30, 2015. 11:33"
"May 30, 2015. 15:01"

1. Variable in SPSS

	A
1	stringdt
2	May 26, 2015. 11:02
3	May 26, 2015. 13:02
4	May 26, 2015. 19:04
5	May 26, 2015. 19:04
6	May 27, 2015. 16:13
7	May 28, 2015. 10:43
8	May 28, 2015. 15:41
9	May 28, 2015. 17:40
10	May 28, 2015. 21:44
11	May 29, 2015. 10:22
12	May 29, 2015. 15:23
13	May 29, 2015. 20:22
14	May 30, 2015. 11:33
15	May 30, 2015. 15:01
16	May 30, 2015. 17:09
17	May 30, 2015. 19:30

2. Variable in Excel

Now, we want to replace the text months with numbers. So, run a find & replace all for each month (May = 05, June = 06, etc.):

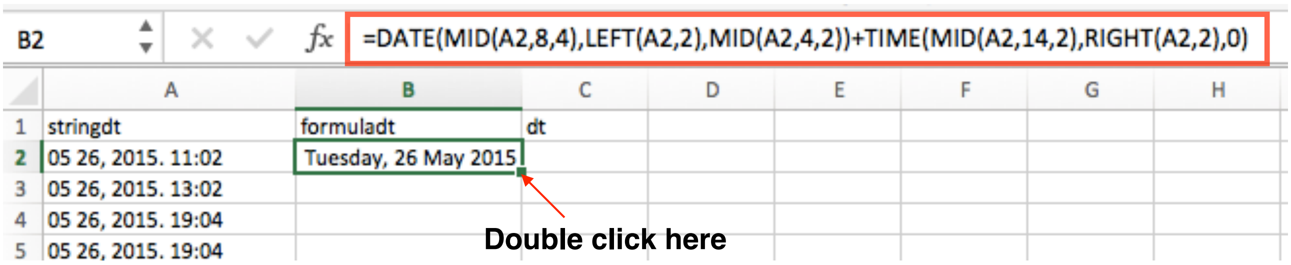


The resulting data should look like this:

	A
1	stringdt
2	05 26, 2015. 11:02
3	05 26, 2015. 13:02
4	05 26, 2015. 19:04
5	05 26, 2015. 19:04
6	05 27, 2015. 16:13
7	05 28, 2015. 10:43
8	05 28, 2015. 15:41
9	05 28, 2015. 17:40
10	05 28, 2015. 21:44
11	05 29, 2015. 10:22
12	05 29, 2015. 15:23
13	05 29, 2015. 20:22
14	05 30, 2015. 11:33

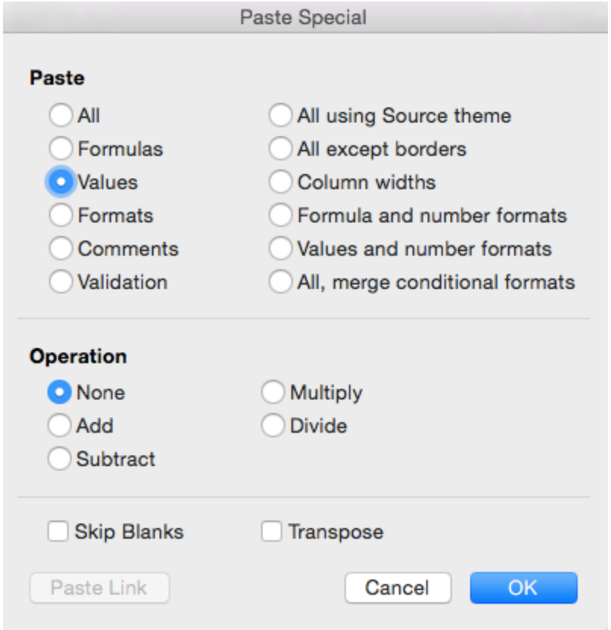
Now, we want to extract the date-time information from these currently meaningless numbers. The formula shown in the screenshot below (in the red box) tells Excel that the first two numbers correspond to the number, the next two correspond to the day, and so on. You'll need to modify the formula if your string date-time variable comes in a different format. For more details on how this works, see http://superuser.com/questions/274494/convert-text-string-to-date-time-format (or Google "convert string to date time Excel" for more guides).

You should get something like this, but it it comes out in a slightly different format (e.g., 26/05/2015), that's also not a problem. The main point is that Excel has information about the date and time.



Double-click on the bottom-right corner of that cell to let Excel automatically fill in that formula down the column.

Now, click on B to select the entire column, and then copy that column (Ctrl+C). Then, right-click on C to select column C, and select the "Paste Special > Values" option.



This is what the data in that column (which I've labelled "dt") should now look like:

	A	B	C
1	stringdt	formuladt	dt
2	05 26, 2015. 11:02	Tuesday, 26 May 2015	42150.4597
3	05 26, 2015. 13:02	Tuesday, 26 May 2015	42150.5431
4	05 26, 2015. 19:04	Tuesday, 26 May 2015	42150.7944
5	05 26, 2015. 19:04	Tuesday, 26 May 2015	42150.7944
6	05 27, 2015. 16:13	Wednesday, 27 May 2015	42151.6757
7	05 28, 2015. 10:43	Thursday, 28 May 2015	42152.4465
8	05 28, 2015. 15:41	Thursday, 28 May 2015	42152.6535
9	05 28, 2015. 17:40	Thursday, 28 May 2015	42152.7361
10	05 28, 2015. 21:44	Thursday, 28 May 2015	42152.9056
11	05 29, 2015. 10:22	Friday, 29 May 2015	42153.4319
12	05 29, 2015. 15:23	Friday, 29 May 2015	42153.641
13	05 29, 2015. 20:22	Friday, 29 May 2015	42153.8486

This is now in the Excel's "General" time format. This counts the number of days since January 0, 1900. Each 1 unit increment represents 1 day, or 24 hours. So, for example, 6 hours would = 0.25.

Now, go back to to SPSS and create a new variable:

```
COMPUTE dt = 0.
EXECUTE.
```

dt
.00
.00
.00
.00
.00
.00
.00
.00
.00
.00

dt
42150.46
42150.54
42150.79
42150.79
42151.68
42152.45
42152.65
42152.74
42152.91

And manually copy column C from the Excel spreadsheet straight into this SPSS variable. Of course, for the date-time information to be matched correctly to each report, I'm assuming that you haven't played around the SPSS dataset and changed the order of any reports since you started the Excel process.

Now, for the second part, we need to sort the SPSS dataset by `id` and `dt`. In other words, we want reports to be separated by participants, and in the order they completed them (e.g., 26 reports from participant A, beginning with their first report, followed by 28 reports from participant B, beginning with their first report):

```
SORT CASES BY id(A) dt(A).
```

This time, open a new sheet in your DateTime file, and copy over the **id** and **dt** variables **from the sorted SPSS dataset** into the Excel spreadsheet:

Now, we're going to apply another Excel formula (in the red box) to the second cell in a new column (which we'll call "startdt"), to extract the time of the first report for each participant. Again, right click on the bottom right corner to fill down. You should get this:

A	B
id	dt
06AAYCZY	42150.46
06AAYCZY	42150.54
06AAYCZY	42150.79
06AAYCZY	42150.79
06AAYCZY	42151.68
06AAYCZY	42152.45
06AAYCZY	42152.65
06AAYCZY	42152.74
06AAYCZY	42152.91
06AAYCZY	42153.43
06AAYCZY	42153.64
06AAYCZY	42153.85
06AAYCZY	42154.48
06AAYCZY	42154.63
06AAYCZY	42154.71
06AAYCZY	42154.81
06AAYCZY	42155.39
06AAYCZY	42155.5
06AAYCZY	42155.6
06AAYCZY	42155.68
06AAYCZY	42155.81

C2 $\quad fx$ =VLOOKUP(A2,$A$2:$B$1842,2,FALSE)

	A	B	C	D	E
1	id	dt	startdt		
2	06AAYCZY	42150.46	42150.46		
3	06AAYCZY	42150.54	42150.46		
4	06AAYCZY	42150.79	42150.46		
5	06AAYCZY	42150.79	42150.46		
6	06AAYCZY	42151.68	42150.46		
7	06AAYCZY	42152.45	42150.46		
8	06AAYCZY	42152.65	42150.46		
9	06AAYCZY	42152.74	42150.46		
10	06AAYCZY	42152.91	42150.46		
11	06AAYCZY	42153.43	42150.46		
12	06AAYCZY	42153.64	42150.46		
13	06AAYCZY	42153.85	42150.46		
14	06AAYCZY	42154.48	42150.46		
15	06AAYCZY	42154.63	42150.46		
16	06AAYCZY	42154.71	42150.46		
17	06AAYCZY	42154.81	42150.46		
18	06AAYCZY	42155.39	42150.46		
19	06AAYCZY	42155.5	42150.46		
20	06AAYCZY	42155.6	42150.46		
21	06AAYCZY	42155.68	42150.46		
22	06AAYCZY	42155.81	42150.46		
23	06AAYCZY	42156.43	42150.46		
24	06AAYCZY	42156.51	42150.46		
25	06AAYCZY	42156.59	42150.46		
26	06AAYCZY	42156.7	42150.46		
27	06AAYCZY	42156.78	42150.46		
28	06AESNCZ	42143.52	42143.52		
29	06AESNCZ	42143.62	42143.52		
30	06AESNCZ	42143.72	42143.52		

In the formula, A2 = UserID, $A$2:$B$1842 corresponds to the range of data that we're using (so, replace "1842" with however many rows you have in that spreadsheet for your data), 2 = column of data being looked up (i.e., the column with dt), and FALSE = exact match only. For more information on how VLOOKUP works, see https://support.office.com/en-us/article/VLOOKUP-What-it-is-and-when-to-use-it-5984e27b-4f0d-431e-83b1-7ab062c75493

Now, let’s go back to our SPSS dataset. Compute a new variable for the start date/time:

```
COMPUTE startdt = 0.
EXECUTE.
```

And as above, manually copy+paste column C from your Excel spreadsheet into this SPSS variable.

Now, we can compute the amount of time elapsed since the start of the study for each participant by subtracting the start time from the time of the report:

```
COMPUTE timeel = dt-startdt.
EXECUTE.
```

Keep this raw variable, but let’s also create a standardised version to use as a covariate (apparently Mplus works best when variables are standardised between -3 & + 3). This creates a Z-scored version of the variable (“Ztimeel”), standardising across all reports and participants:

```
DESCRIPTIVES timeel /SAVE.
```

Let’s also standardise the trait extraversion variable to aid interpretability. Save this file, and then save a new file with just level 2 variables:

```
SAVE OUTFILE='EPA_ESM_8_Final.sav'.
SAVE OUTFILE='EPA_ESM_9_Level2.sav'.

SELECT IF PrimaryLast = 1.
EXECUTE.
```

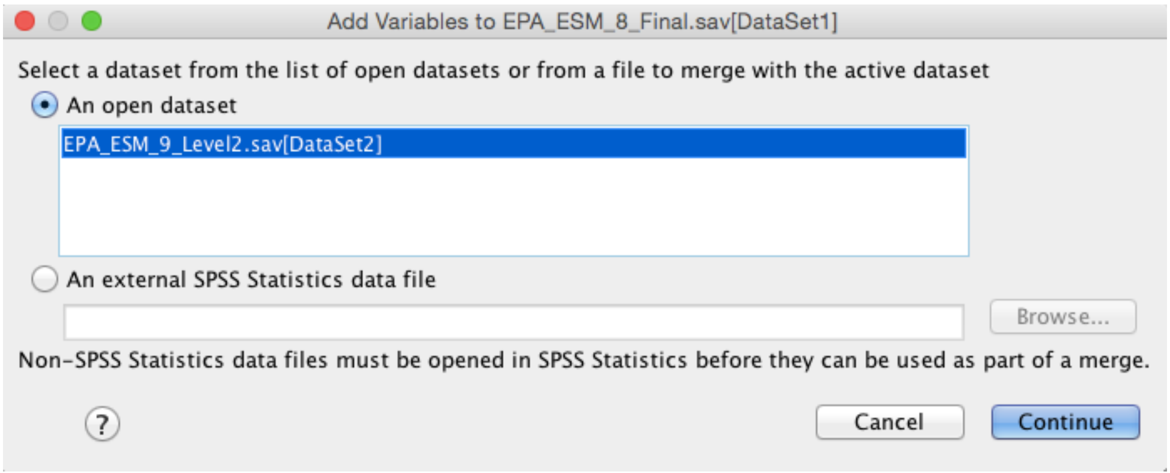
Now standardise trait extraversion (and any other Level 2 variables you want to use in the Mplus analysis):

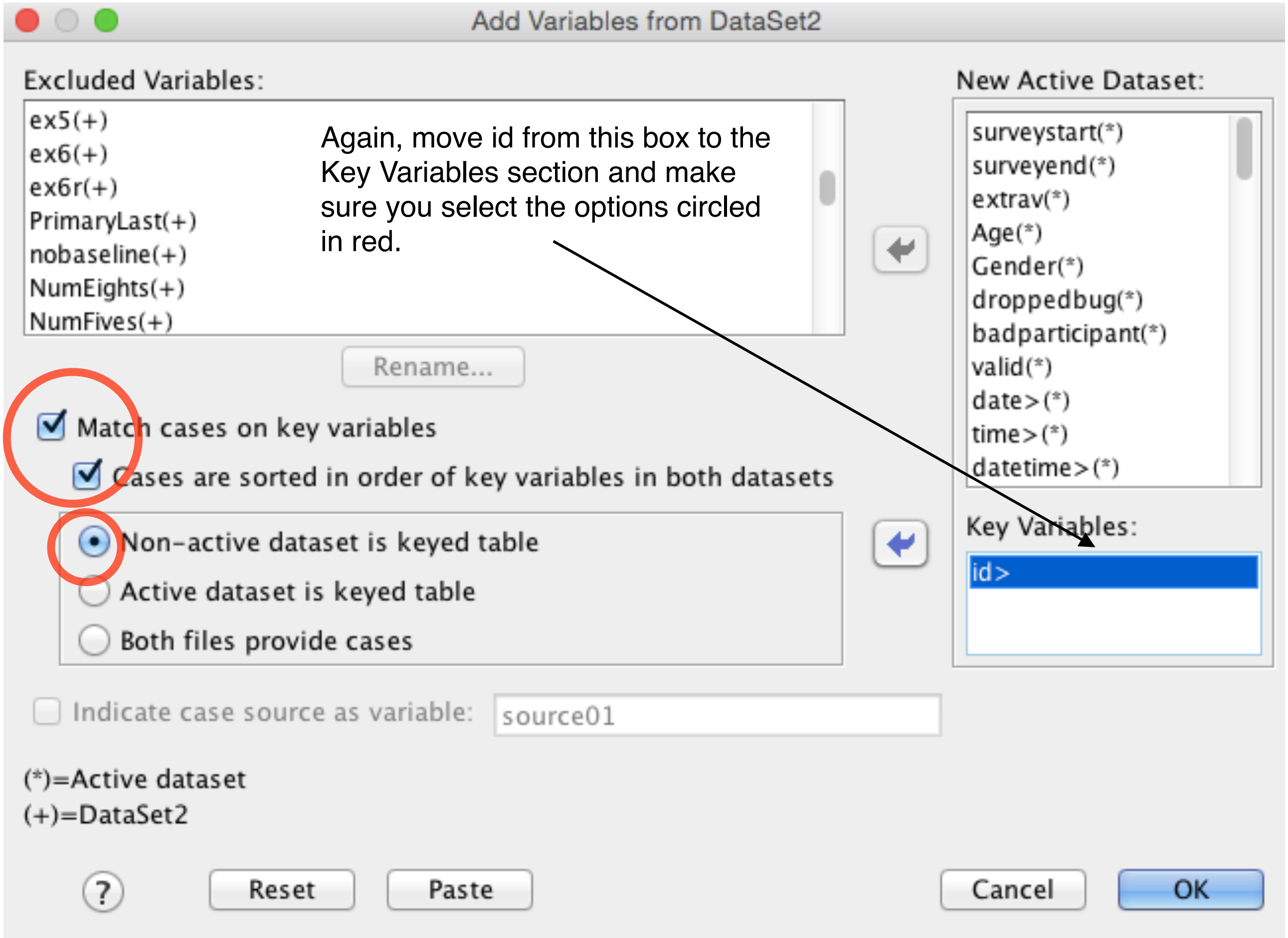
```
DESCRIPTIVES extrav /SAVE.
```

Open up the final file again and sort both files:

```
GET
  FILE='EPA_ESM_8_Final.sav'.
DATASET NAME DataSet1 WINDOW=FRONT.


DATASET ACTIVATE DataSet1.
SORT CASES BY id(a).
DATASET ACTIVATE DataSet2.
SORT CASES BY id(a).
```

Activate the final data file and merge files (Data > Merge Files > Add Variables) to disaggregate the new centred variable, `zextrav`, to the final dataset:





We're now nearly ready to create the data file that we'll use for the Mplus analyses. The final necessary step is to convert the string IDs into numbers. Working from the final dataset, let's rename that variable in case we need it for any reason:

```
RENAME VARIABLES id = stringid.
```

Now, we need to index the observations for each person. This will number each report for each participant, where 1 = a participant's first report:

```
COMPUTE obs = 1.
IF stringid = lag(stringid) obs = lag(obs) + 1.
EXECUTE.
```

Finally, this syntax will use that information to create a numerical id:

```
COMPUTE #x = #x + 1.
IF obs ne 1 #x = lag(#x).
COMPUTE id = #x.
EXECUTE.
```

Drag the new obs and numerical id variables next to the stringid variable to visually check that it worked. As you can see on the right, that looks perfect, so I've saved this final file again:

stringid	obs	id
06AAYCZY	1.00	1.00
06AAYCZY	2.00	1.00
06AAYCZY	3.00	1.00
06AAYCZY	4.00	1.00
06AAYCZY	5.00	1.00
06AAYCZY	6.00	1.00
06AAYCZY	7.00	1.00
06AAYCZY	8.00	1.00
06AAYCZY	9.00	1.00
06AAYCZY	10.00	1.00
06AAYCZY	11.00	1.00
06AAYCZY	12.00	1.00
06AAYCZY	13.00	1.00
06AAYCZY	14.00	1.00
06AAYCZY	15.00	1.00
06AAYCZY	16.00	1.00
06AAYCZY	17.00	1.00
06AAYCZY	18.00	1.00
06AAYCZY	19.00	1.00
06AAYCZY	20.00	1.00
06AAYCZY	21.00	1.00
06AAYCZY	22.00	1.00
06AAYCZY	23.00	1.00
06AAYCZY	24.00	1.00
06AAYCZY	25.00	1.00
06AAYCZY	26.00	1.00
06AESNCZ	1.00	2.00
06AESNCZ	2.00	2.00

```
SAVE OUTFILE='EPA_ESM_8_Final.sav'.
```

Mplus Data File

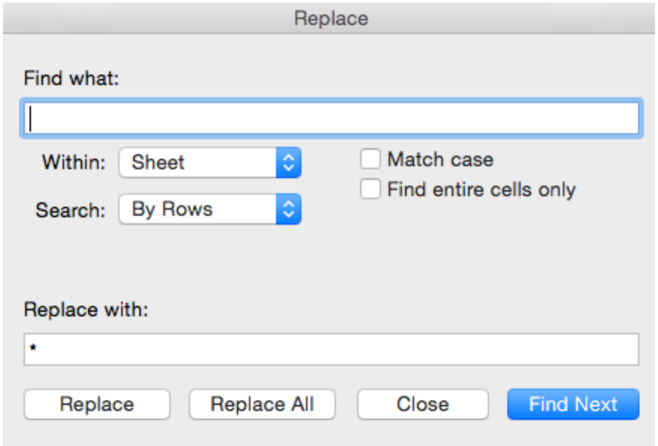
Lastly, we need a .csv file with no variable labels (i.e., data only). You can export a .csv file straight from SPSS, with only the variables that you want to analyse:

```
SAVE TRANSLATE OUTFILE= 'epa_example.csv'
/KEEP= id zextrav e pa cont pow est past powst contst
       Ztimeel social semi
/TYPE=CSV
/ENCODING='UTF8'
/MAP
/REPLACE
/CELLS=VALUES.
```

Remember that e, pa, cont, and pow are uncentred variables for state extraversion, positive affect, social contribution, and social power, whereas est, past, powst, and contst are the person-mean centred versions of these variables.

Just a couple of things to fix up in this .csv file:

1. Sometimes the export process makes the first value in the .csv (usually "1", because it's the first UserID) unreadable (but it looks like a normal number in the spreadsheet)—you just need to manually fix it otherwise Mplus will freak out. I think this issue might be due to using Unicode encoding in SPSS rather than Locale encoding.
2. Replace any blank cells with * or a missing value number[4] (e.g., -99) with Find + Replace in Excel (select the entire sheet). Then, specify your choice of missing value in the Mplus input file (described in the Mplus section).



Finally, note that if your variable names are longer than 8 letters, Mplus will truncate them. It's fine if you used longer variable names in the SPSS file, but you need to assign shorter variable names in the Mplus input file (these don't have to be the same ones that you used in SPSS because the .csv file doesn't contain any variable names; what you are doing in Mplus is assigning new variable labels to columns of data).

Bonus: Creating Lagged Variables

Although you can't really infer causality from naturalistic ESM designs, it is possible to add a time lag to at least meet one of the requirements for causality: precedence (i.e., a change in the theorised cause preceded a change in the outcome of interest). Note that I've added these instructions as a bonus section because we decided that lagged analyses for this particular dataset were not appropriate (reports were usually 3-4 hours apart, whereas we would expect extraverted behaviours and PA to covary within a shorter period of time).

Working from the EPA_ESM_8_Final.sav file, sort cases by time, within each person.

```
SORT CASES BY id(A) dt(A).
```

[4] If you have "system missing values" (i.e., blank cells, which look like this: " . ") in SPSS, you can use syntax to recode these to any number you want. For example:

```
RECODE var1 var2
        (SYSMIS=-99) (ELSE=COPY).
MISSING VALUES var1 var2 (-99).
EXECUTE.
```

The following syntax uses the lag function in SPSS to compute a lagged variable, and tells SPSS to set values on the lagged variable as missing if (1) the id at the current time point (`dt`) is different (`ne`) from the id at the previous time point (`lag(dt)`; this prevents the lagged variable from crossing participants), and if (2) the two reports are more than 1/6 of a day (i.e., 4 hours) apart. If you wanted to exclude reports that were over 6 hours apart, you could change 1/6 to 1/4, and so on.

With this syntax, if a participant responded "1", "2" and "3" for the variable `est` at T1, T2, and T3 respectively, then the lagged variable `elag` will be ".", "1", and "2", where . indicates a system missing variable (because there was no timepoint before the response at T1, "1").

```
IF (id=lag(id)) elag=lag(est).
IF $casenum = 1 or id ne lag(id) or dt-lag(dt) > (1/6) elag =
($SYSMIS).
IF (id=lag(id)) powlag=lag(powst).
IF $casenum = 1 or id ne lag(id) or dt-lag(dt) > (1/6) powlag =
($SYSMIS).
IF (id=lag(id)) contlag=lag(contst).
IF $casenum = 1 or id ne lag(id) or dt-lag(dt) > (1/6) contlag =
($SYSMIS).
IF (id=lag(id)) palag=lag(past).
IF $casenum = 1 or id ne lag(id) or dt-lag(dt) > (1/6) palag =
($SYSMIS).
EXECUTE.
```

Mplus can handle missing data, but it might be helpful to remove the lines of data with missing values. So, create a filter and save a new dataset, then apply the SELECT IF command as we've seen above to remove invalid observations:

```
COMPUTE lagged = 1.
IF elag = ($SYSMIS) lagged = 0.
EXECUTE.

SAVE OUTFILE='EPA_ESM_10_BonusLagged.sav'.
SELECT IF lagged = 1.
EXECUTE.
```

You can then export another Mplus dataset with these lagged variables, and run analyses with the lagged variable (e.g., `elag`, `powlag`, `contlag` at T1) predicting `past` at T2.

Multilevel Modelling of Personality Processes

The conceptual and statistical basis for multilevel modelling (MLM) has been covered extensively elsewhere:

Nezlek, J. B. (2008). An introduction to multilevel modeling for social and personality psychology. *Social and Personality Psychology Compass*, *2*(2), 842–860. http://dx.doi.org/10.1111/j.1751-9004.2007.00059.x

A very readable beginner-friendly introduction to MLM and why we need it.

Fleeson, W. (2007). Studying personality processes: Explaining change in between-persons longitudinal and within-person multilevel models. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of Research Methods in Personality Psychology* (pp. 523–542). New York: Guilford Press.

Helpful reference for the personality context, and has a step-by-step guide to running MLM in SPSS and interpreting the output, in case you need it.

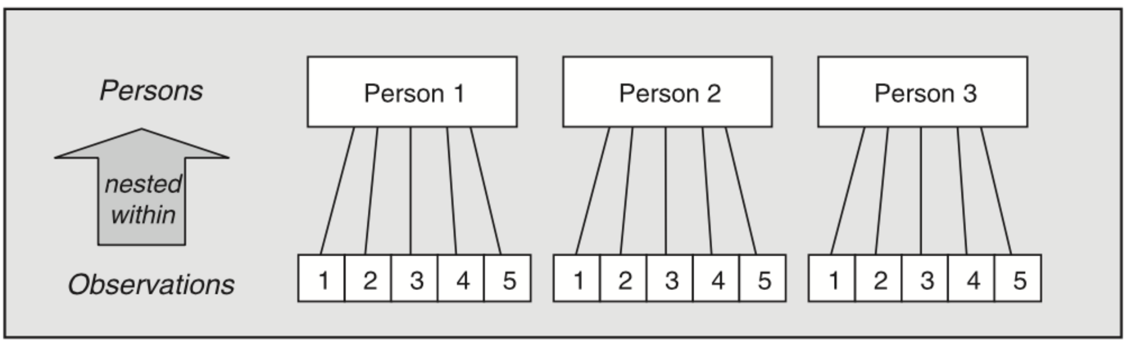
Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, *15*(3), 209–233. http://dx.doi.org/10.1037/a0020141

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: The advantages of multilevel SEM. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*(2), 161–182. http://dx.doi.org/10.1080/10705511.2011.557329

These two papers describe the methods for multilevel structural equation modelling in more technical detail (but are surprisingly readable), with accompanying Mplus syntax available at https://www.statmodel.com/download/Preacher.pdf. The 2010 paper is worth reading in full if you want to get stat-sy, and you will probably be using an approach based on these papers, so cite them both.

But here are a few basics:

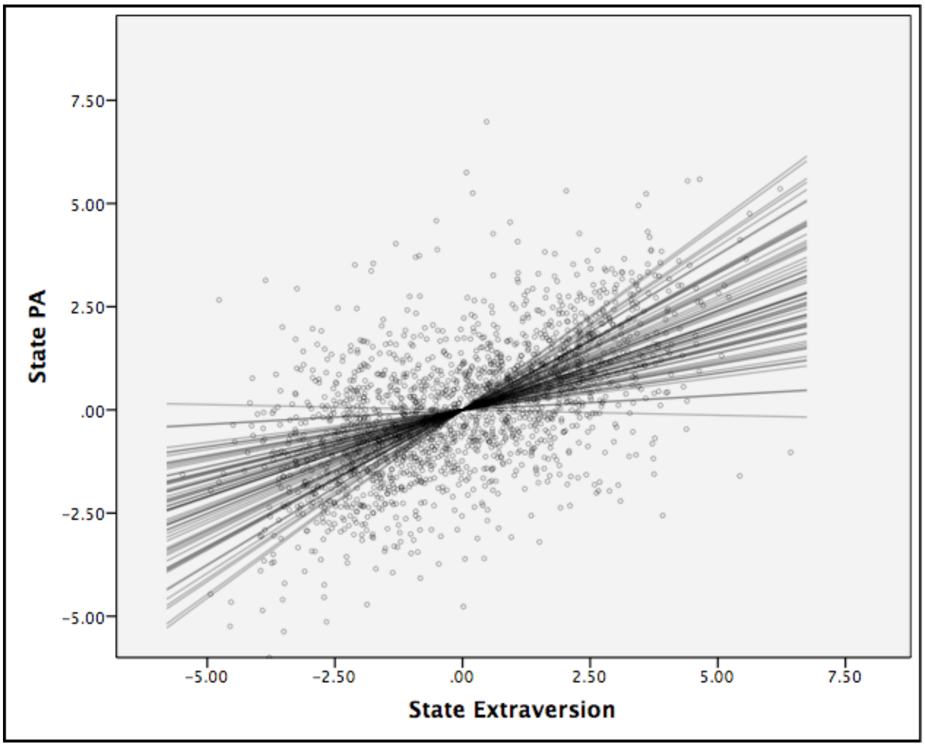
"Levels" refers to how the data are organised, and whether observations are independent. In an ESM study, observations (Level 1) are nested within persons (Level 2).



From Conner, Tennen, Fleeson, & Barrett (2009, p. 7).

We can also create "aggregate" variables that average across all observations for each person. This provides another form of trait (Level 2) data that we can use to answer questions like, "Do trait extraverts report greater average levels of PA across one week?" and "Do trait extraverts report acting more extraverted on average across one week?"

Using random slope models, we can allow the magnitude of the relation between two variables to vary between participants. For example, we don't have to assume (as in single-level modelling) that the relation between extraverted behaviour and PA states is exactly $b = 0.46$ for all participants; it's possible that some participants enjoy extraverted to a greater extent than other participants. Instead, each participant gets their own e-pa slope:



This is how you get this graph:

```
GGRAPH
  /GRAPHDATASET NAME="graphdataset" VARIABLES=est past id
MISSING=LISTWISE REPORTMISSING=NO
  /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
  SOURCE: s=userSource(id("graphdataset"))
  DATA: est=col(source(s), name("est"))
  DATA: past=col(source(s), name("past"))
  DATA: id=col(source(s), name("id"), unit.category())
  GUIDE: axis(dim(1), label("State Extraversion"))
```

```
  GUIDE: axis(dim(2), label("State PA"))
  GUIDE: legend(aesthetic(aesthetic.color.exterior), label("id"))
  ELEMENT: point.jitter(position(est*past),
transparency.exterior(transparency."0.7"), size(size."3"))
  ELEMENT: line(position(smooth.linear(est*past)), split(id),
transparency(transparency."0.7"))
END GPL.
```

From these data, MLM allows us to compute the *mean* relation between two variables, as well as the *variance* around this mean, and whether this slope variance is statistically significant. If this slope variance is significant, we can then test whether another variable might explain some of this slope variance (e.g., is the slope greater for trait extraverts?).

The main reason why we need MLM is that single-level regression assumes **independence of observations**. In an ESM study, observations are not independent, as they share in common the person's characteristics.

Data Screening

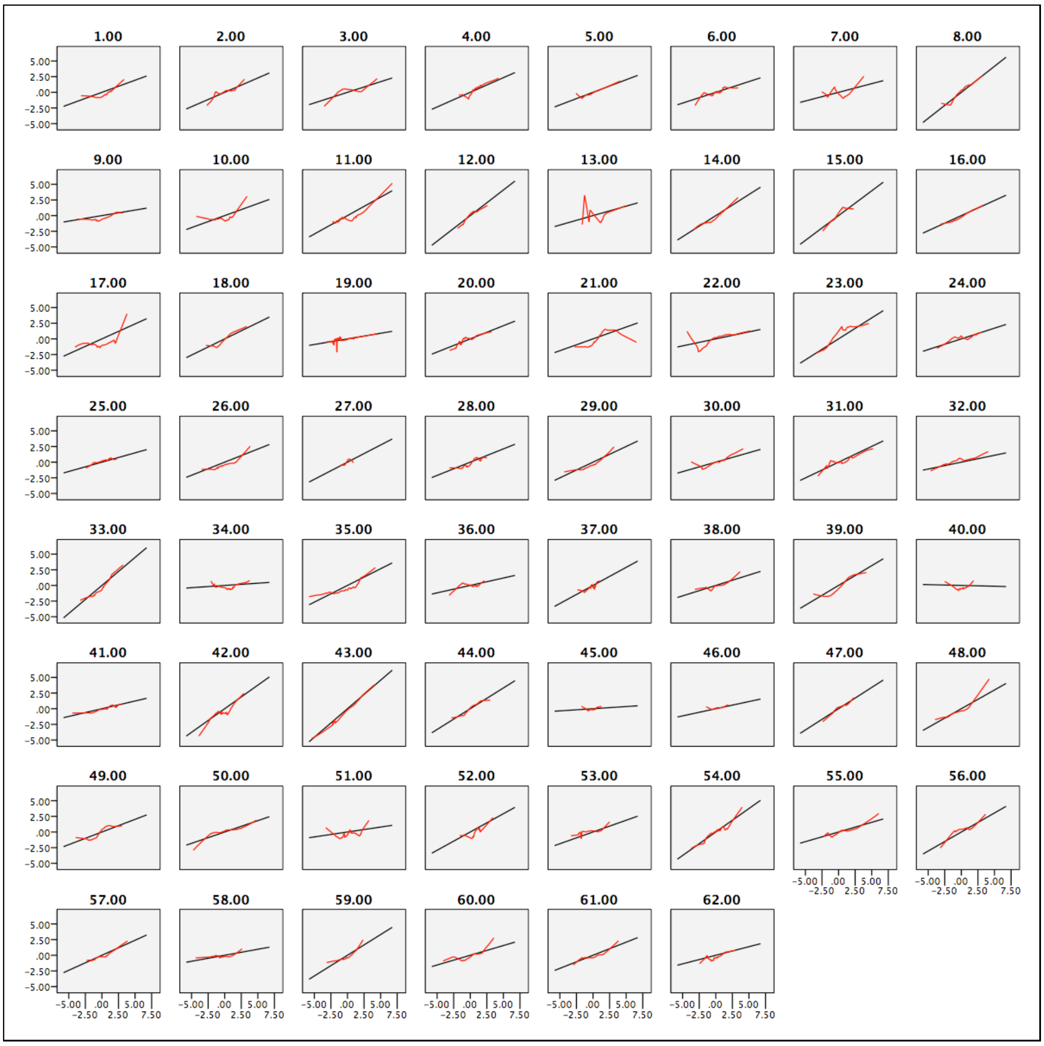
Before we jump into our multilevel analyses, we first need to check that our data are suitable for these analyses. The assumptions of MLM are (1) linearity, (2) normality, (3) homoscedasticity, (4) uncorrelated level 1 and 2 residuals, and (5) uncorrelated errors at the highest level. In practice, you can probably get away with just checking for nonlinearity, because Mplus uses a robust maximum likelihood estimator that does not require the assumption of normality for MSEM analyses (Preacher et al., 2010), and the other assumptions are more difficult to check.

At the trait or aggregate state level, screening for nonlinearity just involves looking at scatterplots. At the state-level, you can create separate plots to look at the relation between pairs of variables, for each participant, using this SPSS syntax:

```
GGRAPH
  /GRAPHDATASET NAME="graphdataset" VARIABLES=est past
id[LEVEL=NOMINAL] MISSING=LISTWISE
    REPORTMISSING=NO
  /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
  PAGE: begin(scale(1200px,1200px))
  SOURCE: s=userSource(id("graphdataset"))
  DATA: est=col(source(s), name("est"))
  DATA: past=col(source(s), name("past"))
  DATA: id=col(source(s), name("id"), unit.category())
  COORD: rect(dim(1,2), wrap())
  GUIDE: axis(dim(1))
  GUIDE: axis(dim(2))
  GUIDE: axis(dim(3), opposite())
```

```
  ELEMENT: line(position(smooth.linear(est*past*id)),
color(color.black))
  ELEMENT: line(position(smooth.loess(est*past*id)),
color(color.red))
  PAGE: end()
END GPL.
```

This gives us cool output like this, which reveals some nonlinearity (e.g., participants 7, 10, 13, 21)…but on the whole, it's not that bad. If it was really bad (again, somewhat subjective…), then you might want to consider excluding some of the problematic participants.



Calculating Scale Reliabilities for Repeated State Measures

Before conducting your key analyses, you'll also want to check the reliability of your measures. For ESM measures, the metric of interest is the reliability with which a measure captures *within-person change*. I won't go into the technical details here (but the references below do, and are excellent); instead, I'll just show you how to compute these omega (ω) reliability coefficients using Mplus (syntax from Bolger & Laurenceau, 2013b).

First, you need to export a separate .csv file that just contains the individual items, as well as the `id` and `obs` variables. Working from EPA_ESM_8_Final.sav, run this syntax:

```
SAVE TRANSLATE OUTFILE= 'epa_items.csv'
/KEEP= id obs ex1 ex2 ex3 ex4r ex5 ex6r
pa1 pa2 pa3 pow1 pow2 cont1 cont2 cont3
/TYPE=CSV
/ENCODING='UTF8'
/MAP
/REPLACE
/CELLS=VALUES.
```

All Mplus input and output files are included in the supplement as annotated files (annotations are preceded by an exclamation mark [!], which tells Mplus to ignore your notes). When you open up the program, the window that opens up is the input file. If you save your input file in the same folder as your .csv data file, you won't need to set the directory (i.e., tell Mplus to find the file somewhere else). When a model runs, Mplus creates an output file in that same folder, with the same name (but ending with .out). Every time you run a new model, you should save it as a new input file, preferably with an informative name (e.g., e_cont_pa.inp, not Model2.inp).

In the Mplus Files/Reliability folder, I've generated input and output files for computing omega reliability coefficients for each of our four variables: extraversion, PA, contribution, and power. I'll just go through the extraversion files (e_reliability.inp & .out) as an example here.

`TITLE` is self-explanatory. `FILE IS` tells Mplus which file to read from. Note that in Mplus, you need to separate all commands with a semicolon (`;`). Under `VARIABLE`, we specify the names of all of the variables (`NAMES ARE`) in the .csv file, in the order that you saved them into the file. Here, using SPSS syntax to prepare your data is again helpful as you can just copy over the variable names from the `SAVE TRANSLATE` command above (p. 41). The `MISSING` command tells Mplus which values are missing values. In the example data set it is *, but you can choose any missing value, but you need to put brackets around it if it's a number, e.g.: "`MISSING ARE ALL (-99)`"

For `USEVARIABLES`, you need to type in all of the variables that you are using in the current analyses; that is, variables that appear in the `MODEL` section. `WITHIN` tells Mplus which variables are within-person (that is, Level 1). `CLUSTER = id` tells Mplus to nest observations within participants (identified by the `id` variable).

Under `ANALYSIS`, we specify that we're running a two-level random slopes model The `MODEL` component begins with `%WITHIN%` to specify the within-person component of the model (I won't go through the `%BETWEEN%` section here). `Ew` is the name of the within-person extraversion factor, and `ex1*` specifies that the extraversion Item 1 loads on `Ew` with a loading that is estimated from the data and assigned the label of `a`. `ex2 ex3 ex4 ex5 ex6 (b-f);` specifies that Items 2–6 load on the `Ew` factor, and are assigned the labels `b`, `c`, `d`, `e`, and `f`, respectively. Next, `ex1 ex2 ex3 ex4 ex5 ex6 (g-l);` gives labels `g`, `h`, `i`, `j`, `k`, l to the error variances of Items 1, 2, 3, 4, 5, 6 respectively. `Ew@1;` fixes the variance of the factor to 1.0.

The syntax under `MODEL CONSTRAINT` uses the estimates of the labelled parameters (`a`, `b`, `c`, etc.) to calculate *ω* for the within-person part of the model.

```
TITLE:    Extraversion reliability;
DATA:     FILE IS epa_items.csv;
VARIABLE: NAMES ARE id obs ex1 ex2 ex3 ex4 ex5 ex6
pa1 pa2 pa3 pow1 pow2 cont1 cont2 cont3;
USEVAR = ex1 ex2 ex3 ex4 ex5 ex6 obs;

WITHIN = obs;
CLUSTER = id;
missing are *;
ANALYSIS: TYPE = TWOLEVEL ;
PROCESSORS=4;
MODEL:    %WITHIN%
            Ew by
          ex1 * (a)
          ex2 ex3 ex4 ex5 ex6 (b-f);
          ex1 ex2 ex3 ex4 ex5 ex6 (g-l);
          Ew@1;
          ex1 ex2 ex3 ex4 ex5 ex6 on obs;

          %BETWEEN%
            Eb by
            ex1*
            ex2 ex3 ex4 ex5 ex6;
            Eb@1;
            ex1 ex2 ex3 ex4 ex5 ex6 ;

          MODEL CONSTRAINT:
            new (omega);
            omega=((a+b+c+d+e+f)**2)/((a+b+c+d+e+f)**2 + (g+h+i+j+k+l));

OUTPUT:   sampstat;
```

To run this analysis, hit Alt + R (or click Run). The key part of the output file is the New/Additional Parameters section, which shows you that the *ω* coefficient for the 6 extraversion items is .927. This suggests that it is possible to reliably distinguish people in terms of their patterns of change over time, using these items and for the particular days in this study.

```
New/Additional Parameters
    OMEGA             0.927      0.006    164.545      0.000
```

Bolger, N., & Laurenceau, J.-P. (2013b). Psychometrics of intensive longitudinal measures of emotional states. In N. Bolger & J.-P. Laurenceau (Eds.), *Intensive longitudinal methods: An introduction to diary and experience sampling research* (pp. 127–142). New York: Guilford Press.

Shrout, P. E., & Lane, S. P. (2012). Psychometrics. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 302–320). New York: Guilford Press.

These two book chapters are very readable. Shrout and Lane (2012) reviews a few different methods, and is worth reading if you're interested in different approaches that have been used in the past. But Bolger and Laurenceau (2013) is probably the most useful as it provides the Mplus syntax for calculating omega reliability coefficients.

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, *19*(1), 72–91. http://dx.doi.org/10.1037/a0032138

A more thorough and technical description of the methods. Worth citing (and reading if you want a better understanding of this).

Within-Person Mediation Models

Now we're finally ready to start answering our key research questions with Mplus!

1. Do people report greater levels of PA in moments when they are acting more extraverted?

Here is the input for a basic two-level model (e_pa.inp).

```
INPUT INSTRUCTIONS

  TITLE:       Do E & PA co-vary within-person?

  DATA:        FILE IS epa_example.csv;

  VARIABLE:    NAMES ARE id extrav e pa cont pow est past powst contst
              timeel social semi;

               MISSING ARE *;

               USEVARIABLES ARE id timeel est past social semi;

               WITHIN ARE est past timeel social semi;

               CLUSTER = id;

  ANALYSIS:    TYPE = TWOLEVEL RANDOM; ESTIMATOR=ml;

  MODEL:       %WITHIN%
               s | past on est;
               past ON timeel social semi;

  OUTPUT:      TECH1 TECH8 CINTERVAL;
```

Under `ANALYSIS`, we specify that we're running a two-level random slopes model, using a maximum likelihood estimator. The `MODEL` component begins with `%WITHIN%` to specify the within-person component of the model. s | past on est tells Mplus to regress state PA (the DV) onto state extraverted behaviour (the IV), and to call the random slope (specified by the |) s (but you can call it whatever you want). The next part of the model, `past ON timeel social semi`, specifies the covariates. Here, we control for any effects of timecourse and social situation on state PA. Finally, we ask for some technical `OUTPUT` and 95% confidence intervals (`CINTERVAL`).

After running this analysis, let's take a look at the important parts of the output that Mplus gives us (e_pa.out; truncated here for brevity). In the `Within Level` section, none of the covariates (`TIMEEL, SOCIAL, SEMI`) have a significant effect on state PA. In the `Between Level` section, we see that the mean relation (unstandardised beta coefficient) between state extraverted behaviour and state PA is $b = 0.46$, and that this relation is statistically significant, $p < .001$. However, the `Variances` output also tells us that there is significant variance around this estimate, $p < .001$. Square-rooting the variance point estimate of 0.027 gives us $SD = 0.16$.

```
MODEL RESULTS
                                                        Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
Within Level

 PAST       ON
    TIMEEL             0.006      0.029      0.202      0.840
    SOCIAL            -0.032      0.053     -0.599      0.549
    SEMI              -0.080      0.044     -1.835      0.066

 Intercepts
    PAST               0.002      0.029      0.081      0.935

 Residual Variances
    PAST               1.529      0.051     29.863      0.000

Between Level

 Means
    S                  0.457      0.030     15.363      0.000

 Variances
    S                  0.027      0.007      3.655      0.000
```

The next section gives us the confidence intervals, which gives us important information about the precision of the effect. The `Lower 2.5%` and `Upper 2.5%` columns give us the lower and upper bounds, respectively, of a 95% confidence interval. From this, we can see that the 95% CI around the point estimate for the extraverted behaviour-PA relation is [0.40, 0.52].

```
CONFIDENCE INTERVALS OF MODEL RESULTS

                  Lower .5%  Lower 2.5%    Lower 5%    Estimate    Upper 5%  Upper 2.5%   Upper .5%

Within Level

 PAST     ON
    TIMEEL          -0.070      -0.052      -0.042       0.006       0.054       0.063       0.081
    SOCIAL          -0.168      -0.136      -0.119      -0.032       0.055       0.072       0.105
    SEMI            -0.193      -0.166      -0.152      -0.080      -0.008       0.005       0.032

 Intercepts
    PAST            -0.073      -0.055      -0.046       0.002       0.050       0.060       0.077

 Residual Variances
    PAST             1.398       1.429       1.445       1.529       1.614       1.630       1.661

Between Level

 Means
    S                0.381       0.399       0.408       0.457       0.506       0.516       0.534

 Variances
    S                0.008       0.013       0.015       0.027       0.039       0.042       0.046
```

2. Does trait extraversion moderate the relation between extraverted behaviour and PA?

As we saw above, there was significant variation in the relation between extraverted behaviour and PA. Could it be that those who are more extraverted enjoy acting extraverted to a greater extent than introverts?

```
INPUT INSTRUCTIONS

  TITLE:      Does trait E moderate the e-pa relation?

  DATA:       FILE IS epa_example.csv;

  VARIABLE:   NAMES ARE id extrav est past powst contst timeel social semi;
              MISSING ARE *;
              USEVARIABLES ARE id timeel est past social semi extrav;
              WITHIN ARE est past timeel social semi;
              BETWEEN ARE extrav;
              CLUSTER = id;

  ANALYSIS:   TYPE = TWOLEVEL RANDOM; ESTIMATOR=ml;

  MODEL:      %WITHIN%
              s | past on est;
              past ON timeel social semi;

             %BETWEEN%
              s on extrav;
```

Most of this input (e_pa_traite.inp) will be the same as for the model above. In this model, however, we now have a new variable, `extrav` (trait extraversion) in the `USEVARIABLES` and `BETWEEN ARE` (because this is a Level 2 variable) commands. In the `MODEL` command we also add a `%BETWEEN%` component to the model. `s on extrav` tells the Mplus to model whether trait extraversion predicts the magnitude of the relation between state extraversion and state PA.

Here's the key part of the output (e_pa_traite.out). The key line of output is the `S ON EXTRAV` estimate. This tells us that trait extraversion does not significantly moderate the relation between state extraversion and PA, *b* = -0.02, *p* = .510. Again, if you scroll down, Mplus also gives us the 95% CI around this estimate. I won't show the output here, but it's [-0.15, 0.08].

```
Between Level

 S          ON
    EXTRAV              -0.018       0.027      -0.658       0.510

 Intercepts
    S                    0.459       0.030      15.391       0.000

 Residual Variances
    S                    0.027       0.007       3.669       0.000
```

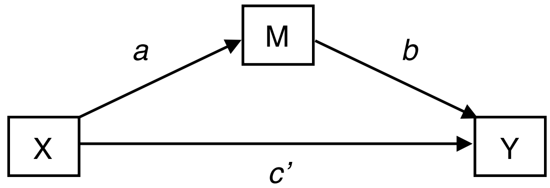
3. Do social contribution and power mediate the relation between extraverted behaviour and PA?

The syntax and interpretations that follow are very closely adapted from Bolger and Laurenceau (2013c):

Bolger, N., & Laurenceau, J.-P. (2013c). Within-subject mediation analysis. In N. Bolger & J.-P. Laurenceau (Eds.), *Intensive longitudinal methods: An introduction to diary and experience sampling research* (pp. 177–195). New York: Guilford Press.

If there is one article on within-person mediation that you absolutely should read, it is this one!

The basic idea in mediation is that an independent variable X (e.g., extraverted behaviour) transmits part of its effect on dependent variable Y (e.g., PA) via a mediating variable M (e.g., social contribution). As shown below, the a path is the effect that X has on M, and the b path is the effect that M has on Y. The c′ path, otherwise known as the direct effect, is the effect that X has on Y, after accounting for its indirect effect via M.



Social contribution. Let's start by testing the mediating effect of social contribution. The new input file (e_contpow_pa.inp) is shown on the next page. You need to add the mediator (`contst`) to the `USEVARIABLES ARE` and `WITHIN` commands.

```
INPUT INSTRUCTIONS
  TITLE:      Does State Social Contribution Mediate the State E-PA Relation?
  DATA:       FILE IS epa_example.csv;
  VARIABLE:    NAMES ARE id extrav e pa cont pow est past powst contst
             timeel social semi;
              MISSING ARE *;
              USEVARIABLES ARE id timeel est past contst social semi;
              WITHIN = est contst past timeel social semi;
              CLUSTER = id;

  ANALYSIS:   TYPE = TWOLEVEL RANDOM; ESTIMATOR=ml;

   MODEL:    %WITHIN%
             a | contst on est;
             b | past on contst;
             cp | past on est;
             past contst ON timeel social semi;
             [contst@0 past@0];

             %BETWEEN%
             cp WITH a b;
             [a] (ma); a (vara);
             [b] (mb); b (varb);
             [cp] (mcp);
             a WITH b (covab);

             MODEL CONSTRAINT:
             NEW(med te pme corr);
             med=ma*mb+covab;
             te=med+mcp;
             pme=med/te;
             corr=covab/sqrt(vara*varb);

   OUTPUT:    TECH1 TECH8 CINTERVAL;
```

In the `MODEL: %WITHIN%` section, the first three lines tell Mplus to model the *a* (M on X), *b* (Y on M), and *c'* (Y on X) paths. The fourth line adds the covariates, as in the previous models, but note that we also add `contst` to the left side of the `ON` command, as we want to remove the effects of the covariates on the mediator as well. Note that we could have written these as two separate lines (`past ON timeel social semi; contst ON timeel` social semi;), but Mplus lets us specify both at the same time (past contst ON timeel social semi;). Finally, [contst@0 past@0] fixes the intercepts of the m and y equations to 0 because both of these variables were subject-mean centred.

In the %BETWEEN% component, cp WITH a b gives you the covariances of the cp random effect with the a and b random effects. [a] (ma) requests the mean (i.e., fixed effect) of the a slope and labels it "ma" , whereas a (vara) requests the variance of the a slope random effect and labels it "vara". The next two lines do the same for the b slope fixed and random effects and the mean of the cp effect. Finally, a WITH b (covab) gives us the covariance of the a and b random effects, labelled as "covab" (see Kenny, Korchmaros, & Bolger, 2003, or Bolger and Laurenceau, 2013, for why we need this term).

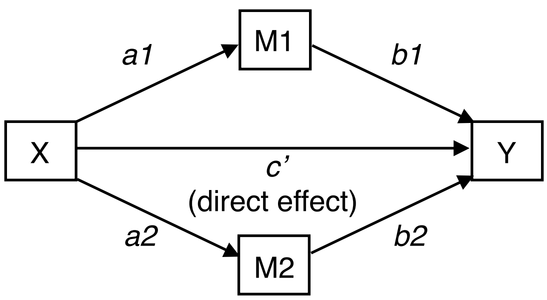
Finally, we ask Mplus to calculate the direct and indirect effects under MODEL CONSTRAINT. We first create new parameters for the indirect (i.e., mediated) effect (med), total effect (te), percent mediated effect (pme), and the correlation between the a and b random effects (i.e., standardises covab between 0-1). Let's look at each of these lines in turn. med=ma*mb+covab asks Mplus to multiply the fixed a and b effects and add the covariance of the a and b random effects, giving us the mediated effect. For the next line (te=med+mcp), remember that the total effect comprises the indirect effect (med, which we just computed) and the direct effect (mcp), and that's exactly what this tells Mplus to compute. The percent mediated effect (pme) is a metric that helps us to communicate the size of the mediated effect, as a proportion of the total effect (see Preacher & Kelley, 2011).

Here are the key parts of the output (e_contpow_pa.out). Under Means, we find the point estimates for each of the *a*, *b*, and *c′* paths. So, for example, the average size of the relation between extraverted behaviour and social contribution is $b = 0.54$. For the purposes of testing mediation, however, our new parameters MED, TE, and PME are more interesting. These reveal a significant indirect effect of extraverted behaviour on PA via social contribution, $b = 0.11$, $p < .001$, which accounts for 23.1% of the total effect. All of these parameters have a corresponding 95% confidence interval, if you scroll down in the output file.

```
MODEL RESULTS

                                                   Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
Between Level

 CP       WITH
    A                  0.000      0.006     -0.032      0.975
    B                 -0.007      0.008     -0.872      0.383

 A        WITH
    B                 -0.005      0.007     -0.795      0.427

 Means
    A                  0.537      0.030     17.723      0.000
    B                  0.206      0.032      6.396      0.000
    CP                 0.351      0.032     11.117      0.000

 Variances
    A                  0.026      0.008      3.322      0.001
    B                  0.028      0.011      2.488      0.013
    CP                 0.025      0.009      2.871      0.004

New/Additional Parameters
    MED                0.106      0.019      5.665      0.000
    TE                 0.457      0.030     15.388      0.000
    PME                0.231      0.040      5.718      0.000
    CORR              -0.190      0.233     -0.814      0.416
```

Social power. To test the mediating effect of social power, all we need to do is copy over the syntax from the input file and just replace `contst` with `powst` for every line after the `NAMES ARE` command. You can check out the output (e_pow_pa.out) yourself, but the basic result is that the indirect effect via power is also significant, accounting for 12.5% of the total effect of extraversion on PA.

Social contribution and power. Finally, let's see what happens when we enter social contribution and power as *simultaneous* mediators. This will allow us to establish the extent to which social contribution and power *independently* mediate the relation between extraverted behaviour and PA.



Although this input file (e_contpow_pa.inp) looks a bit longer than the ones we've seen so far, it's really a natural extension of the syntax for testing a single-mediator model. So I'll go through this by comparing it to the single-mediator syntax.

```
TITLE:      Do State Social Contribution and Power Mediate the State E-PA
Relation?
DATA:       FILE IS epa_example.csv; ! Name of the file
VARIABLE:   NAMES ARE id extrav e pa cont pow est past powst contst
             timeel social semi;
            MISSING ARE *;
            USEVARIABLES ARE id timeel est past contst powst social semi;
            WITHIN = est contst powst past timeel social semi;
            CLUSTER = id;

ANALYSIS:   TYPE = TWOLEVEL RANDOM;

MODEL:    %WITHIN%
              a1 | contst ON est;
              a2 | powst ON est;
              b1 | past ON contst;
              b2 | past ON powst;
              cp | past ON est;
              past contst powst ON timeel social semi;
              contst WITH powst;
             [contst@0 powst@0 past@0];

           %BETWEEN%
           cp WITH a1 b1 a2 b2;
           [a1] (ma1); a1 (vara1);
           [b1] (mb1); b1 (varb1);
           [a2] (ma2); a2 (vara2);
           [b2] (mb2); b2 (varb2);
           [cp] (mcp);
           a1 WITH b1 (covab1);
           a2 WITH b2 (covab2);

MODEL CONSTRAINT:
         NEW (med1 med2 te pme1 pme2 corr1 corr2);
         med1=ma1*mb1+covab1;
        med2=ma2*mb2+covab2;
         te=med1+med2+mcp;
         pme1=med1/te;
         pme2=med2/te;
         corr1=covab1/sqrt(vara1*varb1);
         corr2=covab2/sqrt(vara2*varb2);

 OUTPUT:   TECH1 TECH8 CINTERVAL;
```

First, we need to make sure that all of the variables we're using in this analysis—including both mediators—are in the `USEVARIABLES ARE` command. We also need to add the second mediator, powst, to the `WITHIN` command line.

In the `MODEL: %WITHIN%` section, we again need to estimate the *a* and *b* paths, but this time, we do this for both mediators. In this case, `a1` refers to the *a* path for the first mediator (`contst`), whereas `a2` refers to the *a* path for the second mediator (`powst`). Same thing for the *b* paths. In the line where we're controlling for covariates, remember to add powst to the left of the `ON` command. The new line, `contst WITH powst`, estimates the residual covariance between social contribution and power. Finally, we also need to fix the intercepts of the powst equations to 0.

In the `%BETWEEN%` section, everything we previously did for `a` and `b` now need to be repeated for `a1`, `a2`, `b1`, and `b2` (as shown in the preceding diagram). Same thing applies for the `MODEL CONSTRAINT` section: we now need to calculate two indirect effects (`med1` and `med2`), two percent mediated effects (`pme1` and `pme2`), and two correlations between the `a` and `b` random effects for social contribution (i.e., `a1` and `b1`) and social power (i.e., `a2` and `b2`). And the total effect now comprises the two indirect effects, plus the direct effect.

Again, we'll skip to the Between Level: Means section of the output file (e_contpow_pa.out). These estimates reveal a significant effect of extraversion on social contribution (A1), extraversion on social power (A2), and social contribution on PA (B1), but no significant effect of social power on PA (B2). Jumping down to the New/Additional Parameters, we can see that in this two-mediator model, social contribution explains a significant 22% of the relation between extraverted behaviour and PA, *b* = 0.10, *p* < .001, but power is not a significant mediator, *b* = 0.02, *p* = .296.

```
MODEL RESULTS

                                                          Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

Between Level

 Means
    A1                 0.538      0.028     19.027      0.000
    A2                 0.659      0.038     17.547      0.000
    B1                 0.191      0.036      5.251      0.000
    B2                 0.052      0.027      1.908      0.056
    CP                 0.337      0.028     11.902      0.000

 Variances
    A1                 0.022      0.008      2.967      0.003
    A2                 0.051      0.013      3.789      0.000
    B1                 0.025      0.018      1.416      0.157
    B2                 0.012      0.006      1.835      0.066
    CP                 0.015      0.012      1.230      0.219

New/Additional Parameters
    MED1               0.100      0.021      4.754      0.000
    MED2               0.019      0.018      1.046      0.296
    TE                 0.456      0.029     15.784      0.000
    PME1               0.220      0.047      4.728      0.000
    PME2               0.042      0.040      1.062      0.288
    CORR1             -0.113      0.176     -0.638      0.523
    CORR2             -0.614      0.272     -2.257      0.024
```

Example Write-up

That's it! Here's an example of how to describe the analyses and clearly report the results.

Method

…

Data Analyses

The resulting data had a multilevel structure, with reports (Level 1) nested within individuals (Level 2). Descriptive statistics and correlations for Level 2 variables were computed using SPSS Version 22. Omega (ω) reliability coefficients (McDonald, 1999) for trait measures (a less biased alternative to Cronbach's alpha; Dunn, Baguley, & Brunsden, 2014) were computed in R (R Development Core Team, 2015) using the MBESS package (Kelley & Lai, 2012). ω reliability coefficients for state measures (Bolger & Laurenceau, 2013b; Shrout & Lane, 2012) were computed using Mplus Version 7 (Muthén & Muthén, 1998–2012).

Mediation hypotheses were examined using a multilevel structural equation modelling (MSEM) framework with effects modelled as random slopes (Muthén & Asparouhov, 2011; Preacher, Zhang, & Zyphur, 2011; Preacher, Zyphur, & Zhang, 2010). Given the null correlations between trait and aggregate state extraversion (described below), all MSEMs had a 1–1–1 structure, with extraversion, hypothesised mediators, and PA each assessed at the state level. State measures were centred around each individual's mean (i.e., aggregated states) to remove trait influences on state variables. Multilevel analyses were then deployed using Mplus Version 7 (Muthén & Muthén, 1998–2012) with syntax adapted from Bolger and Laurenceau (2013c, p. 188).

Given that extraverted behaviours and most of the mediators refer to social processes, all multilevel models partialled out the effects of social context on state PA and the mediating variables. Following Bolger and Laurenceau's (2013a) strong recommendation, all multilevel models also controlled for time course (time elapsed since each individual's first ESM report). The resulting unstandardised b coefficients quantify the average amount of change in the dependent variable associated with a one-unit change in the independent variable, independent of social situational and time course confounds. For transparent reporting, supplementary analyses in Appendix B demonstrate that omitting these controls does not affect the conclusions.

Results

Consistent with Hypothesis 1, individuals tended to report higher levels of PA in *moments* when they reported greater levels of extraverted behaviours, $b = 0.46$, $p < .001$, 95% CI [0.40, 0.52]. The magnitude of this within-person relation varied significantly between persons, $SD = 0.16$, $p < .001$. However, consistent with all previous studies in this literature, extraverts and introverts both reported similar increases in PA associated with extraverted behaviours, as revealed by a nonsignificant interaction between trait extraversion and extraverted behaviours, $b = -0.02$, $p = .510$, 95% CI [-0.07, 0.04].

Consistent with Hypotheses 2 and 3, contribution (Model 1) and power (Model 2) both mediated the relation between extraverted behaviours and PA states in single-mediator models (see Table 1). Contribution and power respectively explained 24.1% and 12.5% of the total relation between state extraversion and PA. When entered together as simultaneous mediators (Model 3), however, contribution remained a significant mediator (PME = 23.1%), but power no longer had a significant independent mediating effect.

Table 3

Mediators of the Relation Between Extraverted Behaviours and PA States

Mediating Variables	IV → M (*a* path)	M → DV (*b* path)	Direct effect (*c'* path)	Indirect effect (*a* x *b* path)	95% CI of indirect effect
Model 1					
Contribution	0.54***	0.21***	0.35***	0.11***	[0.07, 0.15]
Model 2					
Power	0.66***	0.11***	0.40***	0.06**	[0.02, 0.09]
Model 3					
Contribution	0.54***	0.20***	0.34***	0.11***	[0.06, 0.15]
Power	0.66***	0.05		0.02	[-0.02, 0.06]

Troubleshooting and Further Questions

There lies the extent of my Mplus knowledge, but I hope it's enough to give you a basic idea of what Mplus can do. Here are some resources that are likely to help you to figure out how to implement the analyses you have in mind, and to troubleshoot issues and error messages:

1. **Google!!!** With the magic words "multilevel mediation syntax mplus", I was able to track down the syntax I needed to perform my analyses within minutes. That is often the case (all you need to do is to adapt it), so Google away! You can also Google any error messages that Mplus gives you; this will likely direct you to the…
2. **Mplus forums.** The Muthéns and other users have likely answered any questions that you have: http://www.statmodel.com/discussion/messages/board-topics.html
3. **User Manual.** Describes all the features of the program, with several syntax examples. Available here: http://www.statmodel.com/ugexcerpts.shtml
4. **Tutorial videos.** From the creators of Mplus: https://www.statmodel.com/course_materials.shtml
5. **UCLA resources.** Includes seminar notes and data files, FAQs about Mplus, annotated output, and more: http://www.ats.ucla.edu/stat/mplus/
6. **UT resources.** General Mplus tutorial: https://stat.utexas.edu/images/SSC/documents/SoftwareTutorials/MPlus_Tutorial.pdf; MLM tutorial: https://stat.utexas.edu/images/SSC/documents/SoftwareTutorials/MultilevelModeling.pdf

References

Allen, T. A., & DeYoung, C. G. (2015). Personality neuroscience and the five factor model. In T. A. Widiger (Ed.), *Oxford handbook of the five factor model*. New York: Oxford University Press.

Bolger, N., & Laurenceau, J.-P. (2013a). Fundamentals of intensive longitudinal data. In N. Bolger & J.-P. Laurenceau (Eds.), *Intensive longitudinal methods: An introduction to diary and experience sampling research* (pp. 27–39). New York: Guilford Press.

Bolger, N., & Laurenceau, J.-P. (2013b). Psychometrics of intensive longitudinal measures of emotional states. In N. Bolger & J.-P. Laurenceau (Eds.), *Intensive longitudinal methods: An introduction to diary and experience sampling research* (pp. 127–142). New York: Guilford Press.

Bolger, N., & Laurenceau, J.-P. (2013c). Within-subject mediation analysis. In N. Bolger & J.-P. Laurenceau (Eds.), *Intensive longitudinal methods: An introduction to diary and experience sampling research* (pp. 177–195). New York: Guilford Press.

Conner, T. S., & Barrett, L. F. (2012). Trends in ambulatory self-report. *Psychosomatic Medicine*, *74*(4), 327–337. http://dx.doi.org/10.1097/PSY.0b013e3182546f18

Conner, T. S., & Lehman, B. (2012). Getting started: Launching a study in daily life. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 89–107). New York: Guilford Press.

Conner, T. S., Tennen, H., Fleeson, W., & Barrett, L. F. (2009). Experience sampling methods: A modern idiographic approach to personality research. *Social and Personality Psychology Compass*, *3*(3), 292–313. http://dx.doi.org/10.1111/j.1751-9004.2009.00170.x

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychological Methods*, *12*(2), 121–38. http://doi.org/10.1037/1082-989X.12.2.121

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, *80*(6), 1011–1027. http://dx.doi.org/10.1037//0022-3514.80.6.1011

Fleeson, W. (2007). Studying personality processes: Explaining change in between-persons longitudinal and within-person multilevel models. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of Research Methods in Personality Psychology* (pp. 523–542). New York: Guilford Press.

Fleeson, W., & Gallagher, P. (2009). The implications of Big Five standing for the distribution of trait manifestation in behavior: Fifteen experience-sampling studies and a meta-analysis. *Journal of Personality and Social Psychology*, *97*(6), 1097–1114. http:// dx.doi.org/10.1037/a0016786

Fraley, R. C., & Vazire, S. (2014). The N-pact factor: evaluating the quality of empirical journals with respect to sample size and statistical power. *PLOS ONE*, *9*(10), e109019. http://dx.doi.org/10.1371/journal.pone.0109019

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, *19*(1), 72–91. http://dx.doi.org/10.1037/a0032138

Green, A. S., Rafaeli, E., Bolger, N., Shrout, P. E., & Reis, H. T. (2006). Paper or plastic? Data equivalence in paper and electronic diaries. Psychological Methods, 11(1), 87–105. http://dx.doi.org/10.1037/1082-989X.11.1.87

Kelley, K., & Lai, K. (2012). MBESS (Version 3.3.3) [R package]. Retrieved from https://cran.r-project.org/web/packages/MBESS/index.html

Kenny, D. A., Korchmaros, J. D., & Bolger, N. (2003). Lower level mediation in multilevel models. *Psychological Methods*, *8*(2), 115–128. http://doi.org/10.1037/1082-989X.8.2.115

Kreft, G., & de Leeuw, J. (1998). Introducing multilevel modeling. Thousand Oaks, CA: Sage.

Maas, C. J., & Hox, J. (2005). Sufficient sample sizes for multilevel modeling. *Journal of Research Methods for the Behavioral and Social Sciences*, *1*(3), 86–92. http://dx.doi.org/10.1027/1614-1881.1.3.86

McCabe, K. O., Mack, L., & Fleeson, W. (2012). A guide for data cleaning in experience sampling studies. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 321–338). New York: Guilford Press.

McDonald, R. P. (1999). *Test theory: A unified treatment.* Mahwah, N. J.: L. Erlbaum Associates.

Mehl, M. R., & Conner, T. S. (2012). *Handbook of research methods for studying daily life*. New York: Guilford Press.

Mehl, M. R., & Robbins, M. L. (2012). Naturalistic observation sampling: The electronically activated recorder (EAR). In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 176–192). New York: Guilford Press.

Mehl, M. R., Vazire, S., Holleran, S. E., & Clark, C. S. (2010). Eavesdropping on happiness: Well-being is related to having less small talk and more substantive conversations. *Psychological Science*, *21*(4), 539–541. http://dx.doi.org/10.1177/0956797610362675

Muthén, B. O., & Asparouhov, T. (2011). Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. In J. Hox & J. K. Roberts (Eds.), T*he Handbook of Advanced Multilevel Analysis* (pp. 15–40). New York: Taylor & Francis.

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide.* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Nezlek, J. B. (2008). An introduction to multilevel modeling for social and personality psychology. *Social and Personality Psychology Compass*, *2*(2), 842–860. http://dx.doi.org/10.1111/j.1751-9004.2007.00059.x

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: The advantages of multilevel SEM. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*(2), 161–182. http://dx.doi.org/10.1080/10705511.2011.557329

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, *15*(3), 209–233. http://dx.doi.org/10.1037/a0020141

R Development Core Team. (2015). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org/

Reis, H. T. (2012). Why researchers should think "real-world": A conceptual rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 3–21). New York: Guilford Press.

Richard, F. D., Bond, C. F., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology*, *7*, 331–363. http://dx.doi.org/10.1037/1089-2680.7.4.331

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*(5), 609–612. http://dx.doi.org/10.1016/j.jrp.2013.05.009

Shrout, P. E., & Lane, S. P. (2012). Psychometrics. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 302–320). New York: Guilford Press.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 word solution. *SPSP Dialogue*, *26*(2), 1–4. http://dx.doi.org/10.2139/ssrn.2160588

Snijders, T. A. B. (2005). Power and sample size in multilevel linear models. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science* (Vol. 3, pp. 1570–1573). Chichester: John Wiley & Sons, Ltd. http://dx.doi.org/10.1002/0470013192.bsa492

Personality in Everyday Life

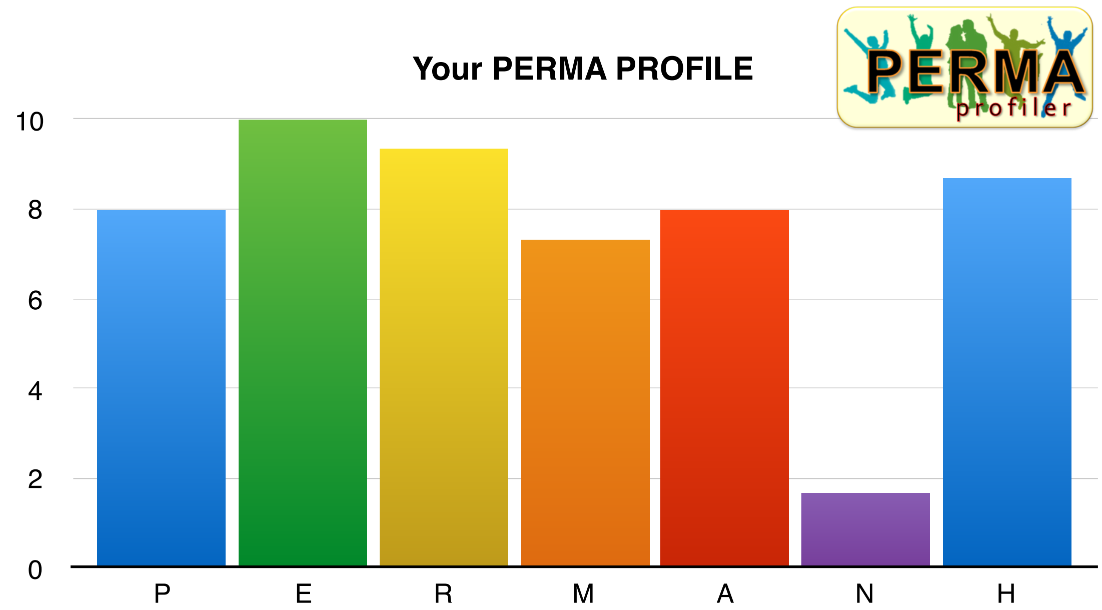
A Research Study conducted by Dr Luke Smillie and Ms Jessie Sun
Personality Processes Lab, University of Melbourne

THE UNIVERSITY OF MELBOURNE

Thank you for completing the Personality in Everyday Life Study. The primary goal of the study was to investigate the reasons why people tend to feel happier when they are acting more extraverted in everyday life, but we are also interested in learning more about the relationships between personality and other aspects of wellbeing.

Although there are many other definitions of wellbeing, Dr Seligman, founder of the field of positive psychology, defines wellbeing in terms of five pillars: Positive emotions, Engagement, Relationship, Meaning, and Accomplishment (PERMA). The PERMA-Profiler measures these five domains of wellbeing.

You completed the PERMA Profiler during the initial lab session, as well as the final survey. Here is your personal PERMA Profile, based on your scores during the initial lab session:



The profile includes the five PERMA domains, plus two others: negative emotions and health. Typically higher scores on P, E, R, M, A, and H, and lower scores on N reflect greater flourishing, but different profiles might be best for different types of people. Dr Margaret Kern (www.margaretkern.org), who developed the PERMA-Profiler, is currently researching optimal profiles.

Descriptions of each of these domains can be found on the next page.

P and N = Positive and Negative emotions

Emotions are an important part of our well-being. Emotions can range from very negative to very positive, and range from high arousal (e.g., excitement, explosive) to low arousal (e.g., calm, relaxed, sad). For Positive emotion, the PERMA-Profiler measures general tendencies toward feeling contentment and joy. For Negative emotion, the Profiler measures tendencies toward feeling, sad, anxious, and angry.

E = Engagement

Engagement refers to being absorbed, interested, and involved in an activity or the world itself. Very high levels of engagement are known as a state called "flow", in which you are so completely absorbed in an activity that you lose all sense of time.

R = Relationships

Relationships refer to feeling loved, supported, and valued by others. Having positive relationships with others is an important part of life feeling good and going well. Other people matter!

M = Meaning

Meaning refers to having a sense of purpose in life, a direction where life is going, feeling that life is valuable and worth living, or connecting to something greater than ourselves, such as religious faith, a charity or a personally meaningful goal. Meaning provides a sense that life matters.

A = Accomplishment

Accomplishment can be objective, marked by honours and awards received, but feelings of mastery and achievement are also important. The

For more information on Positive Psychology, see www.authentichappiness.com.

If you have any comments or questions about this study, please feel free to contact Jessie Sun at j.sun@student.unimelb.edu.au.