

Methodological Considerations and Assumptions in Social Science Research

(to be submitted to the *Journal of the British Academy* as JBA#2)

Daniel B. Wright
University of Nevada, Las Vegas

Abstract

The papers in this special issue are based largely on results from social surveys on beliefs and behaviours related to COVID-19. I describe the considerations and assumptions that we used when conducting this research and when analysing the resulting data. Data from a recent British Academy project on differences between the UK and US, and by ethnicities, with respect to COVID-19 beliefs and behaviours, are used as illustration. The aim is to provide a backstage view of how this approach to social scientific questions occurs. The scales appear to measure the constructions (e.g., Identity Resilience and Trust in Science) as intended, and these appear to influence reports of COVID-19 preventative behaviours.

When COVID-19 began spreading throughout the world, social scientists began studying the psychological effects of the pandemic on both individuals and societies. They attempted to measure behavioural changes due, in part, to guidelines from health researchers and politicians, and attempted to account for these behaviours using hypothesized psychological constructs.

There are several methods that social science researchers have used, but the primary one for much social science COVID-19 research has been some form of survey. This is the approach that our group has used for most of our research, and some of this is discussed in this special issue. Research findings—including our own—are usually disseminated through brief research articles that include a handful of statistics and a couple of plots with arrows connecting the key constructs. This succinct approach to sharing the results, which is efficient for many purposes, can make the decisions involved in conducting such research appear uncontroversial and it de-emphasises the considerations and assumptions underlying the approach. Here, I report some key findings to complement others in this special issue, but focus on the considerations that researchers faced when doing this type of research. These seldom are part of typical journal articles. As such, this paper will focus on more methodological concepts, and less on the psychological theories and their impacts on policies. See the first article in this issue (“JBA1”, 2023) for discussion of the theories.

The purpose of this paper is to describe, to the broad audience of the *Journal of the*

British Academy, the steps that social scientists who rely on surveys and questionnaires go through while conducting research. By making the steps explicit I hope to provide readers from other disciplines an understanding of not just our research, but also that of others using this approach. Further, I hope that the paper will bring these assumptions to the forefront for social science survey researchers and encourage them to question and to justify their processes/assumptions.

The following are, broadly, the considerations that our group to conduct our research and to analyse our data. I use these to structure the paper. This structure will not be appropriate for reporting all research, but it provides a useful way for readers to understand the research beyond a few summary statistics.

1. How to create a research team and to define the research questions/problems to be addressed?
2. What underlying theories and perspectives will inform the research?
3. What are the populations of interest?
4. What research design, including sampling, is appropriate?
5. How will theoretical constructs be estimated?
6. How will the relationships among these constructs be explored?
7. How to summarize/disseminate the implications for different audiences?

These considerations are listed in sequence, so for example exploring the relationship among theoretical constructs will only be appropriate for some research questions. All of these are inter-connected. The data used are from a British Academy funded project (CRUSA210025 to JR, GB, JB, and DW) where we examined the role of identity and other psychological constructs on COVID-19 related beliefs, like vaccine positivity, and behaviours, like hand-washing.

Research Team and Research Questions

Deciding the research team and research questions are often done in concert, and these affect how the rest of the research pans out.¹ A research team may exist and decide what research questions to address, or an individual (or funding organization) may describe some broad research questions and a team will coalesce around these and fine-tune the research questions. In most cases it is a combination of these. Our research team was based on having people with particular social psychological knowledge and different methods skills. We had worked on several research methods projects (e.g., Breakwell, Wright, & Barnett, 2020) and psychological theory focusing on identity (e.g., Breakwell, 2023b), risk (e.g., Breakwell, 2014), ethnic and cultural differences (e.g., Jaspal & Breakwell, 2023), and other pandemics including HIV/AIDS (e.g., Jaspal & Bayley, 2020). We had also begun work on psychological constructs and COVID-19 (e.g., Breakwell, Fino, & Jaspal, 2021; Jaspal, Fino, & Breakwell, 2022). When the British Academy offered funding on social science research related to ethnicity and COVID-19, our team came together.

¹I refer to research questions broadly, and mean simply seeking some information from the research that will change what the researchers believe on some topic. If a study turns out exactly as expected, the change would be greater certainty in the original beliefs. These might be applied problems that exist in the world, or specific questions that the researchers have about some theory.

Our primary research question for this project was how well Identity Process Theory (IPT) (e.g., see papers in Jaspal & Breakwell, 2014) could account for individual differences in vaccine positivity and self-reported likelihood of being vaccinated among different ethnic groups in the UK and the US (see “JBA3”, 2023). This topic was chosen because of the expertise of our team with respect to this theory and with respect to the methods needed to conduct the research. We believe identity resilience is a critical factor that influences people’s beliefs and behaviours. This, as discussed in the next section and in detail in “JBA1” (2023), is a perspective we bring to the research. The role of identity resilience and other constructs in predicting outcomes is complex and this framework allows us to build a model of these dependencies. It allows us to examine several research questions simultaneously. Our main research questions are about estimating the size of different relationships.

In some social science circles it is common to distinguish quantitative and qualitative approaches. This distinction is problematic and I believe it has harmed social science by making people think that there is a schism between the “camps” and this has limited what researchers can learn from different approaches. It is worth stating that these phrases are used differently in other sciences and in ways that benefit research in those sciences and could benefit the social sciences more if they were used this way in these areas. For example, sometimes people refer to quantitative hypotheses that predict specific outcomes (e.g., light bends a precise amount when passing a massive object as proposed by Einstein and evaluated by Dyson et al., 1920) and qualitative hypotheses that predict directions (e.g., the circumference of the earth is greater at the equator than around the poles, as predicted by Newton and evaluated by Boscovich, see Stigler, 2016). The later are more common in the social sciences, and this has consequences for how research should be done. Meehl (1967) describes how using null hypothesis significance testing, the p -value approach is used, is ill-suited for evaluating qualitative hypotheses. The qualitative/quantitative distinction is also used in statistics (e.g., Bartholomew, Knott, & Moustaki, 2011; Goodman, 1978) where the distinction refers to the type of data. Quantitative data have some underlying metric and qualitative data do not, and this informs the choice of statistical procedures (Stevens, 1946). The common social science use of the terms quantitative and qualitative research is less well-defined. These phrases are often used to describe broad characteristics that are often associated with one of these phrases (e.g., inferential statistics for quantitative; focus groups for qualitative), but there is disagreement about classifying research into this dichotomy (e.g., inferential statistics can be used with focus group data). Instead research should focus on the characteristics themselves.

An important point to make is that typical academic research questions are fairly specific. Particularly within the context of COVID-19 research, there is the realization that hundreds of other research groups are conducting projects addressing related research questions. The hope is that together, these form a tapestry to provide a more complete perspective. Further, because these studies occur with different samples and at different times, with different government guidelines and different information on social media, having multiple patches of this tapestry existing at each location allows more confidence in consistent findings.

Team. Social psychological theory. Experience in other crisis research. Methodology.

Questions. Using IPT to frame the relationship between psychological constructs and behavioural intent. Examining the individual components, and looking for differences by ethnicity in the UK and US.

Our Perspectives/Biases

All research is influenced by the perspectives of the researchers. Francis Bacon (e.g., 2019 [1620]) recognized this, describing how the *idols* of mind could distort how we interpreted the world. His advice was to avoid these prejudices, and that nature would more truthfully reveal itself. Our beliefs affect how we perform research and interpret research (e.g., Breakwell, 2023a), but completely removing all biases is neither possible nor desirable as we would be unable to perform research or interpret results (e.g., Popper, 1994, Ch. 4). The legacy of Bacon’s desire to be without bias, coupled with the observation that society has achieved some (perhaps much) of his vision of the science-produce industrial society prophesized in the *New Atlantis* (Bacon, 2020 [1626]), has lead to the myth that scientists have successfully removed their beliefs, prejudices, and biases from the scientific process. In the social sciences, where this myth is even more far-fetched and where there seems to be the desire to appear to do research more like how some believe the natural sciences work, this has shaped how research is discussed.

For decades social and natural science students have been told to write in a formal style, for example saying “The author did ...” or “*AuthorName* did ...,” rather than “I (or We) did.”² The intent was to distance the research from the researchers, downplaying the influence of the researchers. It was not about making the research more objective, but instead to describe the research as if it were more objective, as if Bacon’s goal of removing the idols of mind had been achieved. Consider for example de-personalizing research papers. Campbell summarized the value of the writer owning their statements:

Here a moral issue is raised. If we are not prepared to make a personal statement in a personal form, are we justified in making it at all? Can true modesty, or any other virtue, permit me to occupy valuable space in airing my views and yet forbid me to call them mine?

Campbell (1928, p. 1021)

Like Bacon, I recognize that people bring their beliefs, prejudices, and biases and that these affect their research, but I believe the formality of much scientific writing appears to discount these influences. People choose what to study and how to study it, and interpret the observed data using what they already know. Researchers should be both knowledgeable and interested in what they research, and these biases should be influential. But researchers should remain skeptical of any individual findings, particularly their own. They should be “bending over backwards to show how [they] maybe wrong” (Feynman, 1974, p. 12). They should consider other perspectives. As noted above there are issues with the qualitative/quantitative distinction as often used in the social sciences. So called qualitative research methods books stress that the authors should describe their perspectives, what is sometimes called *reflexivity*. This is not discussed in most, so called, quantitative

²Blind review means authors often refer to themselves in the third person when their papers are originally submitted, and usually this is not changed for later versions.

textbooks. The schism into these camps coupled with the history of telling students, when writing, to pretend that their perspectives do not affect their research, may be a reason that this is seldom discussed in some areas of the social sciences.

People writing about empirical research should describe their beliefs in enough detail for readers both to understand why they are conducting the research that they are, and also how their perspectives may influence how they interpret the findings. Writers cannot describe all their biases both because this would be too great a task and because it would get repetitive across publications. They should use their best judgments to select those that they think are most relevant.

With respect to our group's research, within a survey context, we believe that respondents' answers to our survey questions provide information about their thoughts. This implies that they have some kind of conscious access to this information. It is known that for many tasks humans do not have access to why they make decisions (e.g., Nisbett & Wilson, 1977) and that alternative approaches can be necessary to tap certain information (e.g., Greenwald, McGhee, & Schwartz, 1998). Our assumption is that asking people their views, while not providing perfect access to their beliefs, provides responses that are similar enough to their beliefs to be of use. Specifically, we believe responses to questions that we believe *a priori* are related to a construct can be used to construct an estimate for each person for that construct. This assumption should not be taken lightly and while researchers making this assumption do usually empirically examine some aspects of these assumptions, any conclusions still rely of the basic validity of this assumption.

In addition, it is important to state that our beliefs influence the constructs that we choose to examine. Consider our choice to use a "Trust in Science" construct. We work in academia and while we are critical about all research (scientific or otherwise) and thus want multiple patches for the research tapestry as mentioned above, we believe that the scientific process is better than existing alternatives for allowing wise decision making. In fact, we sometimes get annoyed with those that do not share this belief. Our belief in the value of the scientific method has influenced our choice of including this construct and our *a priori* belief that it plays an important role in adherence to health guidelines.

We trust, to some extent, the scientific literature and rely on this for our methods. For example, we use a trust in science scale by Nadelson et al. (2014) where they report the psychometric properties for their scale with their sample. The psychometrics they report are specific to their sample. A scale does not have, for example, a particular level of reliability. It only has this in relation to a sample. As such, we assume that the scale should have good qualities for our samples, but we do examine some, but not all, aspects these.

Another belief that we have is that IPT provides a good framework to examine the relationships among variables including identity related variables. This includes developing the theory/framework and using it in many prior publications (for a review see Breakwell, 2015). This study was not designed to test IPT as a theory, but to use it as a framework to examine various components across respondents from two countries.

Self-report. While recognizing that self-reported beliefs can be biased, we believe that the responses to the scales we use do reflect associated psychological constructs and can be used to construct useful estimates for these constructs.

Beliefs about Health Guidelines. There are disagreements in both the UK and the US

about both details of the government health guidelines and trust in these institutions. Identity Process Theory. We believe identity, as well as the other psychological constructs that we choose to measure, relate to each other and to vaccine positivity.

Population of Interest

The funding of our project was to examine differences by ethnicity for people in the UK and US, applying IPT to COVID-19 beliefs and behaviours. We are also interested in the relationship among these constructs in other countries, but for this project the UK and US are our main interests, though as part of this project we did examine data from several countries (e.g., Jaspal, Breakwell, & Wright, in press; Wright, Wolff, Jaspal, Barnett, & Breakwell, 2022). We are interested in the general population. We focus on adults *what is on Glynis notes*** in most US states parental consent is required for COVID-19 vaccines (<https://www.kff.org/other/state-indicator/state-parental-consent-laws-for-covid-19-vaccination/>). Avoiding people under 18 years old also makes conducting the research simpler as many ethics guidelines require people under a certain age to provide parental consent to take part in studies. While there is interest in how several specific groups have reacted to health guidelines, vaccine positivity, *etc.*, for example religious groups, right-wing conspiracy advocates, those with existing medical conditions, here our focus is the general population. It is important to note that other projects funded as part of this British Academy initiative did look at more specific groups (see <https://www.thebritishacademy.ac.uk/news/the-british-academy-publishes-studies-examining-covid-vaccine-engagement-in-uk-and-usa/>).

Population of interest. People 18 years and older in the UK and US with large enough sub-samples of different ethnic categories to allow comparisons.

Research design

Target population and sampling

In introductory research methods classes students are told that the statistical procedures often assume that a simple random sample of the population of interest is taken.³ This is not plausible in most research. First, all surveys of this type, in the UK and US, are voluntary and therefore many people will not take part. Further, to conduct a simple random sample you need a list of the population. While electoral roles provide much of the population, many would still be missed.

Because the samples will not be perfectly representative of population of interest, making inference from the sample to the population requires assuming that it is close enough that the conclusions will still be valuable. Importantly, inference in social science is usually not about the absolute level of some variable, like it is when predicting election results from surveys. Usually it is about comparisons among groups or across time, where similar sampling procedures have been used for different groups/time periods. Sometimes experiments are embedded within a survey where respondents are randomly allocated into

³A simple random sample, of size n , is defined as one where all possible samples of size n are equally probable. Most lotteries are designed to be this, though see Fienberg (1971).

different conditions. This allows causal inference about how differences in the ways that the conditions are treated to be made on measures after the manipulations have occurred.

Thus, we conducted our survey knowing that attaining a truly representative sample was impractical. Further, as it was conducted during the COVID-19 pandemic in person interviewing was not an option. We chose to use an online survey. We constructed the survey using the popular tool Qualtrics (<https://www.qualtrics.com/>) used. This allowed a link for the survey to be posted using Prolific (<https://www.prolific.co/>) which allows us to set up filters (e.g., we wanted UK and US respondents, with quotas that allowed us to have enough respondents in several ethnic categories to allow comparisons). Online participants have a reputation for not being as focused as those answering questions in person, though the several studies have shown that they produce data at least of comparable quality to in person administration (e.g., Sheehan & Pittman, 2016). The online aspect of the instrument allows some of the respondents to be excluded. For example, we removed duplicate IP address because people may have talked, we included attention checks, *etc.*

It is important to discuss the exclusion criteria that are used in any study and how missing and odd values are dealt with. With online studies, the IP (internet protocol) address is usually available, and we exclude duplicate IP addresses both because this may be the same person who is using two Prolific accounts or otherwise it is likely this is a housemate who may have discussed the content of the study. Extremely fast responses are also often removed. Attention checking questions, which often include a phrase like “ignore the rest of this question, just tick the option B” (see Gummer, Roßmann, & Silber, 2021, for detailed discussion), are also often included. There are numerous guidelines for constructing online surveys (e.g., Biffignandi & Bethlehem, 2021). With online surveys it is possible to force people to respond to each item. If the question is poorly worded or there is some other reason why the person feels it is inappropriate to respond to one of the items, this can be annoying and affect the quality of all subsequent responses by the person. However, for the types of scales that we use in this study having people provide an answer for all questions is useful. There are methods to address missing data (e.g., Rubin, 1987; van Buuren, 2018), but if it is believed that each person can provide a meaningful response to an item it is worth having that response. Odd values (sometimes called outliers, but the word “outlier” is sometimes used in a more technical way) also arise, particularly when people type in their responses. We had people type their age in years into the survey. Some appeared to write their year of birth. We treat those as missing. Similarly, when we asked people what confused them about COVID-19 guidelines, some responses were difficult to interpret.

Which Demographics?

The demographics that we were most interested in are: ethnicity, sex, age, education, and we asked some questions about political affiliation. We also asked questions that are of particular importance to the COVID-19 guidelines, including the number of people in one’s household, because this is related to number of contacts and therefore the possibility of contagion. How demographic questions are asked and which categories are included in the response alternatives can be very contentious. The meaning of, for example, ethnic categories differs between the UK and the US. We tended to follow the ways in which our government surveys ask these questions as well as the phrasing in both our survey program (Qualtrics) and the sampling program (Prolific). With ethnicity it is clear both that there

is not a clear way to differentiate all people and that there is much ethnic variety within any of the categories that we choose. For ethnicity we had Prolific perform a quota sample for the categories they use for ethnicity. Quota sampling means that you attempt to get a pre-determined number of people (a quota) for each category (Kalton, 2021). The breakdown we achieved is shown in Table ??.

Need to make sure that this is being done the same way as in the ethnicity article.

```
with(p3labels,table(UK_Ethnicity_Categorical))
with(p3labels,table(ETHNICITY))
with(p3labels,table(US_ETHNICITY))
```

The gender breakdown was: 940 (50%) female, 928 (49%) male, 11 (1%) other, and there was one missing value. The survey asked respondents for a number for their age in years. Two said they were nine years old (to have a Prolific account they must be 18 or older), three said they over 100 (listing their likely year of birth), and one left the age variable blank. Excluding these the median was 32 years old and the mean was 34.43 years old. The skew towards younger responses is predicted as Prolific began in universities and was seen as a convenient way for students and recent alumni to earn extra money. More details of the demographics are covered in the articles dealing with those. Here the only demographic comparisons were by country: 1117 were from the UK and 763 were from the US.

Why R?

We use the free statistical environment R (R Core Team, 2022, for a brief description see Chambers, 2009; for a thorough description see Chambers, 2008). This is one of the most used system for data; it has been described by Mizumoto and Plonsky (2016) as a *lingua franca* (a shared or bridging language) for both learning and implementing statistics. There are advantages and disadvantages to all statistical environments, and no system will be best for all situations. It suits our purposes for this paper well. This document I wrote by combining R code within a L^AT_EX document, and used **knitr** (Xie, 2015) to combine these to create the paper in a more readable format. One of the concerns about the statistics in empirical articles is not being able to replicate the findings. We include at [github.***](#) the final submitted L^AT_EX document so anyone (both R and L^AT_EX are freeware) can recreate the document.

Target Population. Prolific sample of UK and US respondents. This will not produce a perfectly representative sample of the population of interest. No practical method will. The choice is pragmatic. Quotas will be used to allow comparisons of ethnicities (see discussion in “JBA3”, 2023).

Items. Several existing scales will be used, as well as some common demographic questions.

Analysis. The environment R will be used for analysis.

Estimating psychological constructs

Scientists construct models that:

1. They believe approximate nature closely enough to be useful and
2. They believe provide a useful framework to interpret their findings.

Their choice is often influenced by the statistical methods they use, but these statistical methods also influence their theories (Gigerenzer, 1991). A popular model that social science researchers assume is the latent variable model and as will be clear in this section this relates to both theories and methods. Here are described two common ways to model data from social science scales.

The scales that I use are: the 16-item Identity Resilience Index (Breakwell, Fino, & Jaspal, 2022), the 12-item version of the Interpersonal Support Evaluation List (S. Cohen, Mermelstein, Kamarck, & Hoberman, 1985), the 10-item COVID-19 Preventive Behaviours Index (Breakwell et al., 2021), six items from the Trust in Science scale (Nadelson et al., 2014), and sets of questions on Trust in six Authorities and Trust in six Sources of Information. Each of these has been discussed elsewhere and been shown for those samples to produce quality responses. However, when social scientists use scales in their own contexts the norm is to check at least some of the qualities of the scale in this different context. Here I describe in more detail how we examined The Trust in Science scale for our research. The items from Trust in Science are listed in Figure 1.

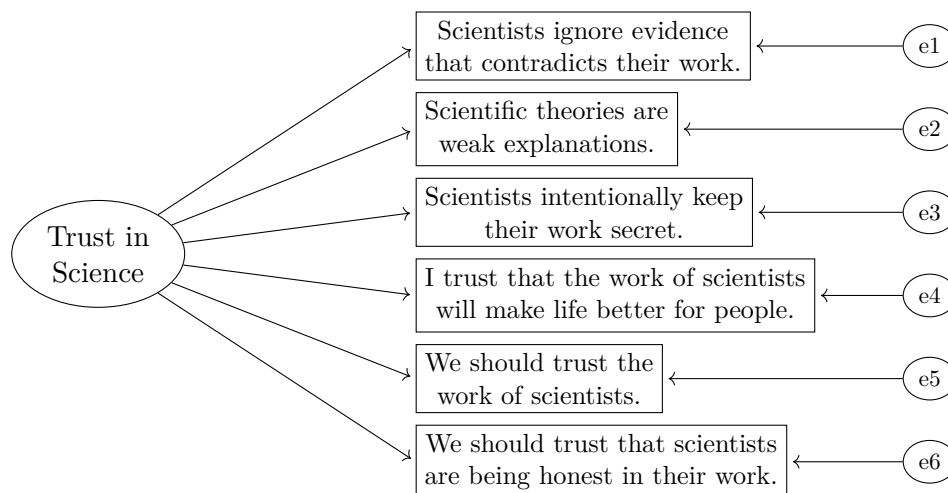
A Latent Variable Conceptualization

An assumption of much social and psychological research is that responses to several related items can be aggregated to estimate a single construct. A latent variable model is often assumed for this purpose. Loehlin and Beaujean (2017) provide an excellent introduction to latent variable models, mathematical details are in Bartholomew et al. (2011), and Spearman (1904) is the seminal historic text. Suppose that you have six variables and that you believed each is related to a particular construct, say *Trust in Science*. Figure 1 shows a latent variable model that might be used for this and the six items that we had respondents answer on a 1 to 5 scale from strongly disagree to strongly agree. The arrows mean that what is described in the node at the neck of the arrow *influences* what is described in the node at the arrow's head. The assumption is that responses to each question, for example the *Scientists ignore ...* rectangle, are influenced by variation in a respondent's *Trust in Science* construct. In addition, responses are also affected by a combination of idiosyncratic aspects of this item and random variation, shown by the *e* nodes to the right of each rectangle. These are often called the error terms associated with the individual items, but it is important to note that they are a combination of error and systematic variation specific to the item. For the model shown in Figure 1, these error terms are assumed independent of each other. This means that after taking into account *Trust in Science* the variables themselves are independent. There are ways to examine if this assumption is justified, discussed below. In these plots, the most popular convention is to have the latent variables shown in ellipses and the observed variables shown in rectangles.

Not all the sets of items fit well with this conceptualisation. Two sets of questions ask respondents how much they trust different authorities and different sources of information. As we discuss below, a different conceptualisation is useful for these.

Figure 1

Assumed relationship between a psychological construct, Trust in Science, and six survey items.



Dimensions versus Categories Assumption

Within the latent variable framework, the latent and observed variables can be based on dimensions or categories (Bartholomew et al., 2011). It is often assumed that values on the underlying constructs vary along continuous dimensions. While there are methods to examine this assumption empirically (e.g., Waller & Meehl, 1998), the dimensions assumption is common in much social science and I make this assumption for the Trust in Science scale. Further, the data are often not able to distinguish these conceptualisations (Bartholomew, 1993), so often the choice of treating the latent constructs as dimensions or categories is decided by what the researchers feel is the better conceptualisation for their purposes.

In addition, it is also often assumed that responses for each of the individual items vary along a dimension, even when the items are presented with a small set of discrete response alternatives as here. There are alternatives appropriate when it is assumed that the item values are categorical or ordinal (e.g., item response theory, see Embretson & Reise, 2000). I make this assumption, but accept that others may not.

When creating scales that other people will use, often those people will have small samples or even just a few individuals, and want to create summary measures. This means that using complex methods that require large samples to estimate values for people's constructs may not be practical. Scale designers take this into account and try to design scales where taking the mean of responses provides a good estimate for the construct. There are advantages to having a simple method for allowing others to estimate these psychological constructs and often scales are developed such that the mean of the responses provides a good estimate. This is like how teachers often report the percentage of correct responses on an assessment for their students. This will likely not have as good statistical properties as the more complex procedures (McNeish & Wolf, 2020), but in some contexts it is a good option available (Widaman & Revelle, 2022).

Measurement Qualities

Researchers are interested in the reliability, validity, and fairness of their scales (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). Reliability is whether the instrument produces similar responses each time it is used in the similar contexts. Validity has a few definitions, but here it is used to say whether the aggregate score is appropriate for the intended purpose (Kane, 2013). Fairness is whether the scale and its items produce similar values for people who have similar values for the latent construct, but are from different groups. For example, a scale designed to measure botany for all plants, but with 90% of the items on the assessment being desert plants, would be unfair as this would be an advantage for people living in the desert regions.

Each of these qualities should be examined for any measurement. The first step for all three of these is to think about the questions and how respondents are likely to answer the questions as opposed to how the designers intended the questions to be answered. Much of this work is done when the scale is being developed. When scales are introduced authors usually publish characteristics of these scales that they found when applied to their data. This does not mean that these characteristics will be the same with other samples. The assumption usually made is that the scale will have similar qualities when used with other samples. When using other scales often only reliability is empirically examined. This is probably because examining reliability is easier to do than validity or fairness. Below I report the most common measure of reliability, Cronbach's α , also called Guttman's λ_3 (Cronbach, 1951; Guttman, 1945), but discuss its assumptions.

Before progressing, it is worth saying that the expected relationships for some of the items may be negative. This is often done by the scale developers in order to identify any respondents who are just ticking all the responses at one end of the scale without reading the questions. It is often desirable to reverse score the items that are designed to be in the opposite direction of the construct. For example, the first three items in Figure 1 should have low values for those who trust science. There is a simple rule for reverse scoring items. If an item goes from m to n , letting $newvar_i = (n + m) - oldvar_i$ means the minimum and maximum are the same as the other items.

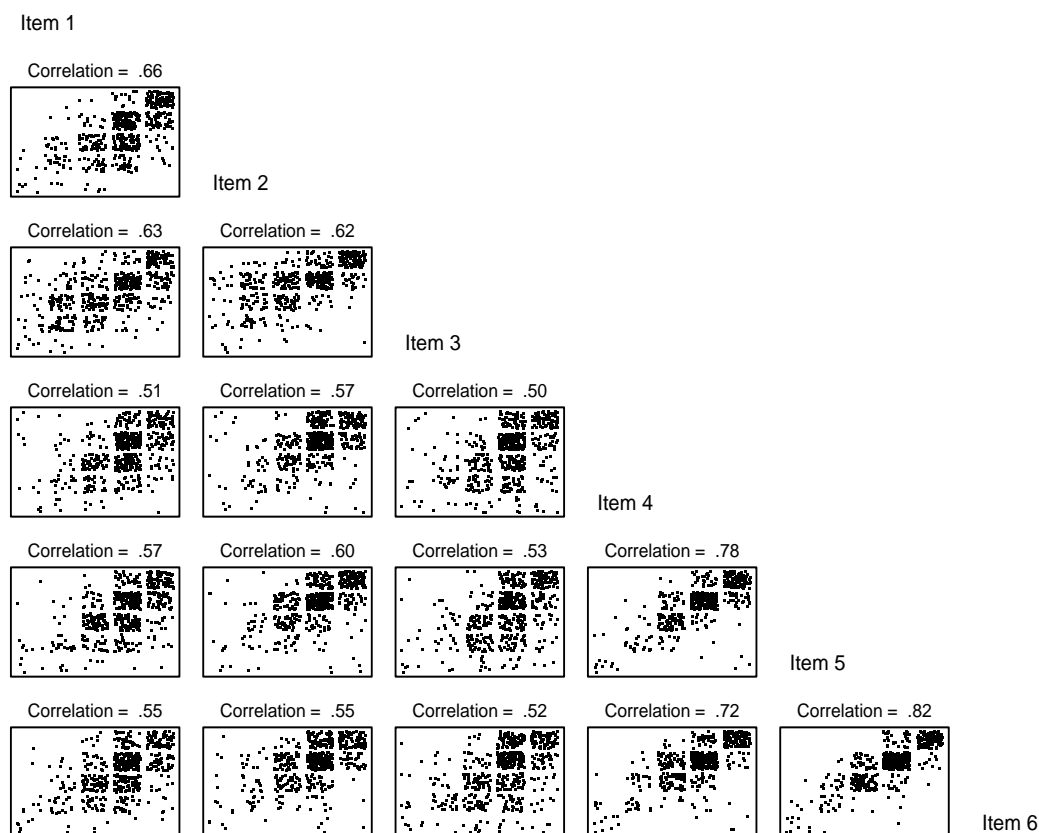
Exploring dimensionality with the scree plot

The assumption in Figure 1 is that a single latent variable influences all of the observed variables. The assumption is that the scale is uni-dimensional. This assumption can be examined empirically. Exploratory data analysis should be conducted, including looking to see if all items are correlated as they should be, prior to creating any latent variables. This can be done with both a statistical test, like Pearson's correlations, and visually with scatterplots. Figure 2 shows the scatterplot matrix for the Trust in Science items with their correlations. The scatterplots allow outliers to be identified and to check when a straight line seems to describe the relationship well. With typical survey items that are measured on discrete rating scales it is useful to add a small random variable to each point so that each point can be seen. This is called *jittering*. In addition, only 600 of the data points are shown in order to make identifying which coordinates have the most values easier. With most social science applications, the data points are spread out so trying to tell

if a pattern is approximately linear is difficult. At this point of the analysis the researcher is usually looking only for clear signs of non-linearity (e.g., is there a floor or ceiling effect) or if the relationship is not monotonic.

Figure 2

Scatterplot matrix of the six Trust in Science items. Six hundred cases were randomly chosen and jittered so that it is easier to see the relationships.



Two things can be concluded from these scatterplots. First, there are more responses above 3 (the mid-point on the 5-point scale) than below 3: 66% compared with 12%. Thus, our sample is more trusting in science than distrusting, but there is a spread in responses. Second, the correlations are all at or above $r = .5$. What does this mean? J. Cohen (1992) describes $r = .5$ as *large*, so in his terminology all of the associations are large, but his terminology is context dependent. Using Figure 1 as a way of describing the size of correlations, suppose that there is a normally distributed latent variable with a standard deviation of 1 and two observed variables that are this variable plus an item specific error variable with a standard deviation of 1. Provided that these two error variables are unrelated, the correlation between these two variables would be about $r = .5$. If there is

less error, say $\sigma = 1/2$, then the correlation would be about $r = .8$. Looking at the spread of the data in the scatterplots also helps you to get a feel for what the different values of r mean with respect to how spread out the data are. You would also want to identify any pairs that had particularly high or low values.

One of the most used methods to examine the number of dimensions of a set of items is a *scree* test (Cattell, 1966). Scree is the geological term for the loose rubble that has accumulated at the base of a steep hillside. The metaphor is that the hillside represents the latent constructs (e.g., Trust in Science) and the rocks in the scree are the e in Figure 1, which are assumed to be a combination of random error and idiosyncratic aspects of those items. Cattell likened the underlying structure of a scale to the hillside and the “scree represents a ‘rubbish’ of small error factors” (Cattell, 1966, p. 249). The statistics to construct the scree are usually calculated using the eigenvalues of the correlation matrix. The sum of the eigenvalues of most correlation matrices is the number of variables. The first eigenvalue, which will be the largest, shows how much of this total can be accounted for by a linear combination of the variables or in lay terms how much of the variation can be accounted for by a single dimension. The second is how much more can be accounted for by a second dimension, and so on. A scree plot is made by drawing a line connecting the eigenvalues. If there are six items, there will be six eigenvalues providing no item is a linear combination of the other five. The amount accounted for necessarily decreases with each dimension. Cattell notes that using the scree “requires the acquisition of *some* art in administering it” (p. 256, *emphasis* in original).

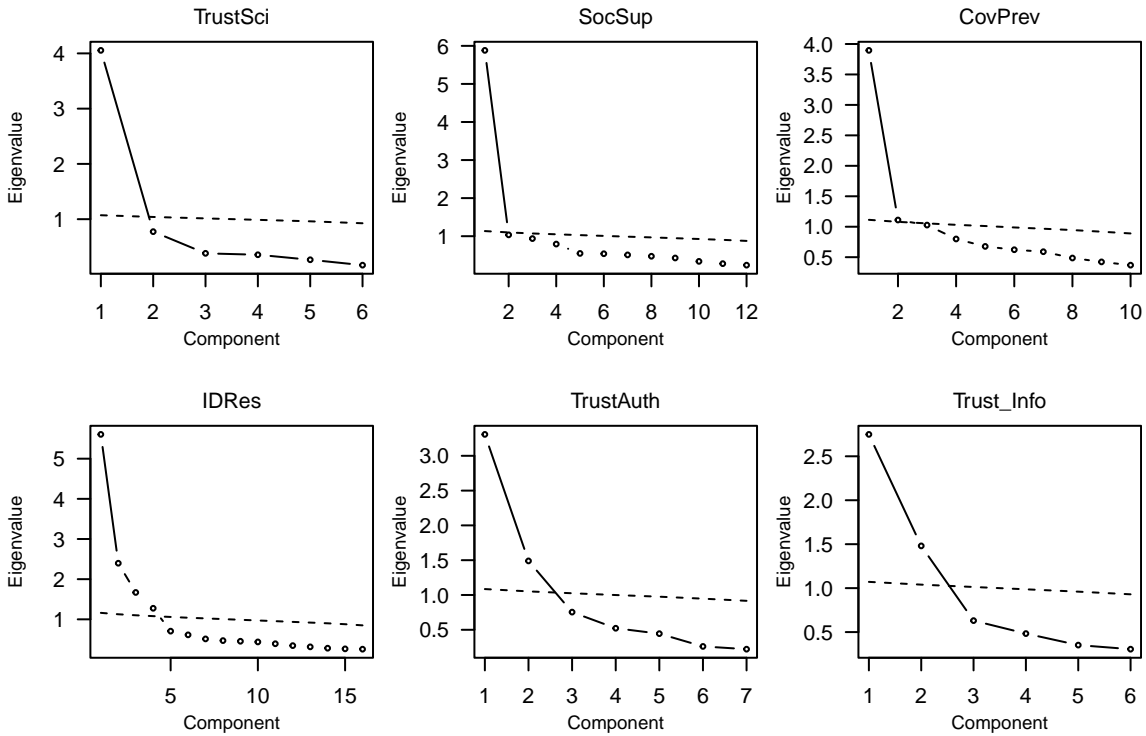
There are several procedures that can be helpful to guide this art. The most useful in my opinion is adding a line for how the scree would look if there were no structure to the data. This is called *parallel analysis*. To be part of the hillside you would want to use the dimensions shown on the scree that are well above the random line. What “well above” means is up to the discretion of the analyst. Velicer, Eaton, and Fava (2000, see also Auerswald and Moshagen, 2019) described several statistical procedures that aim to identify the number of dimensions, and some of these are used below. It is worth noting that reality is much more complex than our models, and that a near infinite number of likely related constructs will inform how people answer any of these questions. Cattell was aware of this: “There is no such thing as ‘the true number of factors to extract,’ since the only possible assumption is that both the number of substantive and the number of error common factors each exceed n , the number of variables” (Cattell, 1966, p. 273). The analyst must decide what is appropriate simplification for their purposes to allow them to make what they believe are wise decisions.

The scree plots are shown in Figure 3 along with lines created to show what the scree would be like for random data. These were made using `fa.parallel` from the **psych** package (Revelle, 2018). Figure 3 shows the Trust in Science, the Social Support scale, and the COVID prevention scale can be well described with a single dimension as a single eigenvalue stands out above the random scree lines. The Identity Resilience scale is, as discussed in earlier manuscripts (e.g., Breakwell et al., 2022), multidimensional with four eigenvalues being above the random scree line. This will be explored in more detail below. The final two, Trust in Authorities and Trust in Sources, are not unidimensional. As discussed below, the way in which respondents likely answer these questions mean a different approach can be useful. The DIMTEST function from the R package **EFA.dimensions**

(O’Connor, 2021) was used to help with the decision. It produces, among other statistics, the empirical Kaiser criterion. It suggests that the Trust in Science scale, the Social Support scale, and the COVID prevention scale can be well described with a single dimension, the Identity Resilience scale with four dimensions, and the Trust in Authorities and Trust in Sources of Information with two dimensions each. These coincide both with my view of the scree plots (and this will often occur as they are based on the same information) and my beliefs about the scales prior to conducting the research.

Figure 3

Scree plots for the six scales examined.



It is worth examining the Identity Resilience measure further. Previous work (Breakwell et al., 2022) has found this scale to have four components, the scree suggests four dimensions, and we used four items designed to tap each of these purposefully. Confirmatory factor analysis (CFA) can be used in this situation. In addition, it is believed that there is still an underlying Identity Resilience construct that influences all the items, but each item is also influenced by one of the four components. Thus, the first item can be thought of as:

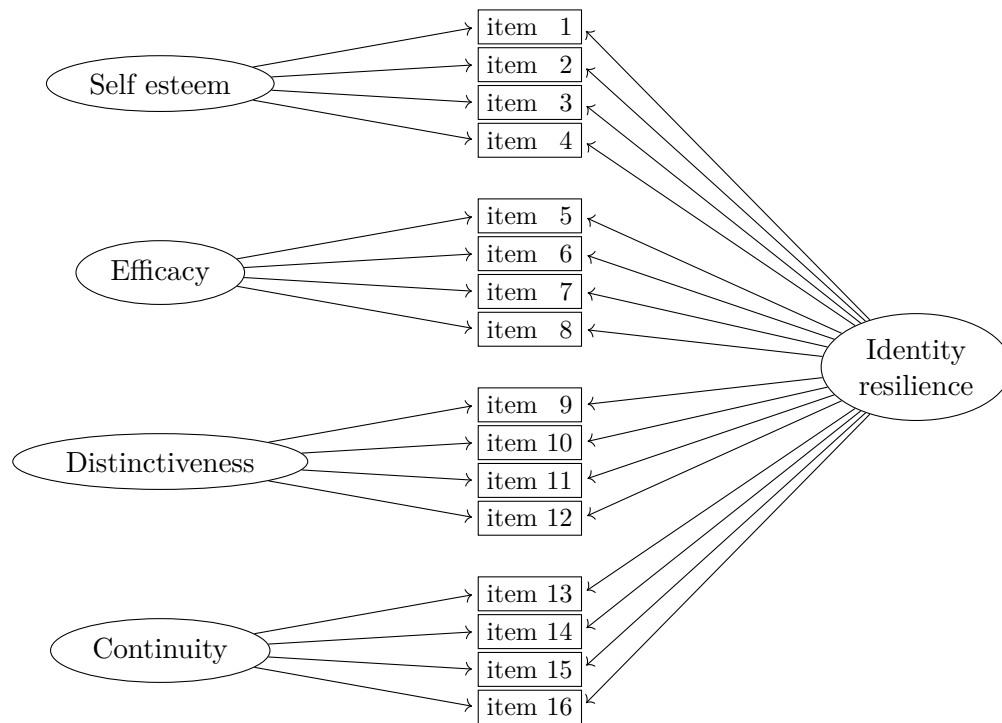
$$\text{item 1} = \text{Identity Resilience} + \text{Self-Esteem} + e_1 \quad (1)$$

This is called the bi-factor model and is depicted in Figure 4. It shows each of the four proposed components of identity resilience influencing four items (listed on the left side of the figure), and all 16 of the items being influenced by some overall identity resilience construct. In addition, each item has a unique e term influencing it. These are sometimes

depicted using curved arrows going from each of the 16 observed items onto itself. These are not shown as that would make the figure look cluttered. I compared the fit of this model with several alternatives, and this model fits better than the alternatives tested. For subsequent analysis in this paper, the single Identity Resilience construct is used. The values for these constructs were estimated using the R package **lavaan** (Rosseel, 2012). There are other ways to conceptualize the four components and an over-arching identity resilience variable. For example, a second order model has an over-arching variable influencing each of the four components, which then influence the observed variables. The bi-factor model fits better with our conceptualization and also fits these data better. There are many statistics that are used to compare the fit of models. Here the Bayesian Information Criterion (BIC Schwarz, 1978) is used. Lower values mean the fit is better. The BIC for the bi-factor model is 72,281, the BIC for the second order model is 72,647. The BIC value is lower for the bi-factor model so according to this statistic it is the preferred model.

Figure 4

The bi-factor plot for the identity resilience scale.



Cronbach's α and alternatives

A popular statistic is called Cronbach's α (Cronbach, 1951), though it was introduced prior to this as Guttman's λ_3 (Guttman, 1945). While often used, it is also often misinterpreted and inappropriately applied McNeish (2018). It assumes that there is only a single underlying dimension/construction and that the construct influences all the items

equally. Psychometricians call this τ equivalence. It provides an estimate for the lower limit of the reliability of the scale. Hayes and Coutts (2020) note that McDonald’s (1978) ω is more appropriate when τ equivalence is not met. Both α and ω are measures of reliability and are reported in Table 1. Neither of these directly measures the internal consistency of the items. They both increase with the number of items, which is expected for measures of the reliability of the whole scale (i.e., all other things being equal, a test of 50 items is more reliable than one with 5 items). The median correlation of the pairs of items is also reported as is the proportion of variation of the first eigenvalue. Based on these, the scree plot, and the statistics based on the scree, single latent variables were estimated for the Trust in Science construct, the Social Support construct, and the COVID Prevention construct. Further, the overall factor from the bi-factor model shown in Figure 4 was estimated.

The mean of the values for each scale (after reverse coding where appropriate) was found and we correlated these with the constructs that we estimated using these latent variable models. The Trust in Science estimate has $r = .966$ with the mean of those responses, Social Support has $r = .990$ with its mean, and COVID prevention has $r = .984$ with its mean. The overall factor Identity Resilience factor has $r = .932$ with its mean. The correlation is lower as the four components account for additional variation (if a single factor is calculated in the same manner as the other it has $r = .961$). Thus, if future users of these scales only had a couple of people filling them out and just took the means for their responses, they would measure similar constructs.

Table 1

Measures for the different scales and the four components of the Identity Resilience scale. The measures are Cronbach’s α , McDonald’s ω , the median correlation, and the proportion of the total variation accounted for by the first eigenvalue (λ_1/k).

	α	ω	median r	λ_1/k
Trust in Science	.903	.891	.573	.676
Social Support	.904	.903	.429	.490
COVID Prevention	.812	.824	.309	.390
Trust in Authorities	.811	.802	.338	.472
Trust in Information Sources	.748	.748	.270	.458
Identity Resilience	.869	.870	.266	.350
Components of IR				
Self esteem	.880	.881	.644	.735
Efficacy	.792	.806	.468	.623
Distinctiveness	.853	.853	.604	.694
Continuity	.808	.808	.532	.635

A Clustered Items Conceptualization

The assumption in Figure 1 is that information flows from the latent variable to the individual items, and not directly between the individual items. This may be an adequate approximation for the *Trust in Science* scale, but is unlikely to be for two of the other scales. Consider two sets of questions that we asked our respondents in Table 2. When deciding

useful models for survey items you are modelling how you think respondents are answering the questions so it is wise to have a cognitive model of the respondent (Belli, Conrad, & Wright, 2007; Loftus, Fienberg, & Tanur, 1985).

Here, it is assumed that in the UK and US respondents are aware that both accurate and inaccurate information is present and that authorities/sources vary by their accuracy. In fact, most will realize that our goal in asking this set of items is to estimate which they believe are the most and least trusted. Respondents likely will believe their task is to compare the different sources. This implies a more complex model than that shown in Figure 1 is likely necessary, and the scree plots (Figure 3) showed simple uni-dimensional models did not fit the responses well. For the authorities, respondents will likely categorize the politicians together, the medical professionals together, the scientists and academics together (within the context of a COVID-19 survey most of the academics expressing opinions in the media have been scientists), and we are unsure where journalists will fit within most respondents' minds. It is worth stating an additional assumption. I am aware, for example, that most people will trust *some* politicians and distrust *some* politicians, and similarly for the other groups. I am assuming that respondents are answering with respect to some prototypical politician or scientist or medical professional or journalist. As for the sources of information, it is likely the the NHS/CDC (the former was used in the UK survey, the later in the US survey) and health profession websites are likely to be classified together and with the medical professionals.

Table 2

Items for the Trust in Authorities and Trust in Information Sources scales.

Trust in Authorities	Trust in Information Sources
Here is a list of different professional groups. For each please indicate how much you generally trust them to tell the truth. (1=not at all, 5=very much)	How much do you trust each of these sources? (1=not at all, 7=completely)
Doctors	WhatsApp
Nurses	Blogs
National politicians	Social networking sites (e.g., Facebook, Twitter)
Local politicians	Professional health information websites (e.g., WebMD)
University academics	The NHS (or CDC for US sample) Website
Journalists	
Scientists	

There are many statistical methods that place people and/or variables into clusters. Many of these are popular in the relatively new branch of statistics called machine learning. Here two older approaches are used to illustrate methods of placing the twelve different variables into clusters.

Conducting Cluster Analysis

Cluster analysis is described in multivariate statistics textbooks including Bartholomew, Steele, Moustaki, and Galbraith (see Ch. 2 2002, as well as numerous other

multivariate statistics textbooks). There are two main types of cluster analysis: hierarchical and non-hierarchical. With hierarchical analysis a distance matrix between each pair of items is entered into the program and a tree-like structure is returned that branches between dissimilar items and keeps similar items together. With non-hierarchical cluster analysis the analyst tells the computer how many clusters to have. Usually this choice is informed by comparing the fit of models with different numbers of clusters. Bartholomew et al. offer two important bits of advice:

1. “we shall recommend that several different methods should be used, since well-defined clusters are likely to show up using any reasonable method” (p. 22).
2. “the usefulness of the method is to be judged by what it produces rather than by the assumptions made along the way” (p. 22).

Both hierarchical and non-hierarchical approaches will be shown here.

We will combine responses for the Trust in Authorities and Trust in Sources of information scales. As the first set was based on a 1–5 scale and the second a 1–7 scale, we standardized these by dividing by each variable’s standard deviation. The Euclidean distance between each pair of the 13 items is calculated and this is entered into R’s hierarchical clustering function `hclust`. Euclidean distance is generally used unless there is a specific reason to use an alternative. Cluster analysis can be used to categorise the variables or the respondents. Both can be useful. Here it is used to categorise the variables.

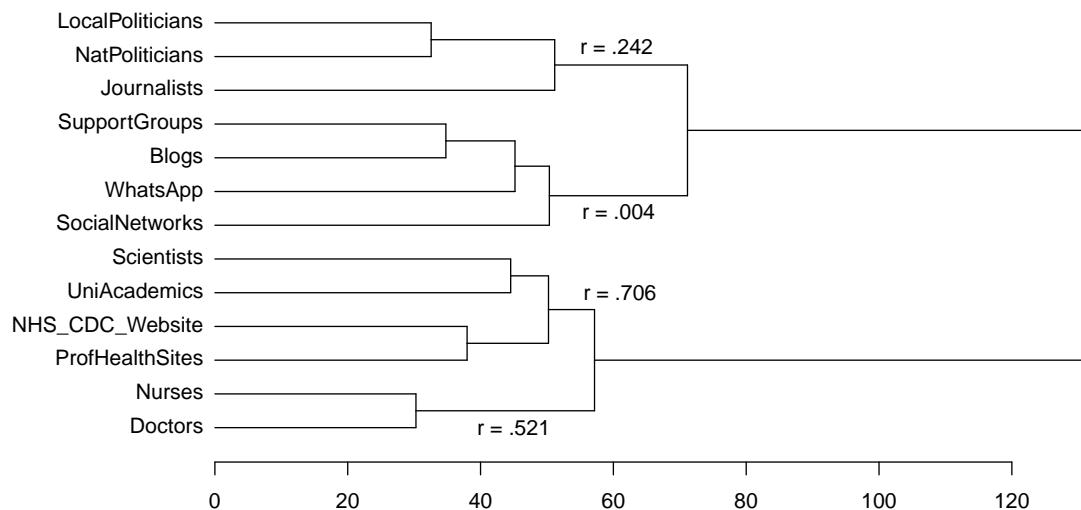
The main plot from a hierarchical cluster analysis, called a dendrogram, is shown in Figure 5. It is made here using the `plot_horiz.dendrogram` from the R package **dendextend** (Galili, 2015). This shows the politicians and journalists are clustered together, as are many of the social media websites. Scientists, academic, and professional medical sites are clustered together. The medical professions (doctors and nurses) comprise a cluster. The correlations between the means for the standardized values for these clusters and the Trust in Science estimated latent variable from above are printed in this Figure. The correlation between Trust in Science and the trust in the science cluster is, as expected, high. There is little association between Trust in Science and how much people trust politicians, journalists, or social media.

An important assumption of hierarchical cluster analysis is that each branch is meaningful. This means that “Support Groups” can be seen as more similar to “Blogs” than to “Whats App,” even though these are clustered together to conduct the correlations with Trust in Science.

For non-hierarchical cluster analysis the analyst determines how many clusters to place the variables into. This could be based on some theory about how people are likely to respond or by examining how well cluster analysis works with different numbers of clusters. One approach is to try several numbers of clusters and measure how much variation there is within the clusters. If the variables in the same cluster are all really similar, the variation should be small, but if they are different there should be much variation. One approach uses the *gap statistic* (Tibshirani, Walther, & Hastie, 2001). Let the within cluster sum of squares variation be denoted W_k for the method with k th clusters. The gap statistic is $Gap_k = E(\ln(W_k)) - \ln(W_k)$, where $E(\ln(W_k))$ is the expected values for random data. This is much like the parallel analysis used with the scree plot described earlier. As with that, the expected value for random data can be calculated several times. This is shown by

Figure 5

Dendrogram of the trust in authorities and sources items.



the error bars in Figure 6. As with the scree, choosing the appropriate number of clusters is part an art but several methods have been devised to assist the analyst in making the choice. The `clusGap` function from the R package **cluster** (Maechler, Rousseeuw, Struyf, Hubert, & Hornik, 2022) calculates the gap statistic and its standard error for different numbers of clusters. The `maxSE` function uses this and finds the “best” of these using six different methods, but as with choosing dimensions from a scree plot the analyst should carefully examine the gap plot and the resulting clusters of variables to see what makes sense. The results from THE six methods were: one suggested a single cluster, one suggested nine clusters, and the remaining four suggested four clusters. I will use four clusters. These are:

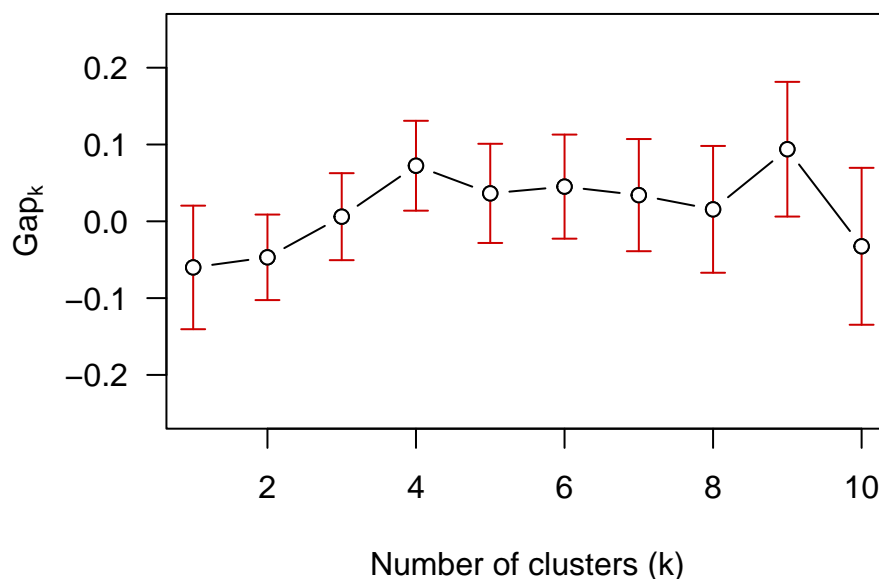
1. National politicians, Local politicians, Journalists
2. Blogs, Support groups, Social networks
3. WhatsApp
4. Doctors, Nurses, University academics, Scientists, Profession health sites, NHS/CDC website

This is slightly different from the four clusters identified in Figure 5. Trust in “WhatsApp” is separated into its own cluster and all the medical and science related sources are placed together. It is important to recognise that the different statistical procedures will often give different answers. There is seldom a single “correct” method for analysing data. The choices the analysts make may affect the answers.

Continous constructs. I assumed that the Identity Resilience items, the Social Support items, the Trust in Science items, and the COVID Prevention items, can be thought of as measuring multiple dimensions.

Figure 6

Gap plot for the trust in authorities and sources. The choice for the number of clusters is to have a few as possible while still having the gap statistic being fairly large. Most of the gap statistic measures suggest four clusters.



Dimensionality. I assumed four factors underlie the Identity Resilience items and a bi-factor model with four additional components fit the data well. The Social Support, Trust in Science, and COVID Prevention items all were consistent with single constructs, as predicted from the papers introducing these scaled.

Trust in Authorities/Sources. Single item construct models did not fit well for the Trust in Authorities and the Trust in Sources of Information items. Cluster analysis was used for these and discussed.

Comparing Scales by Countries

Table 3 shows the means, their 95% confidence intervals, a t test comparing these, and a common effect size for this comparison called Cohen's d . This is the difference in means divided by the standard deviation. The p -values for the individual tests are printed as well as those after adjusting using Holm's method. This is used because having eight t -tests means the probability of getting a significant result (i.e., $p < .05$) on at least one of these, even if there are no differences in the population, is much higher than 5% (Bretz, Hothorn, & Westfall, 2010). An alternative way to consider these differences is looking at the effect sizes. J. Cohen (1992) describes $d = .20$ as a small effect ($d = .50$ as medium,

and $d = .80$ as large). From this the effects for Trust in Science, COVID prevention, and Distinctiveness are all, using his terminology, “small” effects, and the others do not reach this level. Respondents in the US trust science more, have higher scores on COVID prevention, and view themselves as more distinctive. It is important to stress that the labels Cohen uses are arbitrary, which he emphasizes, but also argues for their applicability in social science research:

Although arbitrary, the proposed conventions will be found to be reasonable by reasonable people. . . . Many effects sought in personality, social, and clinical-psychological research are likely to be small effects as here defined, both because of the attenuation in validity of the measures employed and the subtlety of the issues frequently involved.

J. Cohen (1988, p. 13)

There are several assumptions of these t -tests including that the within-group population distributions are normally distributed and with equal variances. These assumptions will never be correct: researchers should “move from [the idea that] all assumptions are right towards all assumptions are wrong” (Tukey, 1986, p. 72). This does not mean that they should be ignored, but that even with relatively small deviations the results will likely still enable wise decision making. Another assumption is that the group variable is measured without error. This should not be an issue for these comparisons as people should accurately know which country they are in, but we bring it up here as it is an issue with some later statistical procedures.

Table 3

Comparing the means for the UK and US. Student t -tests with their associated p -values (without and with Holm’s adjustment for the number of tests) are shown with Cohen’s d . 95% Confidence intervals (CI) are shown below the means and d .

	\bar{x}_{UK}	\bar{x}_{US}	t	df	p	p_{adj}	Cohen’s d
Trust in Science	-0.085	0.124	4.305	1,878	< .001	< .001	0.202
CI	(-0.145, -0.025)	(0.049, 0.199)					(0.110, 0.295)
Social Support	0.036	-0.053	1.811	1,878	.070	.255	-0.085
CI	(-0.022, 0.094)	(-0.132, 0.026)					(-0.177, 0.007)
COVID Prevention	-0.142	0.208	6.954	1,878	< .001	< .001	0.327
CI	(-0.204, -0.080)	(0.130, 0.285)					(0.234, 0.419)
Identity Resilience	0.034	-0.050	2.012	1,878	.044	.222	-0.095
CI	(-0.015, 0.083)	(-0.119, 0.019)					(-0.187, -0.002)
Self esteem	0.018	-0.027	1.424	1,878	.155	.309	-0.067
CI	(-0.020, 0.056)	(-0.077, 0.023)					(-0.159, 0.025)
Efficacy	0.018	-0.027	1.216	1,878	.224	.309	-0.057
CI	(-0.025, 0.062)	(-0.087, 0.034)					(-0.149, 0.035)
Distinctiveness	-0.070	0.103	4.041	1,878	< .001	< .001	0.190
CI	(-0.122, -0.018)	(0.036, 0.170)					(0.098, 0.282)
Continuity	-0.030	0.044	1.856	1,878	.064	.255	0.087
CI	(-0.079, 0.020)	(-0.016, 0.103)					(-0.005, 0.179)

Table 4 shows the correlations among all the constructs. Cohen’s (1992) labels for correlations are: $r = .10$ as small, $r = .30$ as medium, and $r = .50$ as large. Using

this terminology the association between Social Support and Identity Resilience is large, between Trust in Science and COVID Prevention is medium, and some of the components of Identity Resilience have small associations with the other constructs. It is worth noting that a scatterplot matrix of these should also be examined to search for outliers and non-linearity. It is not printed here as no issues were found. When discussing *t*-tests I mentioned the assumption that the grouping variable is assumed to be measured without error. Here, the correlation is interpreted as how closely associated the two variables created using factor analysis are. But, these variables are just estimates of the constructs and have error associated with them. The correlations between the true constructs are likely to be higher than those between their estimates. There are ways to estimate the correlations between the actual constructs and not just their estimate. Structural equation modelling (SEM) is one way and this approach is used in the next section.

Table 4

Pearson correlations between the constructs, includes the four components of the Identity Resilience Index.

	TSci	SucSup	COVID	IdRel	SEst	Effic	Dist	Cont
Trust in Science								
Social Support	.138							
COVID Prevention	.320	.039						
Identity Resilience	.082	.554	.026					
Self esteem	.020	.273	-.010	.371				
Efficacy	.050	.077	-.003	.225	-.347			
Distinctiveness	.081	.039	.058	.082	-.144	-.063		
Continuity	-.042	.037	.016	.131	-.211	-.012	-.065	

The relationships among these constructs

A popular procedure in the social sciences is putting forward a causal model for the relationships among variables, seeing how well the data fit the model, and then focusing on the relationships between the pairs of constructs. This is called *path analysis*. A good introduction is Loehlin and Beaujean (2017). One popular approach is called structural equation modelling. This involves simultaneously fitting a model that incorporates both the measurement of the latent constructs and the relations among them. As our interest here is comparing the UK and US, we include effects for country influencing both the individual constructs and proposed relationships constructs, resulting in a complex model. Alternatively, we could run separate models for the UK and US. This would mean how the constructs are created would differ for the countries and therefore the constructs would not be the same. A third approach, sometimes called the two-step approach is (e.g., Anderson & Gerbing, 1988) where the measurement of the constructs occurs in the first step and the relationships among the constructs, or the structural part of the model, in the second. This is also called the *structural after measurement* (SAM) approach. This means that constructs are made in the same way for the two countries. There are other approaches

that can be used, but this approach also conceptually matches onto how I have described the problem and that makes it easier to convey to readers. It is important to note that it is not always possible to conceptually separate the measurement and structural parts of the model.

One approach to this two-stage approach would be to estimate the constructs as was done above and use these in a set of regression equations. As discussed with respect to t -tests, these would assume that the predictor variables are measured without error. Rosseel and Loh (2023) describes several problems with this approach but note that it is still popular. Like with the correlations it tends to underestimate the associations among the variables. The alternative is to include the uncertainty in these estimates in the model. Rosseel and Loh (2023) show the equations for doing this and provide a function `sam` in newly version of the package `lavaan` (Rosseel, 2012) to do this. This is the approach used here.

The model is simpler than we examine with these data in Jaspal et al. (in press) when comparing the UK and Portugal. This is largely because of wanting to concentrate on the methodological issues of the approach. The Identity Resilience is treated as a single latent variable from the 16 IR measures. As discussed in “JBA1” (2023), this is assumed to be related to other constructs. Trust in Science should influence COVID-19 Prevention, and given the messaging in the two countries differed we expect that COVID-19 Prevention ratings will still vary by country even beyond Trust in Science.

SEM, and more generally using sets of regression models to create path diagrams, is sometimes referred to as causal modeling as if causation can suddenly be determined from the correlations by using fancy statistics. Morgan and Winship (2015) summed up the consequences of believing the statistical procedures are magic.

Naive usage of regression modeling was blamed for nearly all the ills of sociology, everything from stripping temporality and context from the mainstream, ... the suppression of attention to explanatory mechanisms, ... the denial of causal complexity, ... and the destruction of mathematical sociology.

Morgan and Winship (2015, p. 13)

Cartwright (2014) describes the situation succinctly as “no causes in, no causes out.” I assume Identity Resilience, Social Support, and Trust in Science are related, and that Trust in Science influences COVID preventative behaviours, but our statistical procedures cannot show if the direction of causation is accurate. This is a framework in which to test our hypotheses about country differences. Country is treated as exogeneous, none of these other variables will influence it appreciably (not completely because, for example, someone who really trusts science might feel compelled to move to the UK, but we are willing to assume this possibility represents a small enough movement to ignore). Interest is whether Country influences Trust in Science and whether after accounting for the influence of Trust in Science on COVID behaviours, whether Country further influences COVID behaviours. It does. This can be shown by comparing the model in Figure 7 without the dashed line with the model with the dashed line. The difference in fit is: $\chi^2(1) = 39.72, p < .001$. This suggests this effect (the dashed arrow of Figure 7) should be included in the model.

Like the procedure itself, the numeric results from the SAM model are separated into the measurement and structural parts. The measurement part estimates the reliability

Table 5

The path coefficients and related statistics for the model shown in Figure 7.

Path	coef	se	z	p
Identity Resilience → Social Support	-0.551	0.026	-21.546	< .001
Identity Resilience → Trust in Science	-0.084	0.023	-3.726	< .001
Country → Trust in Science	-0.154	0.035	-4.437	< .001
Trust in Science → COVID Prevention	-0.484	0.038	-12.772	< .001
Country → COVID Prevention	0.275	0.048	5.704	< .001

for each construct. These are (for the model including the dashed line in Figure 7):

Identity Resilience	.901
Trust in Science	.927
Social Support	.915
COVID Prevention	.853

The α values listed in Table 1 are often described as lower bounds for the reliability, so it is worth noting where that these values are all higher.

Statistics related to the structural aspects of the model are shown in Table 5. The largest effects are for the Identity Resilience to Social Support path and the Trust in Science to COVID Prevention path. Identifying these is of applied importance. In particular, the path between Trust in Science and preventative behaviours suggests increasing trust in science may improve behaviours that in turn may lessen the impact of pandemics.

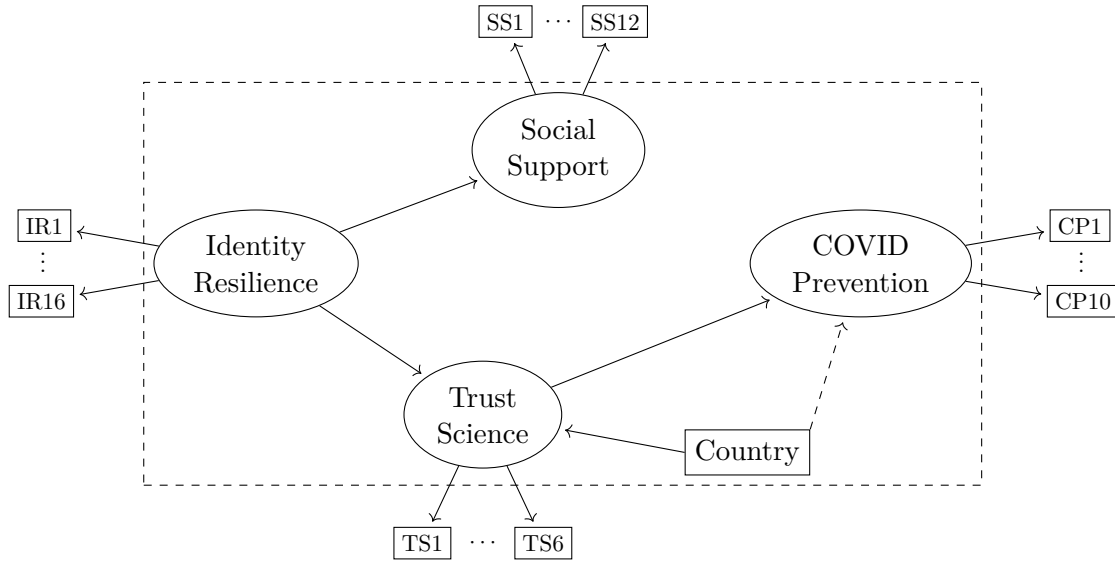
Conclusion/Summary

Considerations and assumptions are part of all research, even when not evident in the descriptions of the research in scientific journals. The aim of this paper, like Toto, is to pull back the curtain that obscures the truth, but unlike the Wizard of Oz (in the film) I want you to “pay attention to those considerations and assumptions behind the curtain [of scientific protocols].” Word limits, masked review, and years of training have led gathering of social scientists to report the main findings from research in a succinct and formal manner that hides the decisions made when conducting the research. The format of journals and the 15-minute conference slots make this inevitable and we do not believe the approach used here is appropriate for all dissemination. For this special issue, focused on several of our research ideas and with a broad audience, it is worth describing our considerations and assumptions in more detail. Further, reflecting on these helps to focus on these decisions and forced me to be explicit why certain choices were made. Doing this had therapeutic value for me.

We created a research team and developed research questions in response to a call from the British Academy. Our aim was to explore different aspects of identity process theory (IPT) in the UK and US with focus on different ethnicities within the context of COVID-19. We focus on IPT in the first paper in this issue and ethnic differences in the third. Our main method, for much of our research, is to present a set of scales to people and make conclusions about how humans think and behave based on the associations among all

Figure 7

A structure after measurement (SAM) model (Rosseel & Loh, 2023) of the relationships among COVID-19 related constructs. The observed variables that make up the constructs and their error terms are not shown. The dashed rectangle encloses the structural part of the model. The measurement is shown by the paths between the constructs inside the rectangle and the items outside it. Each of the observed items, other than Country, also has an implied error term associated with it, but these are not shown as the figure already is fairly complex. The coefficients for this model are shown in Table 5



of their responses. This is a tall order and requires both assumptions and some empirical checks of *some* of these assumptions. We assume that the sample achieved online through Prolific will be similar enough to others to provide useful and meaningful results. We assume that their responses inform us about their beliefs and behaviours consistent with our intent. We assume these can be aggregated and represent the intended psychological constructs. The choice of statistical methods for this aggregation and for looking at the associations among the constructs also require decisions. In the typical article, the authors focus more on *what* they did rather than *why* they did what they did, and *why* they didn't do the alternatives.

What did we find?

- The items for Trust in Science, Social Support, and COVID Prevention could all be represented well with single constructs.
- The Identity Resilience Index was best fit by four components and an overall Identity Resilience construct.
- The Trust in different Authorities and Sources of Information scales could not be summarized well by single measures. The different authorities and sources of information could be categorized depending on if they were from politicians, social media,

scientists, or the medical profession.

- The “structure after measurement” model shows Identity Resilience influencing Social Support and Trust in Science, and Trust in Science influencing COVID Prevention behaviours. Country was related to Trust in Science (US trusting more) and COVID Prevention (US reporting more of these behaviours).

We believe that social science theory and methods can help inform policy and other applications related to societally important issues. The COVID-19 pandemic is an example. While medical and economic research are vital for pandemics, so is understanding how people will psychologically react to health guidance and restrictions. Social scientists have many tools at their disposal. When faced with global crisis the research tools from many disciplines can be useful. Each discipline has tacit protocols. Being explicit about the protocols helps readers to better understand the approaches and the research implications.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational & psychological testing (2014 edition)*. Washington, DC: American Educational Research Association.
- Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, 103, 411-423. doi: 10.1037/0033-2909.103.3.411
- Auerswald, M., & Moshagen, M. (2019). How to determine the number of factors to retain in exploratory factor analysis: A comparison of extraction methods under realistic conditions. *Psychological Methods*, 24, 468-491. doi: 10.1037/met0000200
- Bacon, F. (2019 [1620]). *Novum organum*. Anados Books.
- Bacon, F. (2020 [1626]). *New Atlantis*. Dallas, TX: CTMS Publishers, University of Dallas.
- Bartholomew, D. J. (1993). Estimating relationships between latent variables. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, 55, 409-419. doi: 10.2307/25050951
- Bartholomew, D. J., Knott, M., & Moustaki, I. (2011). *Latent variable models and factor analysis: A unified approach* (3rd ed.). Chichester, UK: Wiley.
- Bartholomew, D. J., Steele, F., Moustaki, I., & Galbraith, J. (2002). *Analysis of multivariate social science data*. Chapman & Hall/CRC.
- Belli, R. F., Conrad, F. G., & Wright, D. B. (2007). Cognitive psychology and survey methodology: nurturing the continuing dialogue between disciplines. *Applied Cognitive Psychology*, 21, 141-144. doi: 10.1002/acp.1333
- Biffignandi, S., & Bethlehem, J. (2021). *Handbook of web surveys* (2nd ed.). Hoboken, NJ: Wiley.
- Breakwell, G. M. (2014). *The psychology of risk* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Breakwell, G. M. (2015). Identity process theory. In G. E. Sammut, E. E. Andreouli, G. D. Gaskell, & J. E. Valsiner (Eds.), *The Cambridge handbook of social representations* (p. 250-266). Cambridge, UK: Cambridge University Press.
- Breakwell, G. M. (2023a). Choosing a research design. In ** (Ed.), (p. **). **: **.

- Breakwell, G. M. (2023b). *Identity, unique and shared*. London, UK: Sage.
- Breakwell, G. M., Fino, E., & Jaspal, R. (2021). The COVID-19 preventive behaviors index: Development and validation in two samples from the United Kingdom. *Evaluation and the Health Professions*, 44, 77-86. doi: 10.1177/0163278720983416
- Breakwell, G. M., Fino, E., & Jaspal, R. (2022). The Identity Resilience Index: Development and validation in two UK samples. *Identity*, 22, 166-182. doi: 10.1080/15283488.2021.1957895
- Breakwell, G. M., Wright, D. B., & Barnett, J. (Eds.). (2020). *Research methods in psychology* (5th ed.). London, UK: Sage.
- Bretz, F., Hothorn, T., & Westfall, P. (2010). *Multiple comparisons using R*. Boca Raton, FL: CRC Press.
- Campbell, N. R. (1928). Personal and impersonal styles in scientific communication. *Nature*, 121, 1021. doi: 10.1038/1211021a0
- Cartwright, N. (2014). Causal inference. In N. Cartwright & E. Montuschi (Eds.), *Philosophy of social science* (p. 308-326). Oxford, UK: Oxford University Press.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 245-276. doi: 10.1207/s15327906mbr0102_10
- Chambers, J. M. (2008). *Software for data analysis: Analysis with R*. New York, NY: Springer.
- Chambers, J. M. (2009). Facets of R. *R Journal*, 1(1), 5-8. Retrieved from https://journal.r-project.org/archive/2009-1/RJournal_2009-1_Chambers.pdf
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159. doi: 10.1037/0033-2909.112.1.155
- Cohen, S., Mermelstein, R., Kamarck, T., & Hoberman, H. M. (1985). Measuring the functional components of social support. In I. G. Sarason & B. R. Sarason (Eds.), *Social support: Theory, research and applications* (Vol. 24, p. 73-94). NATO ASI Series. doi: 10.1007/978-94-009-5115-0
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334. doi: 10.1007/BF02310555
- Dyson, F. W., Eddington, A. S., & Davidson, C. R. (1920). A determination of the deflection of light by the sun's gravitational field, from observations made at the total eclipse of May 29, 1919. *Philosophical Transactions of the Royal Society of London. Series A*, 220, 291-333. doi: 10.1098/rsta.1920.0009
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Feynman, R. P. (1974). Cargo cult science: Some remarks on science, pseudoscience, and learning how to not fool yourself. Caltech's 1974 commencement address. *Engineering and Science*, 37(7), 10-13.
- Fienberg, S. E. (1971). Randomization and social affairs: The 1970 draft lottery. *Science*, 171, 255-261. doi: 10.1126/science.171.3968.255
- Galili, T. (2015). **dendextend**: an R package for visualizing, adjusting, and comparing trees of hierarchical clustering. *Bioinformatics*, 31, 3718-3720. doi: 10.1093/bioinformatics/btv428

- Gigerenzer, G. (1991). From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review*, 98, 254-267. doi: 10.1037/0033-295X.98.2.254
- Goodman, L. A. (1978). *Analyzing qualitative/categorical data: Log-linear models and latent-structure analysis*. London: Addison-Wesley Publishing Company.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74, 1464-1480. doi: 10.1037/0022-3514.74.6.1464
- Gummer, T., Roßmann, J., & Silber, H. (2021). Using instructed response items as attention checks in web surveys: Properties and implementation. *Sociological Methods & Research*, 50, 238-264. doi: 10.1177/0049124118769083
- Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, 10, 255-282. doi: 10.1007/BF02288892
- Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. but *Communication Methods and Measures*, 14, 1-24. doi: 10.1080/19312458.2020.1718629
- Jaspal, R., & Bayley, J. (2020). *HIV and gay men: Clinical, social and psychological aspects*. London, UK: Palgrave.
- Jaspal, R., & Breakwell, G. M. (Eds.). (2014). *Identity process theory: Identity, social action and social change*. Cambridge, UK: Cambridge University Press.
- Jaspal, R., & Breakwell, G. M. (2023). Race/ethnicity and COVID-19. In *The Wiley-Blackwell encyclopedia of health, illness, behavior, and society*. Wiley-Blackwell.
- Jaspal, R., Breakwell, G. M., & Wright, D. B. (in press). Identity resilience, science mistrust, COVID-19 risk and fear predictors of vaccine positivity and vaccination likelihood: A survey of UK and Portuguese samples. *Journal of Health Psychology*.
- Jaspal, R., Fino, E., & Breakwell, G. M. (2022). The COVID-19 Own Risk Appraisal Scale (CORAS): Development and validation in two samples from the United Kingdom. *Journal of Health Psychology*, 27, 790-804. doi: 10.1177/1359105320967429
- Jba1. (2023). *Journal of the British Academy*.
- Jba3. (2023). *Journal of the British Academy*.
- Kalton, G. (2021). *Introduction to survey sampling* (2nd ed.). Thousand Oaks, CA: Sage.
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50, 1-73.
- Loehlin, J. C., & Beaujean, A. A. (2017). *Latent variable models: An introduction to factor, path, and structural analysis*. (5th ed.). New York, NY: Routledge.
- Loftus, E. F., Fienberg, S. E., & Tanur, J. M. (1985). Cognitive psychology meets the national survey. *American Psychologist*, 40, 175-180. doi: 10.1037/0003-066X.40.2.175
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2022). **cluster**: Cluster analysis basics and extensions [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=cluster>
- McDonald, R. P. (1978). Generalizability in factorable domains: Domain validity and generalizability. *Educational and Psychological Measurement*, 38, 75-79.
- McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, 23, 412-433.
- McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research*

- Methods*, 52, 2287-2305. doi: 10.3758/s13428-020-01398-0
- Meehl, P. E. (1967). Theory-testing in psychology and physics: a methodological paradox. *Philosophy of Science*, 34, 103-115.
- Mizumoto, A., & Plonsky, L. (2016). R as a *Lingua Franca*: Advantages of using R for quantitative research in applied linguistics. *Applied Linguistics*, 37, 284-291.
- Morgan, S. L., & Winship, C. (2015). *Counterfactuals and causal inference: Methods and principles for social research* (2nd ed.). Cambridge, MA: Cambridge University Press.
- Nadelson, L., Jorcyk, C., Yang, D., Jarratt Smith, M., Matson, S., Cornell, K., & Husting, V. (2014). I just don't trust them: The development and validation of an assessment instrument to measure trust in science and scientists. *School Science and Mathematics*, 114, 76-86. doi: 10.1111/ssm.12051
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231-259. doi: 10.1037/0033-295X.84.3.231
- O'Connor, B. P. (2021). **EFA.dimensions**: Exploratory factor analysis functions for assessing dimensionality [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=EFA.dimensions> (R package version 0.1.7.2)
- Popper, K. (1994). *The myth of the framework: In defense of science and rationality*. Oxford, UK: Routledge.
- R Core Team. (2022). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Revelle, W. (2018). **psych**: Procedures for psychological, psychometric, and personality research [Computer software manual]. Evanston, Illinois. Retrieved from <https://CRAN.R-project.org/package=psych> (R package version 1.8.10)
- Rosseel, Y. (2012). **lavaan**: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1-36. Retrieved from <http://www.jstatsoft.org/v48/i02/>
- Rosseel, Y., & Loh, W. W. (2023). A structural after measurement approach to structural equation modeling. *Psychological Methods*. doi: 10.1037/met0000503
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: John Wiley & Sons.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6, 461-464. doi: 10.1214/aos/1176344136
- Sheehan, K. B., & Pittman, M. (2016). *Amazon's Mechanical Turk for academics: The HIT handbook for social science research*. Irvine, CA: Melvin & Leigh.
- Spearman, C. (1904). 'General intelligence,' objectively determined and measured. *The American Journal of Psychology*, 15, 201-293. doi: 10.2307/1412107
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103, 677-680. doi: 10.1126/science.103.2684.677
- Stigler, S. M. (2016). *The seven pillars of statistical wisdom*. Cambridge, MA: Harvard University Press.
- Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63, 411-423. doi: 10.1111/1467-9868.00293
- Tukey, J. W. (1986). Sunset salvo. *The American Statistician*, 40, 72-76. doi: 10.1080/

00031305.1986.10475361

- van Buuren, S. (2018). *Flexible imputation of missing data* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC Press.
- Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: a review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment* (p. 41-71). Boston, MA: Kluwer.
- Waller, N. G., & Meehl, P. E. (1998). *Multivariate taxometric procedures: Distinguishing types from continua*. Newbury Park, CA: Sage.
- Widaman, K. F., & Revelle, W. (2022). Thinking thrice about sum scores, and then some more about measurement and analysis. *Behavior Research Methods*, *. doi: 10.3758/s13428-022-01849-w
- Wright, D. B., Wolff, S. M., Jaspal, R., Barnett, J., & Breakwell, G. M. (2022). The choice of response alternatives in COVID-19 social science surveys. *PLOS One*, 17(11), e0263552. doi: 10.1371/journal.pone.0263552
- Xie, Y. (2015). *Dynamic documents with R and **knitr*** (2nd ed.). Boca Raton, FL: Chapman and Hall/CRC.