
USING PEDAGOGICAL AND PSYCHOLOGICAL INSIGHTS TO TRAIN ANALYSTS USING CONFIDENTIAL DATA

ELIZABETH GREEN AND FELIX RITCHIE

University of the West of England, Bristol
e-mail address: elizabeth7.green@uwe.ac.uk

University of the West of England, Bristol
e-mail address: felix.ritchie@virgin.net

ABSTRACT. With researchers increasingly gaining access to confidentiality data through restricted environments, interest has grown in the training of those researchers to protect confidentiality and to use the secure facility effectively.

Researcher training, where it exists, often tends to focus on the ‘chalk-and-talk’ approach or its digital equivalent, the aim is to ensure that the researchers are informed of their legal obligations and so take responsibility for their actions. Although popular, there are multiple problems with this approach. First, it is of limited pedagogical effectiveness. Second, it assumes that information delivery is the purpose of the training. Third, it does not take account of attendees’ attitudes when attending the course. Fourth, it creates an ‘us and them’ barrier between trainers and trainees.

An alternative approach to training researchers has been in place in the UK since 2017. It uses good pedagogical practice to increase the effectiveness of training. It uses psychological models of behaviour and attitudes to engage attendees and shape future behaviours. The aim of the course is to build a shared sense of community and trust, rather than information delivery, in line with good data governance practice.

This paper describes the experience of designing and running the course. Multiple organisations and trainers were involved in design and delivery, improving feedback but creating its own problems in terms of trainers’ different preferences. Overall, the approach has been highly successful, and has become the model for other organisations. However, the model does place higher demands on the trainer than the traditional model.

We also briefly touch on how the move to online teaching in the pandemic has learned from the face-to-face experience.

1. INTRODUCTION

Much of the analytical value from the great advances in social science this century has come from the safe use of identifiable confidential or sensitive government and survey data (Ritchie, 2021). Using identifiable confidential data without direct informed consent comes with risks: ethical, legal and societal. Anonymisation is often seen as one solution in the spectrum

Key words and phrases: training, confidentiality, data governance.
Corresponding author: E. Green.

of access possibilities, but the detail required by anonymisation reduces the value of the data; moreover, vulnerability to technological developments (Ritchie and Smith, 2019) raises concerns over the long-term viability of anonymisation techniques. Hence, data holders are increasingly looking to portfolio governance models such as the ‘five safes’ (Ritchie, 2017). The portfolio model is also reflected in legislation, such as the European General Data Protection Regulation, the UK Digital Economy Act, and the Australian Data Availability and Transparency Act.

In most portfolio approaches, the conduct of the researcher is central. This is particularly true for trusted research environments (TREs), where researchers have great freedom to study and manipulate the data in highly secure virtual environments.

Although the risk of malicious misuse of data by genuine researchers is vanishingly small, the risk of accidental or deliberate misuse (for example, to avoid onerous restrictions) is non-negligible, and needs to be considered in any system design. Moreover, if researchers are trusted then systems can be designed around that trust, allowing for both more efficient controls and greater security (Desai and Ritchie, 2009). That trust needs to be based on evidence, such as qualifications, contractual agreements, or – increasingly – researcher training.

There is a wealth of literature surrounding teaching and learning; and there are guidelines such as Corti et al (2019) on how to work with confidential data management. However, with the exception of Green et al (2017), there appear to be no papers on effectively teaching analysts about confidential data management or statistical disclosure control (SDC). There is a small literature around the need to instil ethical awareness and good confidentiality practices, but none of this covers the way that information is passed on; the assumption is that providing the information to researchers (and in some cases checking that users have read it) is sufficient.

In the past, training has usually been based on the ‘policing’ or ‘rational agent’ model (Green et al, 2017). This treats researchers as potential malicious ‘intruders’; the primary purpose of training is to enforce good behaviour by making researchers aware of the consequences of misbehaviour, such as prison, fines, or loss of access. The fondness for this approach arises from (a) defensive attitudes to data protection (Hafner et al, 2017) (b) limited understanding of the psychology of attendees, and (c) lack of pedagogical experience in trainers.

In 2017 the authors were commissioned by the UK Office for National Statistics (ONS) to devise a new researcher training programme. The requirement was to redesign researcher training from scratch following the EDRU approach (evidence-based, default-open, risk-managed, user-centred; ADSS, 2016). The course was grounded in psychological and educational theories around how to teach professionals, how to motivate researchers, and how to foster a community; the primary goal was not information delivery, but behaviour change. The resulting Safe Researcher Training (SRT) course is now central to UK data governance, and significant elements, particularly the active teaching style, have been adopted in other countries.

This paper considers the lessons learned over the last five years, and compares outcomes with the ambition outlined in Green et al (2017). We show how different psychological and pedagogical concepts were operationalised, and how the course matured. We consider where ambitions were exceeded or missed, and the reasons behind this. Finally, we consider how transferable these lessons are; for example, training support staff, or delivering training online.

While there is little literature specifically on training researchers to use confidential data, there is much on relevant psychological and pedagogical models, and section 2 summarises key findings. Section 3 introduces the design principles, and uses specific exercises to show how these principles were operationalised. Section 4 reflects on the lessons learned and how the course and teaching evolved; it also provides some preliminary perspectives on the move to online learning in the Covid-19 pandemic. Section 6 concludes.

SRT was designed to cover use of confidential data by researchers in all situations. However, it was commissioned to be an access requirement for users of the ONS TRE and similar secure facilities. Hence, we focus on the TRE users.

This paper is concerned with training of authorised users. We do not consider attempts to hack systems or reverse-engineer statistics by unauthorised users.

2. LITERATURE REVIEW

There is almost no discussion of how researchers can be trained most effectively to work securely and efficiently on confidential data. However, there are well-understood models of researcher behaviour and of effective pedagogical practice. In addition, we need to place such professional development courses in in the appropriate institutional context.

2.1. Models of researcher training. Until fairly recently (Ritchie, 2021), models for data access were built on the threat of intruders: deliberate attempts to breach confidentiality. This led to the ‘policing’ model in training users for data access: a user is seen as an untrustworthy self-serving agent (Ritchie and Welpton, 2014). Whilst the policing approach appears superficially logical, Hafner et al (2015) argue that this lacks empirical support or theoretical coherence. Kim and Kim (2015) demonstrate that academic motivation is fundamentally intrinsic (ie driven by personal interest and perceptions of the value of the work), suggesting that extrinsic factors (fines, prison) are effective forms of behaviour control for researchers.

Desai and Ritchie (2009) argue that organisations can use this intrinsic motivation so that researchers become self-governing, improving both data security and efficiency; researcher training is a part of this, although they do not discuss the details of training. Brandt et al (2010) do discuss content, but not the pedagogical model.

ADSS (2016) reviewed training policies in National Statistical Institutes (NSIs) for users requiring access to confidential data in their secure facilities. Most had passive guidelines for researchers; compulsory face-to-face training was not common at that time. Face-to-face training was more focused on the technical aspect of using systems, rather than wider discussion about confidentiality; information delivery was generally the driver.

Moving on five years, Green et al (2021) summarise an international workshop convened to identify common and good practices in confidential data management. This came out strongly (although not unanimously) in favour of training targeting intrinsic motivation, but this was largely driven by the experience of participants rather than any specific conceptual framework.

In summary, there appears to be no literature considering the best way to train researchers in confidential data management.

2.2. Psychological models of researcher behaviour and learning. The broad elements of human psychology, and the implications for pedagogy, are well-understood, and so we provide only a brief summary of key relevant theories here. For a more detailed review see Green and Ritchie (2022).

2.2.1. Behaviourism, conditioning and learning. With its origins in animal behavioural research, the theory of connectionism states that, for learning to happen,

- the learner must be physiologically and mentally ready to learn (law of readiness)
- the more repeated interaction between the stimulus and response (student and exercise), the more ingrained the information becomes (law of exercise)
- learning response are strengthened when positively rewarded but weakened when punishment is used (law of effect)

This model forms the basis of many traditional teaching routes, The role of the teacher is to ensure that their students are prepared and motivated to learn ('readiness'). The teacher sets the students repetitive exercises ('exercise'), and rewards correct behaviour and punishes bad behaviour ('effect'). The outcome is that the student are *classically conditioned* to automatically display correct behaviours (Thorndlike, 1932; Islam, 2015).

The "law of exercise" is focused on the repetition of a learned rule which does not necessarily equate to an understanding of the task and its theoretical underpinnings (Shuell, 1986). A further criticism of this model revolves around the use of punishment as a means to shape behaviour. It is important for the researchers to not see the teachers as police as this prevents relationship development between the teacher (controller) and the student (the researcher).

Operant conditioning focuses on the rewarding positive behaviour (positive reinforcement) and ignoring negative behaviour (negative reinforcement). This model is used to help shape behaviour in areas where negative behaviour does not have an overwhelming detrimental impact or where the negative outcome can be efficiently managed. However, this practice, despite being relatively common in the form of most workplace/ customer service dynamics (Strickler, 2006), fails to explicitly raise the researcher's failings with them; therefore, unless the researcher is made aware of their mistakes the behaviour may continue. Research has demonstrated that integrating operant conditioning is more effective at long-term shaping of behaviour than classic conditioning (Wolfgang, 2001; Skinner 1953).

2.2.2. Cognitive models of learning. The shift away from behaviourist models happened in the early 20th century. Whilst behaviourists focused on the environmental stimuli and conduction for learning, cognitivism was interested in the internal mechanisms of how knowledge is; acquired, processed, stored, retrieved, and activated by the individual (Anderson et al 1997; Greeno et al, 1996).

Cognitive learning models suggest that learning happens best under conditions that are aligned with human cognitive architecture (Thagard, 1996). Learning materials should be made meaningful to the learners (Galavotti, 2019), and the acquisition of new skills is influenced by learners' three internal states: (a) students' knowledge; (b) students' desire to complete the course; (c) learners' self-efficacy (Zimmerman, 1995). Levels of self-efficacy may fluctuate across the course as the student recalibrates their knowledge and aptitude to each new task; therefore the teacher needs to gauge efficacy cues at the end of each task (Schunk, 1989).

Schunk (1987) formulated a self-efficacy model to demonstrate the relationship between internal factors and achievement (Fig. 1).

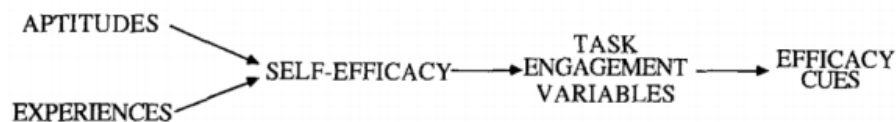


Figure 1: Self-efficacy model of achievement behaviour (Schunk, 1987)

The model can help explain individual differences observed across students; for example, there may be different levels of experience and so forth. This model allows teachers to formulate how students may engage with a task-based on prior knowledge/ experience and their aptitude (capacity to learn). In the context of researcher training, many of the students are established post-graduate researchers or academic staff; therefore we can assume that students at the start of the educational activity will hold an initial high level of self-efficacy and theory engaged within the content. The self-efficacy model implies continuous engagement and motivation, placing the onus on the teacher to ensure sufficient cognitive engagement with the course materials.

When individuals 'fail to learn', one contributory factor is whether the individual at the time of processing deems the information as being relevant (Broadbent, 2013). Processing information into long-term-memory storage depends on whether it has a special and meaningful relationship for the individual; therefore the individual must perceive the knowledge to be both personal and useful to them (Eysenck and Eysenck, 1979).

Cognitive models emphasise the need for the teacher to form individual level bonds with the students alongside a group bond, and continually monitor efficacy cues (Yilmaz, 2011). Alverman et al (1985) found students selectively assimilating parts of information, and distorting it to fit their already existing preconceptions and processes. This means that the teacher needs to continually check-in on students learning. The active use of feedback from social interaction can promote self-regulation and further the group bond and identity (Sadler 2010).

2.2.3. Social identity theory. Social Identity Theory (SIT) combines cognitive and social processes and provides a structure for the analysis of intergroup behaviour whilst also considering individual cognitive processes (Tajfel et al. 1971). Individuals share common experiences, beliefs, and values, and these individuals are clustered with a shared identity group (Turner et al. 1987). SIT within education provides insights into theories of group dynamics particularly within collective action and behaviour.

SIT states if the group feels they have low status with no power or control (in contrast to the other group), this will ultimately create a dynamic situation where group members will actively seek to change and empower their group status (Tajfel and Turner, 1979). If the group does not feel they can mobilise their group status, the group will negatively compete against other groups (Turner and Reynolds, 2010). In an educational context SIT can provide insights into poor course engagement and poor student communication, as the actions of the teacher quickly create 'in' and 'out' groups as well as dissociating the teacher from the students.

When training users in confidentiality, there are distinct groups and characteristics already pre-existing; for example, an 'us' and 'them' dynamic in which the student is perceived within the classroom as holding a lesser status than the teacher. Further dynamics may also be present such as a senior professor being schooled by a junior research officer as a result disengagement perception of usefulness within these students occur.

2.2.4. Passive and Experiential learning. Experiential learning (EL) is a philosophy of education and process of learning whereby knowledge is created through students' active participation with an experience, their reflection on the experience and their role in it, and how it relates to theory (Kolb, 1984). Active learning, a key component of EL, assumes that knowledge is actively constructed by people, rather than passively received (Diamond, Koernig, & Iqbal, 2008). EL holds similarities to SIT, but it emphasizes the learning mechanisms to be achieved through critical self-reflection of completed tasks. Most importantly the educational benefits from EL range from more meaningful learning experiences (Granitz, 2001), increased student interest, motivation, and enthusiasm (Canhoto & Murphy, 2016; Feinstein et al., 2002; Karns, 2006), and the development of students' critical thinking and analytical decision-making skills (Dahl, Peltier, & Schibrowsky, 2018).

The main benefit of using passive learning is that a large amount of material can be delivered in a brief amount of time (Miner, Das, and Gale, 1984), but the value is debated. Engagement with pre-course reading materials are often poor, due to a lack of interest and perception of relevance towards the course (Fitzpatrick and McConnell, 2009). Context can also be lost: Singer and Diab (2020) found that participants who received passive learning on ethics were significantly more likely to perceive an unethical situation as ethical compared to active learners..

2.2.5. Humanistic model of learning. A humanistic model of learning places significance on the teaching environment, to create a social experience that students can relate to and naturally want to engage with. Humans by nature are social animals, driven by an innate desire to be part of a community with which we can identify (Greene & Burke, 2007; McNeill, 2015); the teacher is a catalyst for relationship building across the students, not teaching but developing the environment from which students discuss and work out their answers together. Therefore learning takes place through reflective discussions, applications and analysis of one's own experiences, and critical examination of beliefs and learning materials. These discussions take place within a supportive group environment (Branch, 2015). The humanistic model emphasises the need for students to find the material engaging and relevant to their context and situation.

The humanistic model also considers the different factors affecting group motivation and cohesion: for example, intrinsic and extrinsic influences are most effective during a collective group activity, in contrast to a solitary activity (Browning, 2014). If students are forced into learning through external pressures they are unable to identify with (for example legal threats) then the student is less likely to engage (Branch, 2015). The implication is that telling a researcher you must do this course in order not to go to prison is likely to result in a disengaged and unmotivated student.

2.2.6. *Learning models: summary.* Although the theories differ in details, and sometimes approach, there is much common ground: learning is most effective when it

- is active
- is of active interest to the student
- builds a bond between students and teacher, and
- builds a bond between students

The exception to this is the behavioural conditioning approach, with its relatively simplistic reward/punishment mechanisms. There are, of course, occasions where behavioural methods and/or passive learning is effective – for example, in the large scale delivery of information not requiring context.

Traditionally confidentiality training materials have come in the form of passive learning via operating guidelines or training focused on information delivery, including communication of punishments (Desai and Ritchie, 2009). The student receives material with little interaction with the teacher and is expected to assimilate the knowledge by themselves. Students complete the tasks individually and there is a little comparison of answers amongst peers. The responsibility of learning is placed on the user: the user has been given the information necessary to act correctly, so if something goes wrong it is the user's fault for ignoring the training. The teacher is not accountable for making the information accessible or for the acquisition of knowledge. The psychological model is conditioning, and the underlying theory of human behaviour is the *homo oeconomicus* of neoclassical economics (Ritchie and Welpton, 2014). This is the dominant model which the new training set out to change.

2.3. Institutional factors affecting training success. Integrating change within the workforce and particularly the facilitators is complex- with issues surrounding perception and power. Trewin (2007) noted that cultural change was required to support any form of legislative or procedural change in microdata access. Mohrman, Cohen, and Mohrman (1995) noted that for team-based organizations to function effectively, integration and cooperation across teams is imperative. Furthermore, Oaker and Brown (1986) found that when groups held a strong identification with the organization this lead to better intergroup relations: there is shared group identity across subsectors, which in our context would be the data users and the trainers.

When considering change within the statistical organisation Lewin's (1951) theory of planned change outlines organizational change within three different stages:

1. Unfreezing: reducing resilience to the proposed change and instigating a need for change within the team.
2. Change: integrating the change into practice.
3. Refreezing: 'refreezing' the changed behaviour to preserve the new state of the organization.

There is an invisible structural component consisting of elements encapsulating organizational cultures such as employee values, beliefs, and attitudes (French and Bell, 1990). Drupsteen and Hasle, (2014) outlined the literature surrounding conditions that hindered organisational learning from incidents, such as a lack of trust (Pidgeon and O'Leary, 2000, Chevreau et al., 2006), a blame culture (Dekker, 2009), a limitation in the competences of the people involved (Hovden et al., 2011) or resistance to change (Lundberg et al., 2012). Trust and openness are considered to be necessary values within an effective organization.

The relevance of this is that models of confidentiality management generally are ‘defensive’, ‘policing’ models. The conditioning approach to training fits in with this model well, by training to ensure conformity with this worldview. Training which is based upon community support, social engagement and joint learning does not fit this model well; it fits even less well when, as in the case of the new training, the fundamental but implicit *homo oeconomicus* assumption is being challenged. Training which seeks to change individuals behaviour from A to B, but which then send them back into an organisation which only recognises type A behaviour, will struggle to have long-lasting effects. A key part of effective training is therefore to help the individual change the organisation as well.

3. APPLYING PSYCHOLOGY AND PEDAGOGY TO USER TRAINING

3.1. Design principles.

3.1.1. *Aims and objectives.* The aims and objectives of the Safe Researcher Training (SRT) course were agreed with the two expert groups (one of senior managers in data organisations and academia, and one of individuals involved with running secure facilities). The agreed learning objectives were (see Green and Ritchie, 2021, for the detailed list):

- Building a sense of belonging to a data community
- Helping researchers understand risk assessment and management
- Helping researchers understand how - and why - secure data facilities are designed
- Providing a deep understanding of statistical disclosure control concept.

With the exception of the last, these objectives were designed specifically to build social identity.

An example of how these objectives lead to new outcomes is the training related to legal requirements. It is common in this sort of training to provide specific information on the legal framework. As Green Ritchie and Tilbrook (2022) discuss, this is largely there to comfort data holders; it has no practical value. It is more likely to be counter-productive, irritating course attendees by implicitly accusing them of untrustworthiness. In contrast, the SRT emphasises the way that systems are designed to prevent unlawful activity, and encourages attendees to see themselves as working with the data holders with the shared goal of avoiding unlawful behaviour.

3.1.2. *Pedagogical/psychological model.* As noted above, while there are different theories of learning, there is much agreement on the use of active learning, in the value of social interaction and feedback, in building upon trainees existing knowledge and experience, and on creating a bond between teacher and pupil. All of these indicate a need for training which is heavily activity-led and provides opportunities for reflection. This directly contrasts with the ‘information-delivery’ model often used by other courses.

The link between psychological theory, learning model and exercises can be easily seen:

Theory	Learning activities	Exercise (see next section)
SIT	Group interaction facilitates learning and the formulation of a shared identity	Introductory task
Experiential	Active participation with an experience, reflection on it and their role within it	‘Understanding your role’
Humanistic	Reflective group discussions, analysis of own experiences, and critical examination of beliefs and learning materials.	‘Who is in your team?’
Cognitive	Directly relevant materials designed to encourage assimilation of specific information	SDC

The teacher’s role within data protection training has often reflected the ‘policing’ model of data protection: *”Be grateful, be careful, or you’ll go to jail/be fined/lose your job”* (Desai and Ritchie, 2009). In the SRT, the teacher now adopts a facilitator role which allows them to exchange knowledge; the trainee can provide feedback on the teaching and also discuss potential ideas and solutions. The training also emphasises the role of the researcher’s (future) support team, who will act as facilitators in the research community once the research gets access to the data. This builds the impression that the whole process of confidential data access occurs in a facilitated supportive environment.

Expectations about the trainee also contribute to course design. Confidentiality training is usually formulated under the premise that the trainee is untrustworthy until trained and understands the punishments for misbehaviour. In addition, there is sometimes little acknowledgment of the large profile of existing expertise a researcher holds. In contrast, the SRT aims to enable the professional to effectively engage with the system, and understand and appreciate the reasoning for protocols.

The SRT design acknowledges that researchers are intrinsically motivated; training therefore consists of showing how positive behaviours contribute to that intrinsic motivation. Researchers are treated as partners in data protection, not risks. Researchers are generally keen problem solvers, who are likely to respond well to solution-focused activities. Researchers are also demonstrably able to understand nuanced arguments; in practice, they thrive on being asked to find flaws in arguments and challenge statements. Finally, all activities are rooted in real-life scenarios that researchers can understand and appreciate, even if they had not had direct experience themselves.

Thus the course aims to explicitly create a three-way dynamic (Fig. 2):

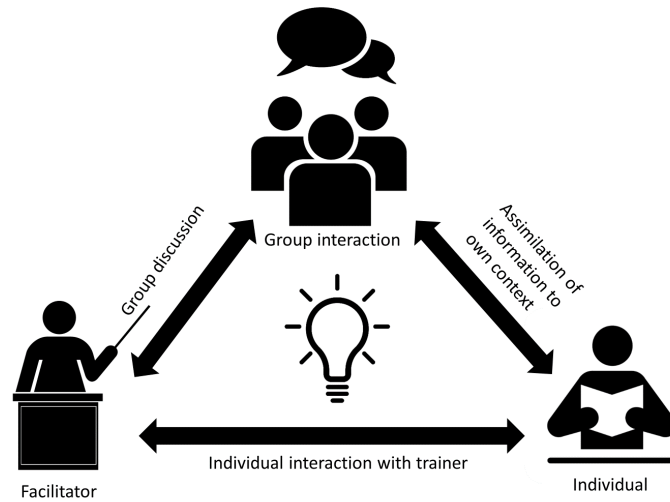


Figure 2: Conceptual framework for classroom learning

3.1.3. *Material preparation.* The material was delivered using PowerPoint™ slides. These were used to provide structure, introduce exercises, and summarise. Minimal information was given in the slides, by design, as the exercises were to be the main teaching tool. However, each slide (save placeholders) contained “Learning objectives” and “Notes” in the notes section of the slide. These discuss the exercises and have detailed information on the interpretations (Fig. 3).

Learning objective for this slide:

- To consider the balancing act of the support team

Notes:

- The support team have an important role to play in balancing the interest of the data providers and researchers
- This should be a two-way process – if the system does not meet your researcher needs, talk to the support team
- It might well be that the support team is unaware of your needs (particularly if they don't have research experience)
- The support team might also have a role to play in communicating the needs of research users to data providers
- this helps to reassure data providers that there is value to society in making data available

Figure 3: Learning objectives and notes on the slides (excerpt)

The slides were circulated after the session. The purpose of this structure was to ensure that trainees were not distracted by text on the slides during the course, but had a detailed and accurate analysis to review post-training.

3.2. **Illustrative exercises.** The following exercises provide a snapshot into the delivery of the course and the theoretical underpinnings behind the activities. We show four examples

- An initial quiz, given without preparation, to introduce ideas of nuance and complexity but primarily to build attendees' confidence and gain engagement

- Exercises to explore how the researcher fits into the research community, to strengthen intrinsic motivation
- Exercises exploring implicit biases, to help researchers examine their motivations
- Exercises in statistical disclosure control, to develop technical skills

The order of delivery of the activities can vary from trainer to trainer, but the first exercise is always the same.

3.2.1. *The impossible quiz: building confidence and engagement.* The SRT begins with a multiple choice quiz about data governance, run in three stages. The trainees complete the quiz on their own; they then discuss with a neighbour and must come to agreed answers, with the trainer circulating between groups; and then the trainer facilitates a broad discussion between all groups.

The quiz is designed to have multiple plausible answers; for example, the first question (and only one answer allowed in this case) is

The best approach is to protect the confidentiality of data...

- ... if the data are sensitive.
- ... unless the data are already in the public domain.
- ... when data are deemed to be personal.
- ... regardless of what the data are about.

The trainer can set up conflicts between trainees and groups. Some questions have no ‘right’ answer, some only have a wrong answer, some have multiple answers. To help, the trainer has a ‘crib sheet’ with prompts and suggestions on how to stimulate discussion; see Green and Ritchie (2021, Annex A) for an example.

The purpose of introducing conflict right at the beginning is not to knock trainees’ confidence, but to build it up: the conflict is resolved by trainees expressing opinions, allowing them to realise that they have valid views to contribute. The questions demonstrate to the trainee that they have meaningful prior knowledge, but which they are unlikely to have spent time considering before. Indeed, this exercise ends by asking the trainees “Have you learned any new information?”. The correct answer to this is “No, I’ve just been forced to reflect anew on things I should already know”. Providing further information through discussion rather than didactic presentation helps assimilation, the self-efficacy model of achievement.

The aims of this exercise are therefore threefold:

- To place a marker right at the beginning that participation is required
- To require participants to begin detailed discussions with at least one new person (participants are encouraged not to sit with their colleagues)
- To illustrate that the reasoning process is more important than getting the right answer

The process of answering the questions independently, then working in pairs or trios, and finally presenting your case to the whole group allows the trainees to formulate their own identity within their groups. Gaining feedback on their responses forms scaffolding and critical analysis of the learning material, and in turn helps motivate, engage and reinforce learning. On completion, the facilitator has a sense of the trainee’s capabilities, and the trainee’s prior expectation of the course (typically, that this will be a dry presentation of legal material) has been confounded and channelled in a positive direction.

Finally, it is also worth noting that this structure allows trainers much freedom to develop a teaching style, perhaps involving humour or a more animated delivery style. In

contrast, the course which SRT replaced, which required trainers to stick very closely to a script, generated resentment of the material amongst both trainers and trainees.

3.2.2. *‘Understanding your role’: seeing yourself as others see you.* Trewin (2007) stated the need to ensure that researchers understand the reasons NSOs are so protective of confidentiality. One key learning outcome for the course is for the researcher to understand that they are part of a wider system; particularly to understand why restrictions may be placed on their behaviour.

In this activity (see Fig. 4) we ask trainees to consider outlining the goals and objectives of each party: data providers, the support team, the researchers themselves, and the ‘research users’ (policy makers and others who might commission or refer to research).



Figure 4: Roles in the community (ONS, 2019)

Trainees are invited to discuss the different goals and interests each party has. We particularly highlight the concerns of the data provider; these are generally “default closed” (no data is shared unless it is provably safe; Hafner et al, 2017). Trainees are prompted to consider why this should be, forcing them to consider the data provider’s perspective. The comparison of the data provider and user allows trainees to understand that data users tend to overestimate their actual and perceived trustworthiness, and often do not consider the organisational steps and risks undertaken to make data available.

This is experiential, in that trainees are asked to imagine themselves in the role of the different actors, and then consider how they would behave. The aim is that, by asking trainees to project their perceptions about each role and its concerns, trainees change their perceptions of themselves and others that they interact with.

3.2.3. *‘Who is in your team’? Exploring implicit bias.* Implicit bias is an unconscious dispositions or biases (Allport, 1954), and so we are not aware how it shapes our decision making. For example, researchers may ignore negative behaviours by co-workers because of the ‘halo effect’: “she shouldn’t be doing that but she’s a top researcher, so it’s probably fine and I won’t complain”. Letting unacknowledged personal feeling override good sense is a potential security risk, and so the SRT seeks to make users aware of their own biases.

Trainee groups are presented with a pack of cards. On each card is a photograph of a film character and individual characteristics (not necessarily related to their film character). The trainees are then asked “who would you want in your team working on confidential data?”, and to group the cards into three categories; definitely yes, unsure, and definitely not. Purposefully some cards hold more positive individual characteristics but a ‘negative’ photograph (a physically unattractive character), or vice-versa. Six of the characters are described thus:

- Intelligent; cold and forbidding manner; sticks to rules but can be persuaded to bend them
- Team player; lazy; takes path of least resistance
- Friendly and positive; well-intentioned; loose talker; will break rules for a good cause
- Intelligent; confident; brushes over mistakes; likes to show off; wants to be admired
- Charismatic; pushes boundaries; intelligent; breaks rules when unhappy with authority
- Sticks to rules; helps people he likes; not friendly or approachable; good at spotting others’ mistakes

The class then gather round the results, and the teacher probes why groups put individuals in particular groups. Individuals often project the film characters onto the images, despite being told explicitly not to. Trainees who did not recognise the characters (hold prior knowledge) were often swayed into agreeing with the other trainee’s decisions based on group dynamics; they trusted the other trainee’s prior knowledge, despite this explicitly being made irrelevant. The aim is for trainees to become cognisant of the role of implicit biases.

For two cards the only description is ‘Professor’ or ‘PhD student’, with no further information; these should clearly be classified into “unsure”, but attendees often allocate to ‘yes’ or ‘no’ based on either preconceived ideas of these roles, or of the film characters. In particular, we note that civil servants are more likely to place ‘professor’ in the trustworthy group, while academics are ambivalent but when pressed are much more likely to see PhDs as more trustworthy than professors.

Finally, the exercise discusses training options – can poor behaviours be changed, if so, how, and if not, what other incentives might encourage positive behaviours. This discussion allows trainees to explore their behaviours alongside considering the behaviours of others.

3.2.4. *Cognitive training in SDC: how do you get the hard facts in?* A required output of the course is to give trainees a basic understanding of statistical disclosure control (SDC). This is concerned with checking that no individual data is being revealed from statistical outputs (such as the wage of a unique, identifiable individual) and taking corrective action. The course mostly uses simple tables of data; these are both the riskiest outputs and the easiest way to illustrate the issue; for some will be the only statistics they produce.

However, most researchers will use much more complex statistics, such as regression models. These are dealt with using the concept of ‘safe’ and ‘unsafe’ statistics (Ritchie, 2008): some things are inherently more risky than others, so focus on the risky ones. The SRT

uses metaphors (lions and rabbits) to represent higher or low risk, providing an anchoring framework to facilitate understanding of new concepts (Evans and Evans, 1989). At the teacher’s discretion, the choice of images (dangerous fully-grown lion, cute lion cubs, regular rabbit, mad killer rabbit from Monty Python) allows the teacher to bring humour into the classroom, which can aid engagement and focus (Ziv, 1988). Note, however, that some tutors are uncomfortable with metaphors, and simply present the technical definitions; we return to this later.

Having introduced the concept, the trainees are invited to take sticky notes and write, one statistic to a note, the sorts of statistics that they will be generating. They are then shown a white board with a spectrum marked “safe” at one end and “unsafe” at the other, and asked to place their sticky notes appropriately. A group discussion takes place about the appropriateness of the placing. One successful way to stimulate discussion is to ask a trainee to identify a sticky note that is in the ‘wrong’ place, move it to a more appropriate place, and explain why.

Unlike most of the other exercises, this is designed to instil specific technical knowledge. The theory of this is technically complex; but the trainees only need to grasp the basics of it to work effectively with confidential data. Trainees can strategically scaffold their current knowledge and perceptions of the SDC review level for their chosen statistic. So the key to the success of this exercise is giving trainees enough information, in a way they can understand, without them being overwhelmed by the information or baffled by complex theory.

4. DISCUSSION

In this section we consider lessons learned in respect of course design, implementation, and the implications for further on-the-job training.

4.1. Course design. The initial course was commissioned in April 2017 and the first courses delivered in July 2017. Given the radical nature of the course, it is not entirely surprising that almost all of the material used in the pilot was either dropped or radically redesigned after the first delivery. The course continued to evolve until September 2019, from which point the ‘canonical set’ of slides have been fixed; hence training material is now dated as ‘ONS (2019)’.

Some of the initial design problems identified were

- Material reflecting the designers’ interests rather than elements of value to the trainee. For example, the initial materials sought to explain risk management strategies of data holders as something intrinsically interesting, rather than as a way to explain the attitudes of data holders
- Intellectual rigour at the expense of user engagement; for example, early materials explained the detailed psychological models, but this was steadily whittled down and finally removed completely in June 2019

The course continued to evolve with feedback from participants and trainers. The latter was more important: participants are less likely to see if exercises work well or not as they have no comparator, whereas the trainers were in a position to review and compare how multiple sessions went. In this way, an early move to a training team helped the development

of the course. Although material shifted and changed, the pedagogical and psychological principles were consistently applied and implemented with each change.

4.2. Implementation and training-the-trainers. Reflecting on the implementation of the new training, we experienced sociological, political and psychological factors that hindered the full integration of the new approach.

First, the ‘facilitated’ learning’ approach differed considerably from the ‘chalk-and-talk’ approach taken before. The latter is straightforward for someone with minimal training to deliver (for the SRT’s predecessor, trainers were told not to deviate from the slides or script). In contrast, SRT requires trainers to step back from leading the conversation, and helping trainees discover their own understanding. Trainers need more self-confidence, letting themselves to be taken to unexpected places in discussions. While the course designers are professional educators, the trainers generally were not, and the adjustment to facilitated learning was hard and needed significant mentoring.

The second issue arises from this. In the ‘chalk-and-talk’ model, the lecturer is clearly the leader of the class and has a job to impart information. In the new model, the trainer now becomes a supporting figure, offering advice but working with the trainees. This is a problem as the trainees were often senior academics, whilst the trainers were mostly relatively junior Civil Service and academic staff. The SRT facilitated approach opens doors for individuals to exert their power, and occasionally abuse it – some trainers felt they did not have the authority to argue with senior academics. Developing the confidence of trainers therefore becomes important. This finding interlinks previously discussed models: our trainers were learning how to train, for example in understanding the need to raise staff confidence and knowledge which in turn promote the individual’s self-efficacy (Schunk, 1987), or understanding the nature of change in organisations and the tensions it can generate (Lewin, 1951).

Third, the new approach required a deeper understanding of concepts than previous similar courses. For example, confidential data users are typically required to apply a threshold rule (minimum number of observations) to ensure that privacy is not breached. Often this is taught as given, without explanation (Ritchie, 2022). The SRT required facilitators to understand how thresholds were derived and help the class to develop that understanding. This was resisted: one trainee facilitator complained “why can’t we just tell them what the rule is?”

Finally, not all exercises were well suited to all trainers’ temperaments. For example, in the implicit bias exercises with film characters, there was resistance. Trainers noted that trainees would drift off and talk about the characters, or even raise questions about the diversity of characters presented. Clearly the underlying issue in these cases is that the trainer has lost the attention and/or confidence of the classroom, and this seemed to arise because the trainer conveyed their hesitancy to the class. The responses are illustrative of how trainers defaulted to the ‘easier’ option. Whilst two organisations continued to use the recognisable faces, others replaced them with silhouettes above the descriptions. Ironically, this did not prevent trainees getting sidetracked, for example by trying to guess which silhouettes were male and which female; ultimately, those organisations replaced the silhouettes with ungendered stick figures. Ultimately this made the implicit bias exercise pointless for those groups, as trainees were now concentrating on the words rather than making judgements based on faces. As a result of this, a number of exercises were designed with optional variants to allow trainers to present in the ways that they felt more appropriate.

Although some trainers adopted the new approach, there was continual push-back. We witnessed a resistance to integrate the new course materials and teaching style; invariably the trainers changed activities to match a traditional teacher-student format with right and wrong answers. This reflects the move from rules-based training to principles-based education: rules are easier to teach and do not require a conceptual understanding of origins.

It was apparent that there were organizational issues and barriers which prevented the successful integration of the new course materials. Although the facilitators were aware of the potential enforcement of the new course materials it was obvious that the activities were not being conducted as outlined in every session. Within the theory of planned change (Lewin, 1951), resilience and instigating a need for change within the team are the components that cause 'unfreezing'- which in turn provides a platform for change. Planning for a contextual change appears equally important as planning for the actual practice.

4.3. Learning outcomes. The course developed over two years, with more trainers came from a variety of organisations with their own institutional preferences. As everyone attending the SRT takes the same test, it is possible to compare results. These show no statistically significant variation in test scores across training organisations. There is also little variation over time in results: the failure rate has stayed at around 5% since the beginning. Moreover, these results have continued to hold into the pandemic, despite widely different approaches to online teaching.

This suggests that the base material is fairly robust, at least in terms of what the test is measuring. However, this could be that the test itself includes a fair degree of learning (see Green and Ritchie, 2022), and so may act to 'smooth out' training variation. Moreover, test performance is not necessarily an indicator of post-test performance. None of the organisations providing data services currently collect formal data on compliance of users with course objectives, partly because the difficulty of defining an effective measure of non-compliance. There is some evidence that the emphasis on attitudes works. The authors used the same approach to train a project team (about fifteen people) in data governance. A formal review of the effectiveness of the training three years later showed that (a) trainees had almost no recollection of the specifics of the training, but (b) they had absorbed and were promoting the attitudinal components, even if unconsciously. There is potential for a substantial research project to consider this.

4.4. Robustness in different contexts. We have delivered all or part of the material to a range of audiences (confidentiality experts, human resources teams, IT managers) and in different cultural environments (Anglo-Saxon countries, European research groups, researchers in low- and middle-income countries). In addition, some organisations have selectively taken exercises and applied in their own context.

The picture that emerges is that reactions to the training materials and exercises are fairly consistent and predictable, irrespective of context. It seems likely that this is because the course focuses on human psychology rather than any specific cultural context. Similarly, the course is required in the UK for all researchers using the Digital Economy Act 2017 as the legal gateway, but the increasing adoption of it for areas where it is not a legal requirement suggests that the generic messages contained in the course are the valuable elements.

Finally, SRT was designed for social scientists. However, in the UK at least it has transferred smoothly to health and other applications. Again, it seems likely that this is because it focuses on human behaviours, and not on specific legal frameworks.

4.5. Online training. In 2020 the pandemic meant that the SRT had to be delivered online. The organisations delivering the training took different approaches: some reduced the amount of exercises and concentrated on information delivery; some focused more on the exercises and provided information through pre-course reading; others took a position somewhere between these two. The ‘canonical slide set’ used in the face-to-face SRT was kept online as reference material.

The SRT appears to have proved fairly robust to these differences in delivery. Attendee satisfaction remains high, and test scores do not appear to have fallen significantly. We suspect that this robustness is a direct outcome of the psychological models underlying the SRT. By focusing very heavily on the relationship between teacher and student, the specific mode of delivery becomes less significant. We will review this in a separate paper.

5. CONCLUSION

Training is just one of the components that make up effective data governance, but it is increasingly recognised as one of the most important elements. For the TREs that are now dominating access to the most confidential data, training researchers is not just about security, but also about efficiency of services.

The SRT was commissioned in response to concerns that previous training would not satisfy incoming legislative changes, as well as concerns in materials and delivery. Previous training was seen as rigid, unengaging, technical but uninformative, and fell below good pedagogic practice. For researchers it was a hurdle to be jumped before access could be given; for some types of trainer, it was an opportunity to make sure researchers knew who was in charge. It took no account of the interests or attitudes of attendees, and was largely designed to shift responsibility onto researchers in line with the ‘defensive’ data access model that is common in data protection decisions. The training channelled common assumptions: that researchers were not interested in this material or in becoming part of the community, that they could not appreciate the nuances of data protection, that it was necessary to provide researchers with a lot of information up front, and that this would always be seen as a burden to researchers. These assumptions also arose from the defensive mindset.

The SRT, in contrast, was embedded in EDRU values: evidence-based, default-open, risk-managed, and user-centred (ADSS, 2016). Long experience of working with researchers showed that past training assumptions were false, and that trust-based models helped build reciprocal trust and engagement.

The SRT was designed from the ground up, with pedagogical and psychological principles first and foremost, to deliver a quite different set of objectives:

- help researchers understand how data governance works
- show researchers their role in the data community
- build a trust relationship between researchers and their support groups/trainers

These even applied to the technical elements such as statistical disclosure control, where researchers were helped to understand how rules were developed, rather than just being given to them.

The focus on how researchers learn, rather than what individuals need to be taught, means that the latter was delivered far more effectively than traditional passive learning methods. The use of active learning, and the emphasis on community and identity throughout, also means that individuals are likely to retain more of the information from the course.

The SRT is a harder course to teach than traditional trainer-led sessions, as it requires a high level of self-confidence in the trainer and a deep understanding of the subject material. This caused problems in implementation, which was partly remedied by allowing more flexibility for trainers to teach in their own way. However, robust test scores suggest that the core material is strong enough to allow substantial variation in practice.

There are two areas which further research would benefit.

First, test scores do not necessarily reflect post-test performance. There is value in reviewing post-test activities to see whether the SRT genuinely manifest itself in 'positive' behaviours.

Second, in this paper we have focused on the face-to-face training operating 2017-2020. The pandemic resulted in a move to online training, with different organisations taking very different approaches. We will be reviewing this in conjunction with other trainers.

6. REFERENCES

ADSS (2016) *Data access strategy: final report*. Australian Department of Social Services. <https://uwe-repository.worktribe.com/output/908255>

Alvermann, D. E., Smith, L. C., & Readence, J. E. (1985). Prior knowledge activation and the comprehension of compatible and incompatible text. *Reading research quarterly*, 420-436. <http://www.jstor.org/stable/747852>

Anderson, J. R., Reder, L. M., & Simon, H. A. (1997). Situative versus cognitive perspectives: Form versus substance. *Educational researcher*, 26(1), 18-21. <https://doi.org/10.3102/0013189X026001005>

Branch, W. T. (2015). Teaching professional and humanistic values: suggestion for a practical and theoretical model. *Patient education and counseling*, 98(2), 162-167. <https://doi.org/10.1016/j.pec.2014.10.022>

Brandt, M., Franconi, L., Guerke, C., Hundepool, A., Lucarelli, M., Mol, J., Ritchie F., Seri G., and Welpton, R. (2010). Guidelines for the checking of output based on microdata research. https://www.academia.edu/38191675/ESSNet_SDC_-_Guidelines_for_the_checking.pdf

Broadbent, D. E. (2013). *Perception and communication*. Elsevier. https://pure.mpg.de/rest/items/item_2300885/component/file_2300884/content

Browning, S.L. (2014). Practitioner application. *Journal of Healthcare Management*, 59(4), 303-304. <https://doi.org/10.1097/00115514-201407000-00010>

Canhoto, A. I., & Murphy, J. (2016). Learning from simulation design to develop better experiential learning initiatives: An integrative approach. *Journal of Marketing Education*, 38(2), 98-106. <https://doi.org/10.1177/0273475316643746>

Chevreau, F. R., Wybo, J. L., & Cauchois, D. (2006). Organizing learning processes on risks by using the bow-tie representation. *Journal of hazardous materials*, 130(3), 276-283. <https://doi.org/10.1016/j.jhazmat.2005.07.018>

Corti L., van den Eyden V., Bishop L. and Woolard M. (2019) *Managing and Sharing Research Data: A Guide to Good Practice* (2nd edition) Sage. <https://doi.org/10.1108/rmj-09-2014-0035>

Dahl, A. J., Peltier, J. W., & Schibrowsky, J. A. (2018). Critical thinking and reflective learning in the marketing education literature: A historical perspective and future research needs. *Journal of Marketing Education, 40*(2), 101-116. <https://doi.org/10.1177/0273475317752452>

Dekker, S. W. (2009). Just culture: who gets to draw the line?. *Cognition, Technology & Work, 11*(3), 177-185. <https://doi.org/10.1007/s10111-008-0110-7>

Desai T. and Ritchie F. (2009) Effective researcher management. Paper presented at UNECE Worksession on Statistical Data Confidentiality 2009. <https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2009/wp.15.e.pdf>

Diamond, N., Koernig, S. K., & Iqbal, Z. (2008). Uniting active and deep learning to teach problem-solving skills: Strategic tools and the learning spiral. *Journal of Marketing Education, 30*(2), 116-129. <https://doi.org/10.1177/0273475308317707>

Drupsteen, L., & Hasle, P. (2014). Why do organizations not learn from incidents? Bottlenecks, causes and conditions for a failure to effectively learn. *Accident Analysis & Prevention, 72*, 351-358. <https://doi.org/10.1016/j.aap.2014.07.027>

Evans, R. D., & Evans, G. E. (1989). Cognitive mechanisms in learning from metaphors. *The Journal of experimental education, 58*(1), 5-19. <https://doi.org/10.1080/00220973.1989.10806518>

Eysenck, M. W., & Eysenck, M. C. (1979). Processing depth, elaboration of encoding, memory stores, and expended processing capacity. *Journal of Experimental Psychology: Human Learning and Memory, 5*(5), 472. <https://psycnet.apa.org/doi/10.1037/0278-7393.5.5.472>

Feinstein, A. H., Mann, S., & Corsun, D. L. (2002). Charting the experiential territory: Clarifying definitions and uses of computer simulation, games, and role play. *Journal of Management Development. https://doi.org/10.1108/02621710210448011*

Fitzpatrick, L., & McConnell, C. (2009). Student reading strategies and textbook use: An inquiry into economics and accounting courses. *Research in Higher Education Journal, 3*, 1. <https://www.proquest.com/openview/9db710ef822e9ed99475a5277d4ae726/1?pq-origsite=gscholar&cbl=38641>

French, W. L., & Bell, C. H. (1990). Organization development (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.

Galavotti, I. (2019). Experience and Learning: Theoretical Perspectives and Effects on Strategic Decision-Making. In *Experience and Learning in Corporate Acquisitions* (pp. 41-90). Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-319-94980-2_2

Granitz, N. A. (2001). Active learning and morality: Incorporating greater meaning into marketing education. *Marketing Education Review, 11*, 25-41 <https://doi.org/10.1080/10528008.2001.11488744>

Green, E., Ritchie F., Newman, J., & Parker, T. (2017, September). Lessons learned in training 'safe users' of confidential data. Paper presented at UNECE/Eurostat work session on statistical data confidentiality - 2017, Skopje, FYR Macedonia. <https://pdfs.semanticscholar.org/548f/4ad0434c0f67183d557fed9661bd8baa2c07.pdf>

Green, E., Ritchie, F., Tava, F., Ashford, W., & Ferrer Breda, P. (2021, July). The present and future of confidential microdata access: Post-workshop report. Presented at The Present and Future of Microdata Access URL: <https://uwe-repository.worktribe.com/output/8175728>

Green, E., & Ritchie, F. (2022) Training analysts in the management of confidential data. DRAGoN working paper no 2022/2. University of the West of England. <https://uwedragon.weebly.com/working-papers.html>

Green E. Ritchie F. and Tilbrook A. (2022) *Teaching researchers about data protection law: a terrible idea*. DRAGoN blog. <https://blogs.uwe.ac.uk/dragon/teaching-researchers-about-data-protection-law-a-terrible-idea/>

Greene, L., & Burke, G. (2007). Beyond self-actualization. *Journal of Health and Human Services Administration*, 116-128. <https://www.jstor.org/stable/41288077>

Greeno, J. G., Collins, A. M., & Resnick, L. B. (1996). Cognition and learning. *Handbook of educational psychology*, 77, 15-46. <https://people.wou.edu/~girodm/library/GCR.pdf>

Hafner, H., Lenz, R., Ritchie, F., & Welpton, R. (2015, October). Evidence-based, context-sensitive, user-centred, risk-managed SDC planning: Designing data access solutions for scientific use. Paper presented at UNECE/Eurostat work session on statistical data confidentiality – 2015. <http://www1.unece.org/stat/platform/download/attachments/109248612/Session%20%20-%20Various%20%28Hafner%20et%20al.%29.pdf>

Hovden, J., Størseth, F., & Tinmannsvik, R. K. (2011). Multilevel learning from accidents—Case studies in transport. *Safety science*, 49(1), 98-105. <https://doi.org/10.1016/j.ssci.2010.02.023>

Islam, M. H. (2015). Thorndike theory and it's application in learning. *At-Ta'lim: Jurnal Pendidikan*, 1(1), 37-47. <https://doi.org/10.1037/11243-008>

Karns, G. L. (2006). Learning style differences in the perceived effectiveness of learning activities. *Journal of Marketing Education*, 28, 56-63. <https://doi.org/10.1177/0273475305284641>

Kim, T.-Y. & Kim, Y.-K. (2015). Elderly Korean learners' participation in English learning through lifelong education: Focusing on motivation and demotivation. *Educational Gerontology*, 41 (2), 120–135. <https://doi.org/10.1080/03601277.2014.929345>

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Englewood Cliffs, N.J: Prentice-Hall. http://www.fullerton.edu/cice/_resources/pdfs/sl_documents/Exp Learning - Experience As The Source Of Learning and Development.pdf

Lewin K. (1951). *Field Theory in Social Science*. New York: Harper <https://doi.org/10.1192/bjp.114.511.794>

Lundberg, J., Rollenhagen, C., Hollnagel, E., & Rankin, A. (2012). Strategies for dealing with resistance to recommendations from accident investigations. *Accident Analysis & Prevention*, 45, 455-467. <https://doi.org/10.1016/j.aap.2011.08.014>

McNeill, D.N. (2015). Social freedom and self-actualization: 'Normative Reconstruction' as a theory of justice. *Critical Horizons*, 16 (2), 153–169. <https://doi.org/10.1179/1440991715Z.00000000045>

Miner Jr, F. C., Das, H., & Gale, J. (1984). An investigation of the relative effectiveness of three diverse teaching methodologies. *Organizational Behavior Teaching Review*, 9(2), 49-59. <https://doi.org/10.1177/105256298400900207>

Mohrman, S. A., Cohen, S. G., & Morhman Jr, A. M. (1995). *Designing team-based organizations: New forms for knowledge work*. Jossey-Bass. <https://psycnet.apa.org/record/1995-98084-000>

Oaker, G., & Brown, R. (1986). Intergroup relations in a hospital setting: A further test of social identity theory. *Human Relations*, 39(8), 767-778. <https://doi.org/10.1177/001872678603900804>

ONS (2019). Safe researcher training, canonical slide set. Office for National Statistics, September release. Available on request. http://www.saferesearchertraining.org/SRT_slides.html

Pidgeon, N., & O'Leary, M. (2000). Man-made disasters: why technology and organizations (sometimes) fail. *Safety science*, 34(1-3), 15-30. [https://doi.org/10.1016/S0925-7535\(00\)00004-7](https://doi.org/10.1016/S0925-7535(00)00004-7)

Ritchie F. (2008) Disclosure detection in research environments in practice. Paper presented at UNECE/Eurostat work session on statistical data confidentiality – 2007 pp 399-406. <https://uwe-repository.worktribe.com/index.php/OutputFile/1023059>

Ritchie F. (2017) The "Five Safes": A framework for planning, designing and evaluating data access solutions. Paper presented at Data for Policy 2017, London, UK. https://uwe-repository.worktribe.com/index.php/preview/880718/99_Ritchie.pdf

Ritchie F. (2021) Microdata access and privacy: What have we learned over twenty years?. *Journal of Privacy and Confidentiality*, 11(1), 1-8. <https://doi.org/10.29012/jpc.766>

Ritchie, F. (2022). 10 is the safest number that there's ever been. *Transactions on data privacy*, 15(2), 109-140. <https://uwe-repository.worktribe.com/OutputFile/9853171>

Ritchie F. and Smith J. (2018) Confidentiality and linked data. In G. Roarson (Ed.), *Privacy and Data Confidentiality Methods – a National Statistician's Quality Review*, 1-34. Office for National Statistics <https://doi.org/10.48550/arXiv.1907.06465> URL: <https://arxiv.org/abs/1907.06465>

Ritchie F. and Welpton R. (2014) Addressing the human factor in data access: Incentive compatibility, legitimacy and cost-effectiveness in public data resources. UWE Department of Economics Working Paper. <https://uwe-repository.worktribe.com/output/807316/>

Sadler, D. R. (2010) Beyond Feedback: Developing student capability in complex appraisal. *Assessment and Evaluation in Higher Education*. 35 (5): pp 535-550. <https://doi.org/10.1080/02602930903541015>

Schunk, D. H. (1987). Self-efficacy and motivated learning. *New directions in educational psychology*, 2, 233-251.

Schunk, D. H. (1989). Self-efficacy and achievement behaviors. *Educational psychology review*, 1(3), 173-208. <https://doi.org/10.1007/bf01320134>

Shuell, T. J. (1986). Cognitive conceptions of learning. *Review of educational research*, 56(4), 411-436. <https://doi.org/10.3102/00346543056004411>

Singer, S., & Diab, D. (2020). Ethics education: the impact of ethics training engagement on unethical decision-making in the workplace. *International Journal of Ethics Education*, 1-16. <https://doi.org/10.1007/s40889-020-00114-y>

Skinner, B.F. (1953). *Science and human nature*. New York: Macmillan.

Strickler, J. (2006). What really motivates people?. *The Journal for Quality and Participation*, 29(1), 26. <https://www.proquest.com/openview/763e6f586b889713c1382cf3815f1a33/1?pq-origsite=gscholar&cbl=37083>

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American* 223, no. 2: 96–102. <https://www.jstor.org/stable/24927662>

Tajfel, H. and J.C. Turner. (1979). An integrative theory of intergroup conflict. In *The social psychology of intergroup relations*, ed. S. Worchel and W.G. Austin, 33–47. Chicago: Nelson-Hall.

Thagard, P. (1996). *Cognitive science*. ISBN: 0262201062, 9780262201063

Thorndike, E. L. (1932). *The fundamentals of learning*. <https://doi.org/10.1037/10976-000>

Trewin D. (2007) *Managing Statistical Confidentiality & Microdata Access; Principles and Guidelines of Good Practice*. United Nations, Geneva. ISBN 13: 987-92-1-116959-1. https://unece.org/fileadmin/DAM/stats/publications/Managing_statistical_confidentiality_and_microdata_access.pdf

Turner, J. C., & Reynolds, K. J. (2010). The story of social identity. In *Rediscovering social identity: Key readings*. Psychology Press, Taylor & Francis. https://openresearch-repository.anu.edu.au/bitstream/1885/24530/8/04_Turner_The_story_of_social_id_2010.pdf

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Basil Blackwell. <https://doi.org/10.1111/j.2044-8309.1987.tb00799.x>

Wolfgang, C.H. (2001). Behavioural analysis. *Solving discipline and classroom management problems*, 5th Ed. P.21-49. Wiley/Jossey. Bass Education. <https://archive.org/details/solvingdisciplin0005edwolf>

Yilmaz, K. (2011). The cognitive perspective on learning: Its theoretical underpinnings and implications for classroom practices. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 84(5), 204-212. <https://doi.org/10.1080/00098655.2011.568989>

Zimmerman, B. J. (1995). Self-efficacy and educational development. *Self-efficacy in changing societies*, 1(1), 202-231. <https://doi.org/10.1017/CB09780511527692.009>

Ziv, A. (1988). Teaching and learning with humor: Experiment and replication. *The Journal of Experimental Education*, 57(1), 4-15. <https://doi.org/10.1080/00220973.1988.10806492>

APPENDIX A. APPENDIX: COURSE OUTLINE

The full course consists of seven sections.

Section	Purpose	Exercises
Introduction	Engage audience	The ‘impossible’ quiz
Understanding data access	Introduce the Five Safes Explain perspectives on data access	‘Uncovering’ the Five Safes
People and attitudes	Discuss the importance of ‘safe people’ Consider the difference between law and procedures, and the importance of the latter Understand the role of error and non-malicious activities in creating risky situations	What causes inappropriate data use? Breach of law or breach of rules?
Statistical disclosure control of outputs - technical	Introduce the concept of SDC: what is it, how to deal with it in simple situations; why organisations have thresholds; how to deal with more complex outputs and graphs	Repeated use of SDC examples; audience input on solutions
Statistical disclosure control of outputs - procedural	Show how to make SDC operationally efficient Introduce ‘principles-based’ SDC	Use of hypothetical release scenarios
Summary and assessment	Summarise; describe materials for circulation Describe assessment procedure and effective ways to prepare for assessment	
Using the service	Practical instruction	

Sections 1 and 6 are fixed at the start and end, but sections 2-5 are modularised so that they can be delivered in a variety of orders, For example, section 4 is specific and technical. Some trainers prefer to get this done first when researchers are fresh. Others think this is best done later when researchers are tired and just want something specific to learn. Section 5 can be done with Section 4, but it works best when delivered after Section 3 as it emphasises researcher responsibility which is covered in Section 3.

Section 7 is delivered by the service as TREs operate differently.

When the course moved online, some services dropped more or less of the elements, to be more suitable for online teaching. However, the above is the structure in the ‘canonical slides’ which are sent to all attendees after the course for review and test preparation.