Training analysts in the management of confidential data
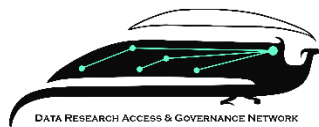
Elizabeth Green and Felix Ritchie

University of the West of England (UWE), Bristol

Training analysts in the management of confidential data

Elizabeth Green* and Felix Ritchie, University of the West of England, Bristol

*Corresponding author: elizabeth7.green@uwe.ac.uk

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 exist, often tends to focus on the 'chalk-and-talk' approach, where the aim is to ensure that the researchers are informed of their legal obligations and so take responsibility for their actions. 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, whereas evidence supports that idea that secure facilities are most effectively and securely run when there exists a positive trust relationship between users and facility managers. Nevertheless, the fondness for this approach arises from (a) a defensive attitude to data protection (b) a limited understanding of the psychology of attendees, and (c) the lack of pedagogical experience in trainers.

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 ensure buy-in from attendees. 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 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 'chalk-and-talk' model.

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

Introduction

Much of the analytical value from the great advances in computing and data availability in this century has come from the safe use of identifiable confidential or sensitive government and survey data (Ritchie, 2021). Anonymisation has a place in the spectrum 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 non-statistical governance models such as the 'five safes' (Ritchie, 2017). This portfolio approach 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 Bill.

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. Trust in researcher behaviour should also be important where data is downloaded under licence; in practice, because there are few practical limits on what researchers can do with downloaded data, data holders are likely to assume a level of data loss and remove detail accordingly.

Although in reality 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 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 largely been designed by data scientists, and usually based on the 'policing' 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. This is formulated on economic models assuming that the individual is a self-interested agent and always takes opportunities advantageous to himself: desired behaviour is achieved when the expected costs of misbehaviour outweigh the benefit. This reflects the 'defensive' attitude to data governance (Hafner et al, 2017) which still dominates government data management, although to a decreasing degree (Ritchie, 2021).

In 2017 the authors were commissioned to devise a new researcher training model ('Safe Researcher Training', SRT) for the UK Office for National Statistics (ONS). The requirement, described in Green et al (2017), was to redesign researcher training from the

ground up along the lines of the EDRU (evidence-based, default-open, risk-managed, user-centred) approach (ADSS, 2016). Rather than abstract models of 'intruders', 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 goal was to fundamentally change the attitudes of researchers. The resulting course has been franchised within the UK, and significant elements, particularly the active teaching style, have been adopted in other countries. The course covers both general approaches to data governance, and the specific 'statistical disclosure control' (SDC) techniques to manage the residual risk in published outputs using confidential data.

This paper considers the lessons learned from four years of teaching and training-the-trainers, and compares the outcomes with the ambition outlined in Green et al (2017). We use examples from the course to show how different psychological and pedagogical concepts are operationalised, and how these changed as 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 in training staff responsible for checking outputs for SDC, and in the move to online training during in response to the Covid-19 pandemic.

The next section describes the conceptual foundations of the new approach. Section 3 discusses approaches to training across countries and organisations, and how this relates to accreditation as a 'safe researcher'. Section 4 introduces the design principles, and how these relate to the underlying theory. Section 5 reflects on the lessons learned and how the course and teaching evolved over three years; it also provides some preliminary perspectives on the move to online learning in the Covid-19 pandemic. Section 6 concludes by reflecting on the sustainability of the new approach.

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 trusted research environment, and as such its influence has largely been on similar government facilities. Hence, although SRT is used for a range of researchers in the UK, we focus on the TRE users.

Confidentiality breaches may arise from both authorised users breaching procedures, and unauthorised users trying to breach data reduction or other protection measures. As this paper is concerned with training of authorised users, we do not consider attempts to hack systems or reverse-engineer statistics by unauthorised users.

Literature review

Until fairly recently (Ritchie, 2021), models for data access were built on the threat of intruders: deliberate attempts to breach confidentiality. Financial incentives (for example, selling data on the dark web or to a competitor) increase the risk of breach, but intruders may also seek to breach confidentiality restrictions for personal reasons, such as checking up on a neighbour.

The standard intruder model has 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). Altruistic behaviour is deemed irrational from the traditional economic perspective which implicitly underlies this model (Ritchie and Welpton, 2014). Whilst on the surface the default reasoning of the policing approach appears to be logical, Hafner et al (2015) demonstrate that this lacks

empirical support or theoretical coherence. Kim and Kim (2015) found that the most academically motivating factors are intrinsic - researchers value the impact of action over the monetary reward of action, suggesting that the threat of monetary fines (an extrinsic factor) is not an effective form of motivation 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.

In summary, there appears to be no literature considering the best way to train researchers in confidential data management. However, there are core psychological theories of learning which may be usefully examined; specifically, behaviourism, cognitivism, humanistic and social learning theory. In addition, we need to consider institutional factors that affect eh effectiveness of professional development courses.

Behaviourism

Hosting its origins in animal behavioural research, the classical conditioning approach to behaviour outlines the method of conditioning behaviours to respond to a particular stimulus through reward and/or punishment. Notable theorists in this area include Pavlov (1910), and Thorndlike, both researchers in animal behaviour. Thorndlike (1932) theory of connectionism believed that learning was the product when a connection had been made between stimuli and response. The theory of connectionism states that three laws are required for learning to happen. The 'law of readiness' states that the learner must be physiologically and mentally ready to learn; motivation and ability to engage are necessary components to a student's learning (Islam, 2015). The 'law of exercise' states that the more repeated interaction between the stimulus and response (student and exercise), the more ingrained the information becomes. Finally, the 'law of effect' states that when the learning response is positively rewarded it will be strengthened, but when it is punished it will be weakened.

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 – the 'law of readiness'. The teacher then needs to set the students repetitive exercises- the 'law of exercise'- and reward correct behaviour and punish bad behaviour – the 'law of effect' (Islam, 2015). The outcome is that the student will learn to automatically display the correct behaviours as they have been classically conditioned to do so (Thorndlike, 1932).

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). As such the individual can fail to assimilate the rule to similar situations as they are unable to identify similarities in the theoretical foundations. This results in a need to continually update and inform students of any new changes to the environment and reinforce the rules. This has implications for teaching addressing behaviours in variable situations. For example, SDC is implemented as 'rules-based' or 'principles-based'. The former is well suited to repetitive learning, but not to research environments (Ritchie, 2007; Ritchie and Elliot, 2015). Principles-based SDC decisions are not binary and it is difficult to generate a comprehensive list of "laws of exercise" which would satisfy the variations observed in individual analysis (Ritchie, 2007). Furthermore "laws of exercise" require repetition of tasks which often is not available within SDC: frequently researchers are working to tight schedules and if working in

a controlled environment have limited trips to complete their analysis and send it for review thereby making repetition of task difficult to achieve.

A further criticism of this model revolves around the use of punishment as a means to shape behaviour. Application of punishment within the controlled environments is always prevalent researchers are reminded of the consequences of a deliberate breach of data these punishments often include legal and financial repercussions (Ritchie *), however, it is important for the researchers to not see the data controllers as jailers as this prevents relationship development between the teacher (controller) and the student (the researcher). This power-divide quashes potential benefits (to the controller) from having researchers using the data, for example, a good relationship can result in students (researchers) relaying information about the development of informal processes, relaying information about concerns or issues within the data.

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. Informally, operant conditioning naturally occurs within many SDC centres that is output checkers know which researchers submit reasonable considered requests- these individuals are usually 'fast-tracked', whilst researchers with a reputation of poorly considered last-minute requests are often placed at the end of the list. 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). Implementing this model within a safe setting would be difficult for numerous reasons. First, a negative behavioural outcome cannot be efficiently managed without contributing further risk; second, it is not a cost-effective strategy at modifying behaviour as it depends on the individual gaining awareness.

Cognitive models of learning

The shift away from behaviourist models happened in the early 20th century. Cognitive models began to emerge, offering explanations to the internal neurological processes and environmental influences that form an individual's behaviour. Tolman in the 1920s noted that rats knew the maze in which they were placed was structured as they held 'internal maps' which the rats were able to use. This provided a foundation for the term 'cognitive maps', a core concept within cognitive models of learning (Greenwood 1999). 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 describe knowledge acquisition as a mental activity involving internal coding and structuring by an active learner (Bruning, Schraw, & Ronning, 1999) and suggests that learning happens best under conditions that are aligned with human cognitive architecture (Thagard, 1996). Therefore learning materials should be made meaningful to the learners, to help assimilate the new information to prior knowledge in memory (Galavotti, 2019). Zimmerman (1995) states that 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. Self-efficacy refers to "People's judgments of their capabilities to organize and execute courses of action required to attain designated types of performances" (Bandura, 1986, p. 391). Schunk (1987) formulated a self-efficacy model to demonstrate the relationship between internal factors and achievement (Fig. 1). This model proposes a reciprocal relationship between self-efficacy, engagement, and behaviour (Schunk, 1989). 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). Applying this model to the context of researcher training it is important to factor in our students' prior experiences of learning: 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. 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). Relevant cues used for the course include performance outcomes, persuader credibility.

APTITUDES → SELF-EFFICACY → TASK ENGAGEMENT VARIABLES → EFFICACY CUES, EXPERIENCES →
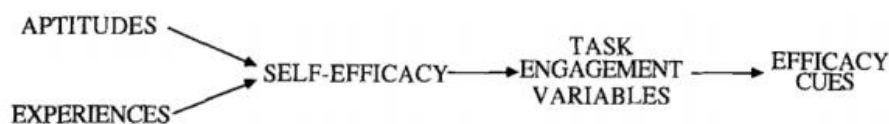
Self-efficacy model of achievement behaviour (Schunk, 1987).

The self-efficacy model allows us to understand the need to continuously engage and motivate the student throughout the course this places onus on the role of the teacher to ensure sufficient cognitive engagement with the course materials. When individuals 'fail to learn' (or, to technically describe it, fail to process information into long-term memory storage,) one core component is whether the individual at the time of processing deems the information as being relevant (Broadbent, 2013). Eysenck and Eysenck (1979) emphasised that for information to be processed into the long-term-memory storage depended on whether we deem it to have a special and meaningful relationship for the individual; therefore the individual must perceive the knowledge to be both personal and useful to them. This emphasises the need for the teacher to form individual level bonds with the researchers alongside a group bond and continually monitor efficacy cues (Yilmaz, 2011). Ensuring that users are fully engaged throughout the course is a difficult but worthy endeavor: Alverman et al (1985) found students who selectively assimilate parts of information, often distorted the information to allow 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 is also promoted self-regulation and students with good self-regulatory skills will seek assistance to clarify any points, and also further the group bond and identity (Sadler 2010).

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 is a key component of EL. Active learning is a highly personalized form of learning as it allows students to link new information to what they already know (Hamer, 2000), and assumes that knowledge is actively constructed by people, rather than passively received (Diamond, Koernig, & Iqbal, 2008). EL holds similarities to social identity theory, but it emphasizes the learning mechanisms to be achieved through critical self-reflection of completed tasks. Most importantl,y 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 later benefit is an imperative learning outcome of the SDC training as students will be required to make independent judgments on risk.

In contrast to EL, traditional confidentiality training was often provided in the form of user guides and other forms of information in which the user is to study and learn on their own. This dissemination of knowledge is known as passive learning. Passive learning is when the student receives material with no interaction with the teacher, the student is expected to assimilate the knowledge by themselves and there is no opportunity for the student to provide feedback (Stewart-Wingfield and Black, 2005). Within the traditional training, model 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's now 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 main benefit of using this approach 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. This could be due to a whole multitude of reasons: for example, participants will state it iss due to workload mismanagement and time constraints. However, Fitzpatrick and McConnell (2009) found that non-engagement with readings was often due to a lack of interest and perception of relevance towards the course.

There are many examples of passive learning tools used within the confidentiality context, for example, user guide manuals outlining different processes to ensure disclosure is controlled for. The difficulty in relying on passive learning for confidentiality training is that it is context and situation dependent, and requires the user often to make their judgments about risk, Singer and Diab (2020) examined the impact of ethics training engagement on decision making: participants were allocated into either a passive learning condition or an active (EL) condition. The study found that participants who received the passive learning condition were significantly more likely to perceive an unethical situation as ethical compared to participants in the active learning ethics training condition. This further evidences the need for experiential learning in confidentiality training.

Humanistic model of learning

A humanistic model of learning places significance on the teaching environment; in particular, group dynamics and interaction as a catalyst for individual attainment and self-actualisation (Maslow, 1970). The aim is to create a social experience that students can relate to and naturally want to engage with. The humanistic model states that when students learn

naturally, they learn more (Branch, 2015). This is a stark contrast to the traditional teaching method which focused on the student and neglected to consider the wider context of the classroom and group dynamics. Within this course, we argue that the teacher is a catalyst for relationship building across the students. The role of the teacher is not to teach, but to develop the right environment from which students discuss and work out their answers together. 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). 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 in which the individual feels secure about discussing their experiences, values, and beliefs, in turn, the security offered by the group allows the individual to consider new perspectives and interpretations on experiences (Branch, 2015). The humanistic model emphasises the need for students to find the material engaging and relevant to their context and situation. Traditionally confidentiality training materials have come in the form of passive materials such as generic examples of outputs for SDC checking; in the new course, interactive group exercises focus on the student's experiences, research projects and past experiences, allowing students to relate and engage with the course aims and materials.

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). This further supports that the idea that, given the ideal environment, a human would prefer to contribute to their community than contribute to himself (Browning, 2014). However, if students are being explicitly forced into learning through external pressures they are unable to identify with (for example legal punishment/ threats) then the student is less likely to engage with the course (Branch, 2015). The implication is that telling a researcher you must do this course in order not to go to prison is likely going to result in a disengaged and unmotivated student.

Social identity theory

Social Identity Theory (SIT) examines the relationship between cognitive processes and larger-scale group dynamics (Brown & Capozza, 2000). 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). Tajfel (Tajfel 1970; Tajfel et al. 1971) performed a series of experiments using what is now called the 'minimal group paradigm.' The aim was to assess how individuals respond to group membership even in the absence of fixed distinguished group identity. Firstly participants were asked to estimate the number of dots projected briefly on a screen. Participants were then categorized randomly into groups, but were told by the experimenters that they had either over or under-estimators on the dot task. Tajfel observed that this simple categorization was sufficient enough to generate intergroup conflict and bias of in-group and out-group members. This phenomenon was coined, 'the minimal group paradigm'. It was determined that individuals share common experiences, beliefs, and values, 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 a group feels able to mobilise their group status then they will use creative and often legal means of inspiring people to value their identity (for example through music, art, communication, etc); however, if the group does not feel they can mobilise their group status, the group will negatively compete against other groups (Turner and Reynolds, 2010). When considering this in an educational context can provide insights into poor course engagement and poor student communication.

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. The longitudinal impact of this form of teaching is that students formulate their own culture of 'not sharing' critical information with the provider as they fear repercussions; avoidance of action results in positive reinforcement as in 'I got away with it that time- I can do it again.' This heuristic spreads like wildfire amongst the student community.

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 because there is shared group identity across subsectors-which in this context would be the data users and the trainers. It is therefore essential to address and consider group dynamics within the training course and build a common group identity with the course delegates- we are all part of the research community.

Institutional factors affecting training success

Integrating change within the workforce and particularly the facilitators is complex- with issues surrounding perception and power. The Trewin report (2006) noted that cultural change was required to support any form of legislative or procedural change in microdata access, it further noted that changes to data access legislation require collaboration with a wide range of different stakeholders. When considering change within the statistical organisation Lewin's (1951) theory of planned change outlines organizational change within three different stages:

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

Within the statistical organisation there is an invisible component of the organizational structure which consists of elements encapsulating organizational cultures such as; employee values, beliefs, and attitudes (French and Bell, 1990). When considering the theory of planned change and the facilitators, it will be interesting to investigate perceptions surrounding the integration of principles, as rules are traditionally easier to teach and do not require a conceptual understanding of its statistical origins. The facilitators were aware of the new course development with many collaborating with the authors and observing the author's course sessions. The authors witnessed a resilience to integrate the new course materials and

principles in the organisation's courses, with some facilitators changing activities to match a traditional teacher-student format with right and wrong answers.

When considering the theory of planned change (Lewin, 1951) and the facilitators, it is apparent that there are organizational issues and barriers which are currently preventing 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.

When considering the failure to learn within organizations (Drupsteen and Hasle, 2014) found although a lack of trust was not explicitly mentioned by their research participants, the limited sense of urgency, lack of motivation were factors that are related to trust in the organization and its management. When reflecting on this finding could it be that facilitators believe if a breach was to occur they would be faced with blame? 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). Especially trust and openness are considered to be necessary values within an organization. Without these values, incidents will be kept secret, investigations will focus only on a selection of factors, and learning opportunities will remain unused. This further elaborates the necessity of a community-based model overarching the training course and the statistical organisation.

Current models of confidentiality training

In this section we consider the level of training and accreditation required by national statistical institutes (NSIs). We focus on NSIs as these are often the bellwether for what is seen as 'best practice' in their respective countries. NSIs sponsored or promoted much of the research into statistical confidentiality, and their actions tend to cast a long shadow over the rest of government in terms of data governance.

Many national statistical organisations require basic statistical knowledge, but not all – the UK for example does not explicitly require it. Often, this component is assessed within the application process. Some organisations (for example, Eurostat) assess the application based on the reputation of the organisation. In Belgium and Bulgaria, researchers explicitly state their experience and training of SDC within their application for microdata access. For some organisations, formal 'training' is interpreted as formal education, such as a University degree in statistics.

The ADSS (2016) report highlighted a wide variety in training practices amongst 12 RDCs run by NSIs from around the world:

Is training provided on security awareness?	9 Yes 1 Optional
How is the security training delivered?	5 Face-to-face 3 Online course 1 Online guide

| Is training provided on using the system? | 7 Yes
4 No |
|---|---|
| Is training provided on statistics eg aspects of data linkage? | 1 Yes
4 Optional
4 No |
| Is refresher training required? | 1 Yes
3 Not recent
7 No |
| Are researchers trained in checking output for disclosure risk? | 8 Yes
3 No |

Data were taken from a survey of 12 RDCs in Europe, N. America, Oceania.
Source: ADSS (2016, Appendix). Not all questions answered by all

Drawing on the ADSS findings, most organisations have passive guidelines for researchers; compulsory face-to-face training is rare. In Europe, only the UK and Dutch appear to run regular mandatory face-to-face training in using the facilities and in SDC (Eurostat report). The Dutch RDC also asks users a question on security and disclosure control at log-in (a wrong answer means login is delayed) to ensure that users get some regular reminders of the operations of the facility, SDC, and good data management. Outside Europe, statistical agencies in Mexico, Canada, New Zealand and all require some form of face-to-face interactive training: Canada has one-to-one training sessions with researchers on their first arrival.

Training may or may not be combined with accreditation, but accreditation of researchers as 'safe' (in whatever way this is interpreted) varies widely. In Hungary, the completion of the application process is also the accreditation. In Slovakia, the completion of the application process plus evidence of statistical qualification is the basis for accreditation. In the UK and Australia, persons are formally accredited as the result of simple checks on qualifications plus mandatory training. In Finland, Canada, and the European Central Bank, the signing of a 'pledge of secrecy' completes the accreditation; this is backed up by training in Finland, stringent personal background checks in Canada, but neither in the ECB. Eurostat is unusual in that the research organisation (such as the university) is the accredited party: the individual researchers need to apply for access to a specific project, but the organisational accreditation means that the project is being assessed, rather than the individuals (as long as they belong to an accredited organisation).

Every organisation provides some written guidance in procedures surrounding SDC. There is much variation. For example, Hungary produces extremely detailed SDC guidance tailored to its researcher. In contrast, for SDC guidance UK RDCs generally refer users to training material and online resources, rather than bespoke publication. When considering the pre-existing face-to-face courses available globally, many have begun to shift away from a rules-based design to a principle-based design- for example, Canada and Australia. The face-to-face training presently offered by these organisations draw upon the present papers examples.

Conceptual framework of course design

Design principles: aims and objectives

The aims and objectives of the 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

- Community: build…
    - Positive sense of research community and one's role in contributing to it
    - Understanding of importance of attitudes and behaviours
    - Understanding importance of presentation
    - Awareness of how good practice develops, and how to get things changed
    - Awareness of link between ongoing good practice and future access
    - Where to go to find more info
    - How to not annoy people and get what you want
- Risk management
    - How to deal with uncertainty
    - Overall view of risks and the evidence base
- Specifics
    - Understanding why the use of data is limited, and why using the data for its proposed purpose matters
    - Understanding how researchers contribute to exploiting the data fully
    - Understanding how attitudes and behaviours of researchers affect the whole community
    - The importance of using systems/following procedures
    - Holding data with the appropriate level of info
    - Understanding SDC and 'safe statistics'
    - Understanding PB/RBOSDC and how it relates to user/system behaviour

These objectives were designed specifically to build social identity. This contrasts with policing models where emphasis is laid on the specific knowledge gained by students, to ensure that they keep on right side of the law.

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 (in one case, testing course attendees on the level of fines and detailed interpretation of clauses in legislation). As Green Ritchie and Tilbrook (2022) discuss, this is solely there to comfort data holders; it has no practical value. In practice it is much 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.

Design principles: pedagogical/psychological model

When developing the new course for the ONS, the authors emphasized the user profile and how best to attain learning outcomes, drawing predominately on Humanistic, Cognitive, and Social Identity Theory models of learning the course aims to provide a group-led inquiry into the application of SDC. Teaching professionals is, to an extent easier than teaching the general population, as within our course profile the professionals all hold similar characteristics. Professionals attending the course are likely to: hold relevant post-graduate qualifications with training in statistics, belong to a research-aware organisation, and be attending the course because they want to use specific datasets. These individuals therefore have extensive experience of education and have completed further study in a higher education system; they are likely to gravitate toward specific learning techniques and therefore shape and consider

course design and participant engagement figure 1 outlines the actors involved and acquisition of knowledge. We applied the theoretical concepts from the different models of learning:

- *Cognitive models of learning:* Students (in this case Professionals) already have a wealth of pre-existing knowledge, therefore course materials need to be framed appropriately to demonstrate relevance and also allow for assimilation of knowledge. The introductory and SDC tasks specifically interlinks to the application of this concept.
- *Humanistic models of learning:* Learning takes place through reflective group discussions, applications and analysis of own experiences, and critical examination of beliefs and learning materials. The "who is in your research team?" exercise is an example of this.
- *Experiential model of learning:* knowledge is created through students' active participation with an experience, their reflection on the experience, and their role within it, the Understanding your role in the research community exercise is an example of this.
- *Social Identity Theory:* group interaction can help facilitate learning and formulate greater shared identity as a research community with a collective mission to protect the data. This is implemented across all the activities with an emphasis on the second part of the first task (the MCQ).
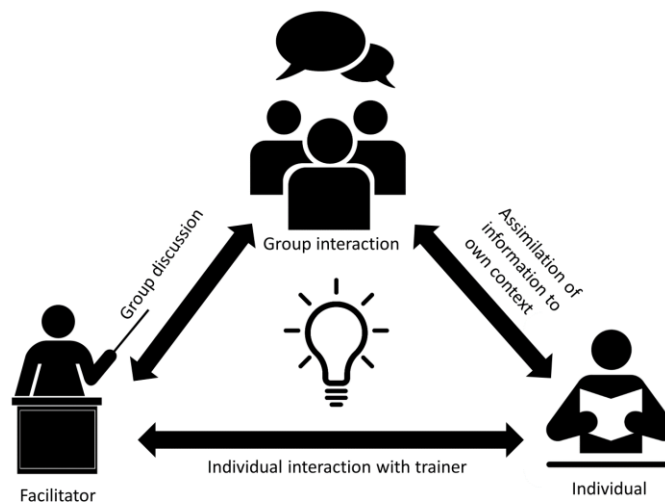

Figure 1 Conceptual framework for classroom learning

When considering the course as we have placed significant emphasis on the human and group learning mechanisms it is important to outline our expectations of the different actors and contrast this to the previous course design assumption.

The role of the teacher

The teacher's role within data protection training often reflects the 'policing' model of data protection: users are told*: "Be grateful, be careful, or you'll go to jail/be fined/lose your job"* (Desai and Ritchie, 2010). In the SRT, the teacher now adopts a facilitator role which allows them to exchange knowledge the student can provide feedback on the teaching and also discuss potential ideas and solutions. There is an emphasis on the facilitator's role within the research community, and a humanisation element of the organisational processes- often the facilitator delivering the course is also a person who checks requested outputs.

Confidentiality training is usually formulated under the premise that the student is untrustworthy until trained. In addition, there is sometimes little acknowledgment of the large profile of existing expertise a researcher holds. In practice, most of the SRT is anti-rocket-science' (see below). This is where this course distinguishes in comparison to other courses in this area, we believe that the learner would be able to engage with the system regardless of whether they have had training (many of them have had similar experiences with microdata and programming before the course), what our course is interested in is not educating the professional about how to do their job, we are interested in enabling the professional to effectively engage with the system and understand the reasoning for protocols.

As such the core foundations on which this course was developed are:

- Researchers' motivation is 'intrinsic' (that is, primarily driven by internal psychic needs rather than external motivations such as money or threats) a direct contradiction to the self-serving agent as proposed by the rational choice theory.
- Researchers like to problem solve and will formulate their strategies to simplify processes and the level of input required from them. Often if this new strategy is perceived as reasonable then over time it will be adopted as a new 'normal' behaviour, and an example of this might include propping fire doors open to gain a breeze in the room or taking down notes of code in the lab.
- Statistically, it is an inevitability that everyone (researchers and support team) will make a mistake at some point and this is not due to malicious purposeful attacks, but more likely due to fatigue, distraction, and lack of motivation.
- Researchers find it easier to rationalise their own mistakes than to admit to making them, this term is coined 'fundamental attribution error' in which we attribute own misbehaviours as a result of external and environmental causes, for example 'I was in a rush and not thinking which is why I submitted the wrong output files for checking', yet if another person did this we would attribute internal causes for this 'they are lazy' or 'intruder alert' (Ross, 1977).

Illustrative exercises

The following exercises provide a snapshot into the delivery of the course and the theoretical underpinnings behind the activities, of course, it is important to note that the delivery of the activities can vary from trainer to trainer however the outlined exercises are how the authors delivered the training. 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

Exercise 1: building initial confidence and engagement

The first warm-up activity in class was a multiple choice quiz about data governance. The exercise consists of 3 distinct stages: the students 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 facilities a discussion between groups. The aims of this are 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

For this the trainer has 'crib sheet' which contains a list of detailed question, prompts, comments to make, and suggestions on how to stimulate discussion and create (and resolve) conflict; see Annex A for an example question and the associated crib.

The questions demonstrated to the student that they did know particular aspects of data governance (thereby acknowledging student's prior knowledge), but which they are unlikely to have spent time considering before. For example, the first question 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.

Students are asked to select only one answer, and the tutor can then set up a conflict when students differ in their views with the tutor acting as devil's advocate to any response from the students – there is no 'right' answer to the question. However the aim is not to knock students' confidence, but to build it up as they realise that they have valid views to contribute to the discussion.

Other questions have 'right' answers, some have 'wrong' answers that the students should avoid, and one is purely technical question designed to show they students that there is a need to gain some specific knowledge. But generally the questions ask the student to identify material they already have a pre-existing knowledge on; providing further information through discussion rather than didactic presentation helps assimiliation , the self-efficacy model of achievement.

A. The process of answering the questions independently, then working in pairs or trios, and finally presenting your case to the whole group allows the students 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. This part of the exercise draws upon both humanistic and social identity theory models of learning, as the learning is formulated from the interaction with peers and also the development of group identity. On completion of this exercise the facilitator has a sense of the student; expectations of the course, the student's prior experiences, and present knowledge surrounding SDC. This information then allows the facilitator to assess if particular individuals require further support. 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. This is in contrast to the course which SRT replaced, which required trainers to stick very closely to a script and so generated resentment of the material.

Exercise 2: Understanding the role of the research community

One key learning outcome for the course is for the researcher to understand that they are part of a wider system, and as so each part is critical in how well the system functions. Trewin (2006) stated the need to ensure that researchers understand the reasons NSOs are so protective of confidentiality, and ensure they are aware of the consequences to them and their institution if there are breaches. The report also notes that researchers should accept they have a shared responsibility to maintain and uphold conditions under which they have been provided access.

In this activity (see figure 2) we ask students to consider outlining the goals and objectives of each party. Students are invited to discuss within their teams and also with the facilitator about what different goals and uses of data each party has.
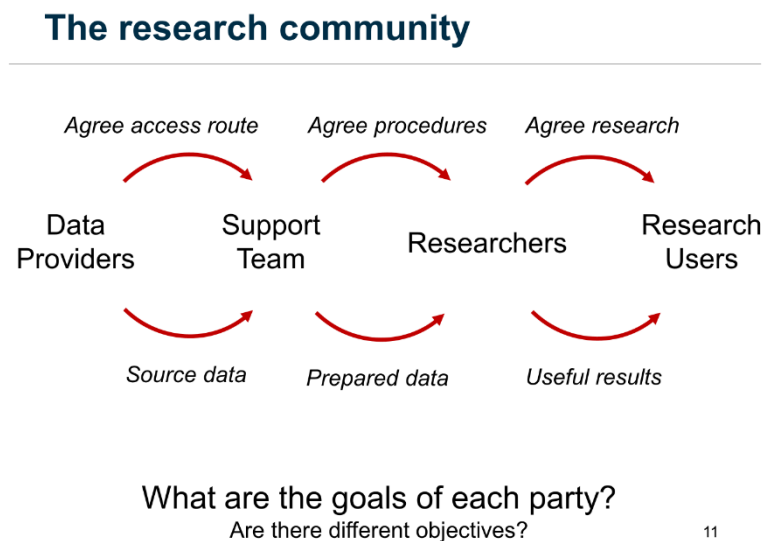

Figure 2 The research community PowerPoint slide (ONS, 2019)

In this exercise, we highlight the concerns of the data provider as they are generally "default closed": meaning their default solution is no access unless persuaded that access is valuable and safe as the data provider's goal is to keep the data safe. Students are then 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 students 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. In providing an overview of the system and different actors involved students gain an understanding of how their actions and behaviours have an impact, and why restrictions that might seem overbearing are in place.

Figure 3 The role of the support team within the research community (ONS, 2019)

This activity is underpinned by experiential learning: we are asking students to project their perceptions about different roles and concerns might be for the different parties alongside the student formulating their

critical reflections about their practices. This is fundamental to the goal o the training, to change students perceptions of themselves and others that they interact with.

Exercise 3: Exploring the role of implicit biases

Implicit bias is typically thought of as unconscious dispositions or biases Allport (1954) was the first to fully explore implicit cognition and the concept that we automatically process information, this model is part of the cognitive models of processing. The difficulty with implicit biases is that by its very nature it is automatic and often we are not aware of how the processes have shaped our decision making (Allport,1954). For an individual to align their desired intents that is our 'morally correct' behaviours they must acknowledge the automatic processing their minds undergo. The SRT includes an exercise to bring awareness to these functions and processes. Students are presented with a pack of cards. On each card is a photograph of a film character and individual characteristics, which do not necessarily bear any resemblance to the character as portrayed in film. The students 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, or vice-versa.Six o 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

On completion of the task the teacher then asks why they selected these cards. The learning outcome is to demonstrate that individuals who knew these characters externally often provide further attributes and higher levels of discrimination for each of these characters. Students who did not hold prior knowledge of these characters were often swayed into agreeing with the other student's decisions based on group dynamics and trusting the other students are more knowledgeable as they hold prior knowledge. The aim is for students to become cognisant of the role of implicit biases.

Following the group discussion, a wider class discussion surrounding the students' decision-making process is then explored. 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 much more ambivalent.

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 students to explore their behaviours alongside considering the behaviours of others.

Technical exercises: high and low review statistics

A required output of the course is to give students 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 individual) and taking corrective action. Most of the course is concerned with simple tables of data, as these are both the riskiest outputs and the easiest way to illustrate the issue; and for some trainees, this will be the only stats they produce. However, most of the researchers are planning to use much more complex statistics, such as regression models; how then can the course meet the needs of the sophisticated researcher as well as the table analyst? And how can a single course aim to cover all of the potential outputs of the former?

The approach taken focuses on a simple division between 'risky' or 'high review' statistics (HRSs) and 'safe' or 'low review' statistics (LRSs). We use a zoo metaphor to consider the risk management for the two different types of statistics;

> *'imagine you are managing a zoo, limited to two kinds of animals: lions and rabbits. You only have limited time to spend on managing these two: which do you focus on?... An angry rabbit can give you a nasty nip; a well-fed sleepy lion can be tickled behind the ears…but…in general, you should be spending your time watching the lions; it doesn't matter if the rabbits escape from their cages but the lions….The lions are HRSs, rabbits are LRSs. What this means in practice is that we spend our time focusing on HRSs where the real risk lies; we ignore the LRSs because by construction they pose no practical risk.'*

The metaphor has been acknowledged as a tool that provides an almost universal conceptualization of basic human tendencies (Danesi, 1993). Metaphors provide an anchoring framework to facilitate understanding of new concepts (Evans and Evans, 1989) they allow humans to make sense of the world in human terms that make sense to others (Danesi, 1993). The use of humour in the classroom is also associated with students' learning and motivation and can aid engagement and focus (Ziv, 1988). James (2001) studied the effect of humour on higher education lectures and found that humour helps students stay focused and engaged during the session. Note, however, that some tutors are uncomfortable with metaphors, and simply present the technical definitions; we return to this later.

Having introduced, the students 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 presented with the image in figure 3.

Defining LRSs and HRSs

- Are your statistics low or high review?
  - LRS means that you don't need to know about the data
  - some LRSs might have conditions

Low
review

High
review

Research results more likely to be LRS

30

Figure 3 Defining Low Review Statistics and High Review Statistics ((ONS, 2019)

Each student is asked to place his or her sticky notes on the board, where they think they are appropriate. A group discussion takes place about the appropriateness of the placing. One successful way to stimulate discussion is to ask a student to identify a sticky note that is in the 'wrong' place, move it to a more appropriate place, and explain why.

This activity provides an example of experiential learning, the student is required to engage and interact with the classroom- each individual is expected to contribute a sticky note. The activity is also conducted in a group with individuals discussing and sharing their opinions about whether a statistic should be classed as high or low, furthermore, some students can formulate scenarios in which a typically 'low review' statistic might prompt further review and vice versa. This allows students to strategically scaffold their current knowledge and perceptions of the SDC review level for their chosen statistic. Students then explore the role of managing a review of output, we ask students to consider themselves as output checkers with limited time and resource, Once again this exercise aims to provide both conceptual and technical knowledge about statistical disclosure control, but also to get the student to consider the wider processes which are being undertaken by the output checkers and how the student's behaviour can influence the efficacy in obtaining outputs.

Discussion

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

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 altered 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 student. 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 data holder's behaviour
- 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.

During the pandemic, training went online from March 2020 onwards. This led to a variety of approaches taken by different training teams to making the base material suitable for online teaching. This is the subject of a separate paper, but interestingly, this seemed to have little effect on test results, suggesting that the core material is relatively robust to delivery style.

Implementation of training

4. 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, and this is what had happened with the SRT's predecessor (in fact trainers were told not to deviate from the slides or script). In contrast, the SRT's model required trainers to step back from leading the conversation, and helping trainees discover their own understanding. This requires trainers to have much more self-confidence, allowing themselves to be taken to unexpected places in discussions. While the course designers are professional educators, the trainers were not; this adjustment to facilitated learning was hard, and needed mentoring and sometimes formal training in facilitated learning skills.

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 with individual behaviour models such as the transtheoretical model of behaviour change (Prochaska and DiClemente, 1986) in understanding the need to raise staff confidence and knowledge which will promote the individual's self-efficacy and understanding of distraction causing motivation and practice change.

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 threshold were derived and help the class to develop that understanding. This was resisted: one trainee facilitator complained "why can't we just tell them?"

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 non-gender 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 the felt more appropriate.

Learning outcomes and robustness

The course develop over two years, and more trainers came from a variety of organisations with their own institutional preferences. However there is very little in test scores across time or individuals. As everyone attending the SRT takes the same test, it is possible to compare results for training organisations. These show no statistically significant variation. There is also little variation over time in results: the failure rate has stayed at around 5% since the beginning. Morevoer, 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, and there is potential for a substantial research project to consider this.

Conclusion

The SRT was commissioned in response to concerns that previous training was not up to the mark required by the incoming Digital Economy Act 2017. That previous training was rigid, unengaging, technical but uninformative, and fell well below good practice; it was seen by researchers as a hurdle to be jumped before access could be given, and by trainers as 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.

The training challenged common assumptions: that researchers were not interested in this material or in becoming part of the community, that they could not appreciates 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 arose from the defensive mindset that is common in data protection decisions.

The SRT, in contrast, was embedded in EDRU values: evidence-based, default-open, risk-managed, and user-centred (Green and Ritchie, 2016). Long experience of working with researchers showed that the assumptions were false, and that trust-based models hlped 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 element of 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 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.

References

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.

Anderson, J. R., Reder, L. M., & Simon, H. A. (1997). Situative versus cognitive perspectives: Form versus substance. *Educational researcher*, *26*(1), 18-21.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice-Hall

Branch, W. T. (2015). Teaching professional and humanistic values: suggestion for a practical and theoretical model. *Patient education and counseling*, *98*(2), 162-167.

Broadbent, D. E. (2013). *Perception and communication*. Elsevier.

Browning, S.L. (2014). Practitioner application. Journal of Healthcare Management, 59 (4), 303–304.

Bruning, R. H., Schraw, G. J., & Ronning, R. R. (1999). *Cognitive psychology and instruction*. Prentice-Hall, Inc., One Lake Street, Upper Saddle River, NJ 07458.

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.

Capozza, D. E., & Brown, R. E. (2000). Social identity processes: Trends in theory and research. In *Small Group Meeting'Intergroup Relations: Current Work and Future Perspective', Sep, 1996; The idea for this book was born during the aforementioned conference.*. Sage Publications Ltd.

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.

Cowan, N. (2014). Working memory underpins cognitive development, learning, and education. *Educational psychology review*, *26*(2), 197-223.

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.

Danesi, M. (1993). *Vico, metaphor, and the origin of language*. Indiana University Press.

Dekker, S. W. (2009). Just culture: who gets to draw the line?. *Cognition, Technology & Work*, *11*(3), 177-185.

Desai T. and Ritchie F. (2009) Effective researcher management. Paper presented at UNECE Worksession on Statistical Data Confidentiality 2009

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.

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.

Evans, R. D., & Evans, G. E. (1989). Cognitive mechanisms in learning from metaphors. *The Journal of experimental education*, *58*(1), 5-19.

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.

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*.

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.

French, W. L., & Bell, C. H. (1990). Orga ni za tion devel op ment (3rd ed.). Upper Sad dle 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.

GDPR (2018) *Data Protection Act 2018*, c. 12 [online] UK Available at http://www.legislation.gov.uk/ukpga/2018/12/contents/enacted last accessed 09/12/2020 [Accessed 9 December 2020].

Granitz, N. A. (2001). Active learning and morality: Incorporating greater meaning into marketing education. *Marketing Education Review*, 11, 25-41

Green E. and Ritchie F. (2016) *Data access strategy: final report*. Australian Department of Social Services. https://uwe-repository.worktribe.com/output/908255

Green E. and Ritchie F. (2022) *Greyscale scoring for training users of confidential data.* DRAGoN working paper.

Green E. Ritchie F. and Tilbrook A. (2022) *Teaching researchers about data protection law: a terrible idea*. DRAGoN blog.

Greene, L., & Burke, G. (2007). Beyond self-actualization. *Journal of Health and Human Services Administration*, 116-128.

Greeno, J. G., Collins, A. M., & Resnick, L. B. (1996). Cognition and learning. *Handbook of educational psychology*, *77*, 15-46.

Greenwood, J. D. (1999). Understanding the "cognitive revolution" in psychology. *Journal of the History of the Behavioral Sciences*, *35*(1), 1-22.

Hamer, L. O. (2000). The additive effects of semistructured classroom activities on student learning: An application of classroom-based experiential learning techniques. *Journal of Marketing Education*, 22, 25-34.

Hovden, J., Størseth, F., & Tinmannsvik, R. K. (2011). Multilevel learning from accidents–Case studies in transport. *Safety science*, *49*(1), 98-105.

Islam, M. H. (2015). Thorndike theory and it's application in learning. *At-Ta'lim: Jurnal Pendidikan*, *1*(1), 37-47.

James, D. L. (2001). Split a Gut and Learn: Theory and Research.

Karns, G. L. (2006). Learning style differences in the perceived effectiveness of learning activities. Journal of Marketing Education, 28, 56-63.

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.

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development.* Englewood Cliffs, N.J: Prentice-Hall.

Lewin K. (1951). *Field Theory in Social Science*. New York: Harper

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.

Maslow, A. H. (1970). *Motivation and Personality* (2nd ed.). New York: Harper & Row.

McNeill, D.N. (2015). Social freedom and self-actualization: 'Normative Reconstruction' as a theory of justice. *Critical Horizons*, 16 (2), 153–169.

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.

Mohrman, S. A., Cohen, S. G., & Morhman Jr, A. M. (1995). *Designing team-based organizations: New forms for knowledge work*. Jossey-Bass.

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

ONS (2019). Safe researcher training, canonical slide set. Office for National Statistics, September release. Available on request.

Pavlov, I. P. (1910). *The work of the digestive glands*. C. Griffin.

Pidgeon, N., & O'Leary, M. (2000). Man-made disasters: why technology and organizations (sometimes) fail. *Safety science*, *34*(1-3), 15-30.

Prochaska, J. O., & DiClemente, C. C. (1986). Toward a comprehensive model of change. In *Treating addictive behaviors* (pp. 3-27). Springer, Boston, MA.

Ritchie F. (2007) Disclosure detection in research environments in practice. Paper presented at UNECE/Eurostat work session on statistical data confidentiality - 2007

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

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. and Elliot M. (2015) Principles- versus rules-based output statistical disclosure control in remote access environments. IASSIST quarterly / International Association for Social Science Information Service and Technology, 2015(Summer), 5-13

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

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/

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In *Advances in experimental social psychology* (Vol. 10, pp. 173-220). Academic Press.

Sadler, D. R. (2010) Beyond Feedback: Developing student capability in complex appraisal. *Assessment and Evaluation in Higher Education.* 35 (5): pp 535-550.

Sadler-Smith, E., Allinson, C. W., & Hayes, J. (2000). Learning preferences and cognitive style: Some implications for continuing professional development. *Management Learning*, *31*(2), 239-256.

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.

Shuell, T. J. (1986). Cognitive conceptions of learning. *Review of educational research*, *56*(4), 411-436.

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.

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

Strickler, J. (2006). What really motivates people?. *The Journal for Quality and Participation*, *29*(1), 26.

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American* 223, no. 2: 96–102.

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.

Tajfel, H., and J.C. Turner. (1986). The social identity theory of intergroup behavior. In *The social psychology of intergroup relations,* 2nd ed., ed. S. Worchel and W.G. Austin, 7–24. Chicago: Nelson-Hall.

Tajfel, H., M.G. Billig, R.P. Bundy, and C. Flament. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology* 1: 149–78.

Thagard, P. (1996). Cognitive science.

Thorndike, E. L. (1932). The fundamentals of learning.

Turner, J. C., & Reynolds, K. J. (2010). The story of social identity. In *Rediscovering social identity: Key readings*. Psychology Press, Taylor & Francis.

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.

Wingfield, S. S., & Black, G. S. (2005). Active versus passive course designs: The impact on student outcomes. *Journal of Education for Business*, *81*(2), 119-123.

Wolfgang, C.H. (2001). Behavioural analysis. *Solving discipline and classroom management problems*, 5th Ed. P.21-49. Wiley/Jossey. Bass Education.

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.

Zimmerman, B. J. (1995). Self-efficacy and educational development. *Self-efficacy in changing societies*, *1*(1), 202-231.

Ziv, A. (1988). Teaching and learning with humor: Experiment and replication. *The Journal of Experimental Education*, *57*(1), 4-15.

DRAGoN Working paper series

2021

2021/01   *A very different time and place: a cross-cultural analysis of the use of the One Minute Paper before and after the pandemic*
Damian Whittard, Elizabeth Green, and Mariyam Shaffau Shareef

2021/02   *Ten is the safest number that there's ever been*
Felix Ritchie

2022

2022/01   *Using the 'five safes' to structure economic evaluations of data governance*
Felix Ritchie & Damian Whittard

2022/02   *Training analysts in the management of confidential data*
Elizabeth Green and Felix Ritchie