

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

**A Validation of the Workplace Dignity Scale**

A thesis presented in partial fulfilment of the requirements for the degree of

Master of Arts

in

Psychology

at Massey University, Manawatū, New Zealand.

Casey Scott-Campbell

2019

## **Acknowledgements**

First and foremost, I cannot give enough praise to my supervisor Matt for his helpful advice and encouragement throughout the year, his methodological expertise, his thoroughness when responding to my page-long questions, and his contagious motivation to do good research. I also want to acknowledge the other members of the Massey Registered Report team for being an enjoyable and approachable group to interact with. I am also thankful for the Massey University Masterate Scholarship which supported me to complete this thesis, and lastly, to Ryan, for being consistently supportive, positive, and always encouraging me to push my limits.

## Abstract

Workplace Dignity has long been the subject of scholarly enquiry, although until recently the body of research has been dominated by ethnographic work. Recently, Thomas and Lucas (2019) developed the first quantitative, direct measure of perceptions of workplace dignity: the Workplace Dignity Scale (WDS). Given the importance of understanding dignity in the workplace, this study sought to replicate the initial scale validation study conducted by Thomas and Lucas, so as to confirm the reliability and validity of the scale prior to its future applied and scholarly use. Moreover, the current study contributes to the ongoing methodological reform of psychology towards a transparent and rigorous science by preregistering the method and analysis script prior to collecting data. A large sample of workers ( $N = 853$ ) from the United States were recruited through Prolific Academic and completed an online questionnaire that included the WDS, as well as theoretically related scales (e.g., workplace incivility). Confirmatory factor analyses indicated that the model specified by Thomas and Lucas had reasonable global fit and estimates of reliability ( $\omega_t$ ) indicated that the two main factors of the scale, Dignity and Indignity, had high internal consistency. Nomological analyses revealed that the Dignity factor of the WDS was significantly correlated in the expected directions with theoretically related variables. Furthermore, the Dignity and Indignity factors of the WDS were found to highly correlate with one another, posing questions as to whether the two factors are qualitatively different phenomena as was argued by Thomas and Lucas. It is concluded that the WDS is a promising tool for measuring workplace dignity although refinement of the proposed measurement model may be necessary.

## Preface

The structure of this thesis differs from that of a traditional thesis in that it is a thesis by publication. At its core is an article-length manuscript that will be submitted for publication to the journal *Collabra*. Thus, this work is shorter than typical for a Master's thesis.

However, I have provided additional material along with the manuscript, most notably, an extended introduction that is presented before the manuscript.

Furthermore, during my enrolment period, the introduction and methods sections of the manuscript were submitted for peer-review where it was intended that the study described in this thesis would be published through a new publication format known as a Registered Report (I describe the outcome of engaging in this route to publication immediately prior to presenting the manuscript and in Appendix A). Within the extended introduction, I provide a description of the Registered Report format and its place within the ongoing methodological reform of scientific psychology. Finally, following the manuscript, I detail a series of reflections on my experience undertaking this novel thesis format.

## Table of Contents

Title page .....	i
Acknowledgements.....	ii
Abstract.....	iii
Preface.....	iv
Table of Contents.....	v
A Validation of the Workplace Dignity Scale .....	1
Background to the Workplace Dignity Scale (WDS).....	4
The Replication Crisis and Rigorous Scientific Psychology.....	17
The Current Study .....	27
Validating the Workplace Dignity Scale: Manuscript.....	29
Purpose of the Study .....	29
Introduction.....	29
What is Workplace Dignity?.....	30
The Workplace Dignity Scale.....	32
Psychometric Research and Reproducibility .....	34
The Current Study.....	36
Method.....	39
Sample Characteristics .....	39
Measures .....	40
Procedure .....	43

Data Analysis .....	43
Results.....	48
Discussion.....	53
Conclusion .....	58
Reflections on Completing a Registered Report-Based Master’s Thesis .....	59
Registered Reports and Psychometric Validation.....	59
Systematic Replications of Psychometric Validation Studies .....	61
Personal Reflections.....	64
References.....	66
Appendix A – Summary of Peer-Review Comments .....	80
Appendix B – Low-Risk Ethics Application .....	82
Appendix C – Participant Information Sheet.....	85

## A Validation of the Workplace Dignity Scale

The concept of human dignity has long been the focus of discussion in a range of disciplines, from its underpinnings in religion and philosophy, to the necessity to legislate its protection. In particular, the experience of dignity in workplace contexts has been the subject of intellectual enquiry over the last 200 years, from the sociological writings of Marx (Healy & Wilkowska, 2017) to Weber (Titunik, 1995) to Durkheim (Hodgkiss, 2013). It is unsurprising that *workplace* dignity has had such prominence, given that dignity has been positioned as the ultimate human value and is closely tied to the right to work as well as the rights of workers (Sison, Ferrero, & Guitián, 2016). This positioning is based on the Christian ideal that each man is created *imago Dei* – in the image of God – and furthermore that the experience of dignity depends on being able to engage in work, to fulfil one’s call for self-giving. As Pope Francis posted on Twitter: “Where there is no work, there is no dignity” (Pontifex, 2015).

Dignity also forms part of the basis of secular claims for the right to *decent* work. The global Decent Work Agenda, which has the objective of achieving decent work for all women and men by 2030, describes decent work as being work that provides dignity (International Labour Organization, 2017). Decent work involves providing people with opportunities to engage in productive work for a fair income, while being afforded security in the workplace, prospects for development, freedom to speak up with concerns, equality of opportunity and treatment, and freedom to participate in organisational decisions that influence their lives. In contrast to decent work, many features of work or the workplace can lead to denials of dignity.

In the field of industrial/organisational (IO) psychology, there is a vast literature on topics related to workplace dignity such as workplace bullying (Verkuil, Atasayi, & Molendijk, 2015), harassment (Nielsen, Glasø, & Einarsen, 2017), prejudice (Jones et al., 2017), incivility (Schilpzand, Pater, & Erez, 2016), and organisational justice (Silva &

Caetano, 2016). In fact, it has been suggested that these topics, and other aspects of the work environment evoke implicit understandings of what workplace dignity is (Bolton, 2007). That is, dignity is usually taken for granted as something that exists and is important, and acts as a guiding standard for making moral judgements (e.g., that workplace incivility is ‘bad’ because it can deny one’s dignity). Moreover, the United Nations declaration on human rights invokes the notion of human dignity in an attempt to prevent actions which violate it, without describing exactly what dignity means (United Nations, 1948). It seems that dignity, and by extension, workplace dignity, is understood and justified by its utility. However, this has led to vague conceptualisations of what workplace dignity is as a phenomenon in isolation.

Workplace dignity has been asserted as being subtle and complex, expressed in a multitude of ways, simultaneously: as respect, as a recognition of humanity, as status, as autonomy, as contribution, as trust, as an essential need of the human spirit (Bolton, 2007; Sayer, 2007). However, the positioning of dignity in such broad terms makes it difficult to study *explicitly* and in a systematic way. The seemingly all-encompassing nature of dignity has been criticised in that it is not only impractical to study, but that at an ontological level, it may be no different than respect (Macklin, 2003). Such discussion is valuable because it helps move workplace dignity from a background level of analysis to the foreground, so that the construct of workplace dignity itself may become the subject of investigation.

When workplace dignity has been studied as the topic of interest, researchers have mainly employed ethnographic methods of enquiry and, thus, created a rich and diverse suite of insights (e.g., from the lived experiences of Indian security guards interfacing with precariousness [Noronha, Chakraborty, & D’Cruz, 2018], to the early experiences of being a police woman on patrol [Martin, 1980]; also see Hodson, [2001]). However, there has been a lack of quantitative work to complement these findings and to further understanding of what workplace dignity is as a construct. Establishing a quantitative measure of workplace dignity

would allow for a more comprehensive understanding not only of what workplace dignity is, but also its antecedents and consequences, such that subsequent insights can be used in applied settings. For instance, a quantitative measure of workplace dignity could form part of a baseline assessment of an organisation's climate, from which the effects of an intervention (such as a diversity training programme) could be measured.

Of the few studies that have quantitatively explored dignity in the workplace, none have included a direct measure of workers' perceptions of workplace dignity. In fact, existing quantitative studies have represented dignity by using combinations of variables derived from Hodson's (2004) Workplace Ethnography Project (however, the variables from this project were created by content-coding over 200 workplace ethnographies with no variables created as direct measures of dignity). For example, using survey items and content-coded variables from the 2004 project, Hodson (2007) grouped three variables under the term 'working with dignity': organisational citizenship behaviours, lack of conflict with management, and lack of co-worker infighting. Regression analyses revealed that management competence positively predicted each of the three dignity variables while the size of the organisation negatively predicted the dignity variables. More recently, Lucas, Manikas, Mattingly, and Crider (2017) grouped other variables together under the titles 'dignity in work' and 'dignity at work', informed by Bolton's (2007) untested multidimensional framework of dignity at work. Of these variables, meaningful work and job satisfaction (dignity in work), as well as employee voice (dignity at work) positively predicted employee engagement. In neither of the above studies, nor in other quantitative investigations of dignity (e.g., Crowley, 2014; Roscigno & Hodson, 2004) was dignity a measured variable itself. Therefore, there is a need for a measure of workplace dignity that is both valid and reliable. Indeed, only recently has there been a scale developed that aims to directly quantify perceptions of workplace dignity (Thomas & Lucas, 2019).

## **Background to the Workplace Dignity Scale (WDS)**

The following sections detail the journey of Lucas, one of the developers of the WDS, to create the scale, documenting her interpretation of the literature and subsequent theory development of what workplace dignity *is* as a construct. These insights provided the background for the development and validation of the WDS (Thomas & Lucas, 2019). I outline Lucas' journey here for two main reasons. Firstly, the workplace dignity literature is vast and primarily ethnographic, and there is not yet a widely accepted definition of workplace dignity; it is thus possible that others could have interpreted the same literature and come to a different understanding of what workplace dignity is (e.g., Bal, 2017). Relatedly, the purpose of the current study is to *replicate* the initial validation study of the WDS. In doing so, the theoretical foundations that guided the development of the WDS, as identified by Lucas, will be tested; that is, this study includes confirmatory analyses that test Lucas' proposed theory of workplace dignity. Thus, I will provide a descriptive overview and commentary of Lucas' interpretation of the literature, the theoretical foundations that she proposed underpin the literature, and the subsequent principles of workplace dignity that guided the construction of the WDS.

**Descriptive overview of Lucas' literature review.** Lucas suggests that ethnographic research into workplace dignity has primarily focussed on critiquing problematic workplaces, exploring workers' responses to dignity threats, and exploring the experiences of vulnerable populations (Lucas, 2017).

***Critiques of problematic workplaces.*** Research into workplace dignity has most often examined the role that problematic workplaces play in violating workers' dignity. There are hundreds of ethnographies covering a range of contexts, each of which provide insights into the lived experiences of *particular* workers within *particular* workplaces. However, it is difficult to make generalisations about the characteristics of problematic workplaces that

influence dignity because ethnographic methods are designed to explore phenomena as experienced by individuals within their lived context (Hodson, 2001). Nonetheless, Hodson (2001), in a qualitative comparative analysis (QCA) of over 200 ethnographies, highlighted several aspects of problematic workplaces that challenge workers' dignity. For instance, dignity can be challenged when employees are mismanaged and abused, overworked and exploited, and have their autonomy invaded. Individual studies are, nonetheless, beneficial in that they highlight specific workplace conditions that can negatively impact dignity, such as when hotel staff are required to show servility to guests in their interactions (Kensbock, Jennings, Bailey, & Patiar, 2014), or when engaging in jobs characterised as 'dirty work': occupations that society generally considers to be physically, emotionally, or morally tainted (Ashforth & Kreiner, 2014; Rivera, 2015).

Across the problematic workplace literature, dignity is often understood in relation to how it has been violated, although suggested meanings of dignity vary between ethnographies. Indeed, because ethnographic research typically focusses on how the particular individuals under study come to derive meaning in the context of their lived experience (each participant is the "expert"), multiple conceptualisations of dignity are to be expected across ethnographies. However, the lack of an established definition of workplace dignity makes it difficult to discover patterns across studies, which is important for developing a generalisable measure of workplace dignity. Furthermore, the research on problematic workplaces does not position dignity as the primary phenomenon-of-interest, but instead uses "dignity as a standard for passing judgement on ethically questionable or otherwise harmful work practices" (Lucas, 2017, p. 7). Therefore, while the body of ethnographic work offers rich idiographic insights about the lived experiences of (violations of) workplace dignity, further understanding of workplace dignity as a construct (and its antecedents and consequences) will be stilted until it becomes the focus of research itself.

***Responses to dignity threats.*** Another focus of the workplace dignity literature has been on how workers respond to threats to their dignity. Lucas (2017) suggests that two separate strategies for responding to threats to dignity have been emphasised within the literature: one that involves engaging in identity work and one that involves resisting dignity threats.

*Identity work.* Occupation typically forms part of an individual's self-identity, given that it is a major determinant of status and prestige. Threats to dignity can occur for those in occupations that might be considered low in status and prestige, although workers often respond to these threats by reframing their work in an attempt to establish a positive professional identity for themselves (Baker, 2014). Engaging in identity work is one way that those involved in dirty work might curtail dignity threats associated with the stigma of their work or workplace (Ashforth & Kreiner, 1999). For example, Chiappetta-Swanson (2005) found that nurses who manage the abortion procedure for pregnancies where there is a high risk of medical abnormality reframed what could be seen by some in society as a morally tainted job (and thus dirty work) to focus on the meaningful aspects of the job. Specifically, they emphasised the opportunity to exercise a high level of care for patients, with some suggesting that it was, in fact, a privilege. In this way, workers responded to dignity threats associated with stigmatised work by reframing the work itself, and thus, formed a positive professional self-identity. Similarly, garbage collectors' discourses suggested that they saw their *physically* dirty work as heroic and paternalistic, which was dignity affirming (Hamilton, Redman, & McMurray, 2019). However, while many researchers evoke the concept of identity work, it has been criticised in that there has been a lack of systematic exploration into exactly what identity work is (A. Brown, 2015); thus, its contribution to workplace dignity may yet be challenged.

*Resistance.* Workers can also *resist* dignity threats, with research identifying both collective and individualised forms of resistance. An example of collective resistance occurred when exploitative termination practices led vulnerable migrant poultry workers to organise a union-led campaign aimed at generating public outcry (Stuesse, 2010). Further, the designated theme for this campaign was one of ‘justice and dignity’. A similar form of collective resistance might come in the form of striking (Rosigno & Hodson, 2004). Individualised forms of resistance refer to actions that individuals take to resist dignity threats. This resistance might come in the form of absenteeism, sabotage, gossiping, withholding effort, resigning, or other resistance behaviour (Hodson, 2001).

Resistance efforts have also been characterised as being subtle, creative, or productive in nature (Lucas, 2015). An example of subtle resistance occurred when workers displayed cynicism in response to an offensive paternalistic management culture (Fleming, 2005). Cynicism helped employees retain a sense of self that emphasised adulthood strength and maturity in response to dignity threats that arose from being treated as if they were irrational children. Narratives of organisational misbehaviour often contain examples of creative resistance efforts, such as when employees purposely wore ugly ties to oppose the company dress code (Karlsson, 2011). Finally, as an example of a productive resistance effort, Baker (2014) found that LGBTQ employees experiencing dignity threats associated with their sexual orientation or gender expression advocated for ‘safe spaces’ for similarly threatened employees in other workplaces, as one means of restoring their own dignity.

*Experiences of vulnerable populations.* Much of the workplace dignity literature has also focussed on the experiences of vulnerable populations. Belief in an *earned* component of dignity suggests that one’s perceived value is related to the contribution that they make in the workplace (Lucas, 2015). That is, the greater the instrumental contribution that one makes, the greater the dignity they are afforded. However, for those on the bottom rungs of the

organisational ladder, which are often overrepresented by those from vulnerable groups in society, it can be more difficult to achieve dignity based on competence and contribution. According to Lucas (2017), this is partially because people from vulnerable groups tend to occupy low-skilled jobs, which may provide comparatively little opportunity to contribute to the organisation than high-skilled jobs. This lack of opportunity to achieve workplace dignity may be exacerbated by an absence of receiving messages from salient others in the organisation that acknowledge the contribution that workers *do* make (Lucas, 2017).

Inequality in opportunity to earn dignity might be considered a type of *identity-indifferent inequality*: one of two types of inequality that can be a basis for dignity injury. According to Sayer (2007), identity-indifferent inequalities are organisational inequalities that cause dignity threats and that are not relevant to the job holder's personal identity. Indeed, economic resources, such as pay and benefits, are unequally distributed in the workplace, and while these economic inequalities do not directly discriminate based on workers' identities, they disproportionately affect vulnerable populations (Lucas, 2017). Where economic remuneration is seen as a signal of worth, this can lead to dignity violations for those vulnerable populations at the lower end of the economic inequality. In contrast, *identity-sensitive inequality* refers to a basis of dignity threat that arises from an inequality in treatment by organisational others that is based on a person's identity (e.g., their ethnicity, religion, or gender), which again, typically affects those in vulnerable or marginalised groups (Sayer, 2007). Harassment, discrimination and other forms of negative treatment directed at workers based on their identity (e.g., gay slurs directed at LGBTQ employees; Baker, 2014) can lead to dignity violations. In sum, much of the workplace dignity research has focussed on vulnerable populations, because dignity is often experienced in relation to equality, and vulnerable groups are often at the negative end of workplace inequalities.

**Theoretical foundations of workplace dignity.** Based on her understanding of the workplace dignity literature, Lucas (2015) suggested that there are four theoretical foundations from which the research on workplace dignity is built upon. However, she emphasises that these foundations by no means constitute a comprehensive theory of workplace dignity, but rather that they should act as a starting point for future theorising.

**1. Normative expectations.** Firstly, Lucas (2015) suggests that one of the core assumptions underlying the workplace dignity literature is that people expect that human dignity should always be upheld unconditionally, and that this normative expectation necessarily extends to the workplace. Thus, violations of dignity that occur in the workplace are viewed as problematic, which may partially explain why a substantial body of research has been dedicated to exploring the experience of workers who face dignity violations (Hodson, 2001). This normative expectation was evident, for instance, among care workers facing the impending closure of their workplace who expressed the desire to be treated with dignity as being more important than the job loss itself: “the absence of dignity served as their definition of the situation” (Gunn, 2011, p. 73). However, one of the unique characteristics of workplace dignity is that normative expectations of dignity are sometimes at odds with the structure of the employment relationship.

**2. Incongruence between the employment relationship and dignity.** While entering into an employment relationship can be congruent with dignity pursuits – employment can promote flourishing (e.g., through meaningful work), and it can be a source of identity and self-esteem – at its core, the employment relationship is a contract based on an economic exchange between employee and employer that is usually characterised by unequal power relations (Bélanger & Edwards, 2013; Bolton & Laaser, 2013; Sayer, 2007). In particular, the nature of the economic exchange relationship, being unequal and instrumental, provides a context in which workers are susceptible to dignity violations. This economic exchange basis

of the employment relationship means that employees' worth (which is signalled by pay) is primarily determined by the instrumental contribution they make towards achieving organisational goals, rather than their inherent, unconditional value as human beings (Sayer, 2007).

Similarly, the economic exchange relationship is often incongruent with dignity pursuits because the workplace is characterised by inequalities. Inequalities in power within work relationships (e.g., between managers and subordinates), pay, flexible work availability, type of contract (e.g., fixed-term vs. permanent), benefits (e.g., favourable sick leave policy being offered to permanent staff only), and other aspects of the workplace can impinge upon workers' dignity because they may be treated with a lack of respect and have their individual agency and autonomy denied (Newman, 2000; Sayer, 2007, 2011). However, organisational practices and interactions between workers that *conceal* the unequal nature of the workplace have been shown to remediate dignity injuries (whereas practices that further *reveal* inequalities "add insult to injury") (Lucas, 2015, p. 634). The distribution of civility to members of an organisation, for instance, can either conceal or reveal inequalities.

**3. Subjective nature of dignity.** A third theoretical foundation of workplace dignity is that it is subjectively experienced. While the economic exchange relationship suggests that people's dignity may be threatened partially because their workplace involves inequalities and emphasises their instrumental value, it is not always the case that workers will experience dignity violations. There are a multitude of factors likely to influence whether one's dignity is affected in the workplace. For example, despite inequality in power between manager and subordinate, their relationship may be characterised by respectful interaction (or at least the subordinate may perceive it to be), which may offset potential power-related dignity threats to the subordinate. Therefore, ultimately it is the circumstances surrounding the individual

worker, and specifically, the individual's *perception* that shapes the experience of dignity (Lee, 2008; Thomas & Lucas, 2019).

**4. *Inherent and earned dignity.*** Lucas (2015) suggests that there are at least two separate meanings of workplace dignity, each of which offer a route for dignity pursuits to be attained or obstructed.

*Inherent dignity.* *Inherent dignity*, which is often described as human dignity, is the belief that without exception, all people have an equal level of unconditional and absolute worth simply because they are human (Brennan & Lo, 2007). In the workplace, inherent dignity is upheld through respectful communication from other workers. This is consistent with Sayer's (2007) assertion that human dignity relies on 'words and deeds' that signal that one's importance is beyond the value of their labour and that they deserve respect because they are human. Furthermore, this expectation is consistently conveyed by workers themselves. In focus groups, workers from a range of occupations frequently cited instances of disrespectful interaction, and near zero instances of respectful interaction, when asked to recall concrete examples of when they had experienced dignity or indignity at work (Lucas, 2015). Lucas interpreted this pattern of responses as suggesting that, at a minimum, people *expect* to be treated as human through respectful interaction; thus, instances to the contrary are more salient and memorable. Moreover, dignity violations (again, which are thought to be subjectively determined) were experienced with greater intensity when disrespectful interactions were thought to involve hostile intent.

This finding suggests that inherent dignity is undermined through disrespectful interaction in the workplace and is consistent with other studies that have positioned disrespectful communications as violators of human dignity. For example, immigrant workers at a meatpacking organisation were made to feel worthless when their employer expressed that they were worth no more than disposable plates or cups that get thrown in the garbage

(Apostolidis, 2005). Here, the inherent worth of workers was reduced to that of non-human objects via explicit communication. This example reflects a common finding in the literature of a particular type of disrespectful interaction that undermines inherent dignity: that which is dehumanising or objectifying. Other examples include: American Football players being referred to by alphanumeric codes that made them feel that they were “a piece of furniture”, or part of a “slave market” during drafting evaluations (Dufur & Feinberg, 2007, p. 517); models being subject to humiliating comments about their bodies, creating unrelenting pressure to maintain bodily capital such that they feel like “paper dolls” (Mears & Finlay, 2005, p. 318); and garbage workers regularly facing abuse from members of the public, positioning them as “scum” and “lower than a snake’s belly” (Hamilton et al., 2019, p. 894).

*Earned dignity.* Lucas (2015) theorises that a second component of workplace dignity is *earned dignity*. Whereas all people have inherent dignity that is unconditional and equal because they are human, earned dignity, in contrast, is conditional and unequal because it is afforded on the basis of an individual’s efforts and abilities in the workplace (Hodson, 2001). Similarly, whereas inherent dignity is affirmed or violated through communication (of respect or disrespect), earned dignity is affirmed or violated through messages relating to competence and contribution. For example, praise of one’s work can be dignity affirming, whereas highlighting mistakes and insulting one about their job performance can be dignity violating (Lucas, 2015). Moreover, while doing a job well can be somewhat intrinsically rewarding, it is (according to Lucas) essential that messages from others that acknowledge the quality of the worker’s contribution are received to generate dignity. Evidence from focus groups indicates that a lack of acknowledgement for one’s contribution can lead to a denial of dignity (Lucas, 2015). Even token messages of competence and contribution can be dignity affirming. For example, one participant, a fast-food cook, had his dignity affirmed when he was given a free pizza by his manager after completing a busy shift on his own.

In addition to expressions of work quality, Lucas suggests that workers need to be provided with *opportunities* to demonstrate their competence in the workplace to achieve dignity. Earned dignity operates on the principle that greater contributions in the workplace lead to greater dignity. Not only do workers believe in this principle, they are also aware of how much instrumental value they provide, and can provide, to their organisation (which depends on the opportunities offered by their particular job). Indeed, many people not only have a desire to demonstrate their competence, but also to increase it (Paloniemi, 2006). Opportunities to do so are critical for upholding dignity; a lack of these opportunities, which is often characteristic of jobs held by those in vulnerable groups, can lead to a denial of earned dignity (Lucas, 2015).

*Inherent and earned dignity interact in lived experience.* On the surface, the notion of workplace dignity being composed of these two sources of dignity is contradictory and illogical, because the idea of dignity as being equal and unconditional (inherent dignity) contradicts the idea of dignity being unequal and conditional (earned dignity). In other words, “positioning some people as worthy of greater dignity necessarily means that all people cannot have equal dignity” (Lucas, 2017, p. 3). However, Lucas posits that people may simultaneously express a belief in inherent dignity *and* a belief in earned dignity, despite the two components seeming contradictory in theory (also see Pirson, 2017). In practice, in some cases, the expectation of earned dignity may be in tension with the expectation of inherent dignity. Although participants’ accounts in Lucas (2015) implied an understanding that people who made greater instrumental contributions in the workplace deserved greater dignity, their experiences also indicated a resistance to the corollary relationship that those who made smaller contributions deserved less dignity. In these instances, dignity threats occurred when workers observed that they were valued comparatively less than other employees. Inequality in contribution, which led to inequality in (earned) dignity, seemed to

run counter to people's expectations of equality of (inherent) dignity. In this way, equality expectations based on inherent dignity may have been transferred over into earned dignity situations. Again, this threat might seem at odds with the belief in earned dignity, but it illustrates how inherent and earned dignity are intertwined in lived experience (Lucas, 2015).

Lucas (2015) also found that dignity affirmations were usually described in relation to events involving evaluations of contribution and competence, rather than events where workers' unconditional value was communicated. Thus, while inherent dignity has traditionally been positioned as the ultimate human value (Sison et al., 2016), for workplace dignity, inherent dignity may be subjugated by other components of dignity, at least in certain situations. Lucas suggests that earned dignity may, thus, have a more prominent role in workplace dignity than conventional thinking might dictate. However, I argue that people may simply be more likely to remember dignity affirmations related to the earned component, because they are more salient and memorable than inherent dignity affirmations, which are less noticeable because they are expected. Nonetheless, in sum, a key insight of Lucas' work is that earned and inherent dignity interact in lived experience; a finding which extends the theory that workplace dignity simply involves those two components. Moreover, the assertion that workplace dignity involves a component that is specifically derived from engagement in workplace activities, as well as an inherent dignity component (the two of which seem to interact), suggests that *workplace* dignity should be distinguished theoretically from *human* dignity (i.e., just inherent dignity) operating in a workplace context.

**The Workplace Dignity Scale (Thomas & Lucas, 2019).** Based on Lucas' review of the literature (Lucas, 2015, 2017) and subsequent proposed theoretical foundations, Thomas and Lucas (2019) sought to develop a quantitative measure of people's perceptions of workplace dignity (for each of the scale items see Thomas and Lucas, 2019). Here, I will outline how Lucas' theorising guided the development of the Workplace Dignity Scale,

although I will not be describing the methods used for the scale's development because these are described in the manuscript for journal submission, which involves a replication of the initial validation of the WDS (see section "Validating the Workplace Dignity Scale: Manuscript"). The focus on theory here is important because scale development should be theory-driven, and theory should guide decision making throughout the scale validation process (Kline, 2015).

The WDS was developed based on Lucas' (2017) definition of workplace dignity: "the self-recognised and other-recognised worth acquired from (or injured by) engaging in work activity" (p. 2549), as well as four principles of workplace dignity identified following her review of the literature (note that the four principles are *not* the same as the four theoretical foundations discussed earlier). The first principle is that dignity is communicatively bound, and so many of the scale items ask about interactions with other people (e.g., "People at work treat me like a second-class citizen"). Secondly, while communication is a necessary precondition for the experience of dignity, dignity is experienced *subjectively*. Thus, the WDS was designed as a self-report measure (and hence taps into perceptions of workplace dignity). Thirdly, workplace dignity has both *inherent* and *earned* components, which are reflected in the item wording. For example, "At work, people talk to me like an equal, even if there are status differences between us" alludes to the inherent component, specifically referencing the importance of equality, and "People at work communicate with me respectfully" acknowledges the role of respectful interaction in upholding inherent dignity. On the other hand, items such as "People at work show they appreciate my work efforts" allude to the earned component, specifically referencing the importance of messages of contribution, and items such as "People at work recognize my competence" and "At work, I have the chance to build my competence" reflect the importance of competence being recognised and the opportunities to demonstrate

competence, respectively. Finally, the fourth principle guiding scale development was that workplace dignity is bivalent in nature. Thomas and Lucas (2019) suggest that workplace dignity operates in a dual-continuum manner (where dignity and indignity fall on separate continuums). It “appears to be analogous to Herzberg’s (1986/2003) model of hygiene and motivating factors in that it has both positive elements that promote dignity and negative elements from which dignity must be protected” (Thomas & Lucas, 2019, p. 77). To this end, the WDS includes positively and negatively valenced items, with the negative items later treated as an *Indignity* factor that was separate from *Dignity* in a confirmatory factor analysis (CFA).

However, the negatively valenced items refer to both negative experiences of *dignity* and experiences of *indignity* and I argue that there is ambiguity surrounding the conceptual definitions of dignity versus indignity. Lucas (2017) implied that *indignity* was the experience of the absence of dignity, though Thomas and Lucas (2019) stated that “the opposite of low indignity does not necessarily equate to high dignity and vice versa (c.f. Herzberg, 1968/2003)” (p. 102). In fact, while the first three principles underlying the WDS development clearly draw on Lucas’ background review of the literature, Lucas does not provide sufficient reasoning nor enough conceptual clarity to regard dignity and indignity as being separate continuums rather than just two poles of the same dimension. Nonetheless, there was a need for the development of a quantitative measure of workplace dignity, and it is the goal of the present study to replicate Thomas and Lucas’ (2019) validation work. More broadly, it is critical that studies – especially those involving scale assessment – are replicated, given the current replication crisis in psychology.

## **The Replication Crisis and Rigorous Scientific Psychology**

In 2011, the *Journal of Personality and Social Psychology* published a study in which the author claimed to have found evidence for the existence of precognition in humans (i.e., the ability to ‘see’ into the future; future events affecting people’s thinking and behaviour in the past) (Bem, 2011). These implausible claims attracted extensive attention (Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011), not only because the suggestion that mental time travel was conceivable was published in a peer-reviewed journal, but also because the findings were produced using conventional methods in psychology. Moreover, 2011 was especially significant because a respected social psychologist, Professor Stapel, was found to have fabricated data used in at least 30 research publications and manipulated data in other publications as far back as 2004 (Levelt Committee, 2011).

Following these events and failed attempts to replicate Bem’s work (Ritchie, Wiseman, & French, 2012), psychology researchers began to question other published findings in psychology. Attempts to replicate well-known findings revealed that a worryingly large number of findings in the field were irreproducible. In a wide-scale reproducibility test, researchers from the Open Science Collaboration (2015) found that while 97% of 100 original studies to be replicated displayed statistically significant results, only 36% of the replications displayed significant results. Thus, the field of psychology is now considered to be undergoing a ‘replication crisis’, and while the cause of the crisis is debated, it is commonly thought that many published findings are the result of methodological issues, not true effects (John, Loewenstein, & Prelec, 2012; Shrout & Rodgers, 2018).

Replication is a foundation of a healthy science. If the world operates according to stable laws, as is assumed by the scientific method, then if a truth is discovered it should therefore be reproducible (Radder, 1996). Furthermore, because research findings are probabilistic – that is, they rely on statistical methods that establish not certainties, but

likelihoods of particular results – discoveries or truths can be claimed only after an accumulation of evidence rather than from the result of any single study (Chambers, 2017). Thus, the act of replicating prior research allows for theories to be refined and made precise in order to advance knowledge. Indeed, Schmidt (2009) suggests that within the social sciences, replication studies have five main functions: controlling for sampling error, experimental artefacts, and fraud, and enabling generalisability and verification of hypotheses.

In principle, replications are fundamental to scientific psychology. However, within the field, there is an attitude of indifference towards replications (Chambers, 2017). In fact, Makel, Plucker, and Hegarty (2012) found that as of 2012, only 1% of the published psychology literature involved replication attempts. This may stem from a publication bias within academic journals whereby research that is original and that includes positive findings (studies that show a significant ‘effect’) is prioritised over replication attempts and negative findings, which are equally important if the research is methodologically sound (Francis, 2012; Nosek et al., 2015). In fact, Ritchie et al. (2012) initially sent their manuscript detailing three unsuccessful attempts to replicate Bem's (2011) findings to the same journal that the original work was published in, and were immediately rejected without peer-review because it was in the journal’s policy not to publish *any* replication studies (Chambers, 2017). Furthermore, because publications are key to an academic career (so that academics ‘publish-or-perish’), this common assessment criteria for publication (i.e., manuscripts that detail positive and novel findings) encourages questionable research practices (QRPs) (John et al., 2012). In turn, this limits the field’s ability to generate replicable evidence, which is central to scientific discovery.

QRPs can be thought of as research practices that lead authors to publish misleading evidence about the size and replicability of effects. One such QRP involves submitting results

with low statistical power (i.e., without having conducted a power analysis for required sample size). Not only does low power reduce the probability of detecting an effect when there is an effect to be detected (i.e., a true-positive), it also inflates the probability that a result is due to chance rather than a true effect (i.e., a false-positive). When a study detects a positive result (i.e., the null hypothesis is rejected), that result can either reflect a true- or a false-positive. The true-positive rate can be calculated by multiplying power by the prior probability that the effect being studied is true. The false-positive rate ( $\alpha$ ) is typically set at .05, the threshold for deeming an effect as 'significant'. The positive predictive value (PPV), then, is the proportion of true positives to *all* positives (true-positive rate plus  $\alpha$ ), indicating the probability that a given positive result reflects a true-positive instead of a false-positive. Thus, for a given prior probability, as power decreases, the PPV decreases, and, hence, the probability of the result being a false-positive increases (Chambers, 2017).

Another QRP involves collecting data until a significant result is found, or deciding whether more data should be collected by constantly stopping and testing for significance (John et al., 2012). This is referred to as optional stopping and is problematic because it can result in a false positive rate of 100%; even when the null hypothesis is true, by repeatedly collecting observations and running significance tests after each new observation, one will *always* eventually find a statistically significant result that rejects the null hypothesis (Kruschke & Liddell, 2018). More generally, QRPs tend to come about when researcher flexibility in data collection and analysis is exploited and decisions are made because they produce favourable outcomes. In fact, researcher flexibility can lead to impossible results being presented as significant (i.e., statistics from experiments that do not match what is known about space and time). For instance, Simmons, Nelson, and Simonsohn (2011) demonstrated how researcher flexibility could lead to a finding indicating that people's chronological age decreased following listening to music. This flexibility, in turn, is afforded

by a lack of firm guidelines spanning the many steps involved in research, from data collection to analysis.

The data analysis procedure has been likened to a “garden of forking paths” with each decision the researcher makes based on the collected data leading them down a different path, potentially to a different result (Gelman & Loken, 2013, p.1). Even seemingly innocuous and reasonable decisions can lead to different findings, given the same data set (Steen, Tuerlinckx, Gelman, & Vanpaemel, 2016). Thus, it is vital that data analysis is *not* contingent on the data. When researcher degrees of freedom are exploited, such that data analysis options are selected based on the data (i.e., decisions are made after exploring *p*-values in the data, with the aim of ending up with significant results), this is referred to as *p*-hacking (Chambers, 2017; Head, Holman, Lanfear, Kahn, & Jennions, 2015). Another example of this is when the researcher excludes reasonable data points under the guise of being outliers so that an effect becomes significant and more likely to be published. A similar QRP to *p*-hacking is known as Hypothesising After the Results are Known (HARKing) (Kerr, 1998). HARKing involves first exploring a data set for significant effects and then generating hypotheses *post-hoc* about those effects but reporting the hypotheses as if they were *a priori*.

Both *p*-hacking and HARKing are problematic because they increase the likelihood that published findings reflect Type I errors: illusory effects that are taken to be genuine. In the case of HARKing, this is because the researcher tests many possible relationships, and in doing so, increases the probability that a significant relationship will emerge due to chance (even if there are no true relationships). Hypotheses are then made about these significant relationships, which are obviously supported by the data, to form the basis for publication (Nosek, Ebersole, DeHaven, & Mellor, 2018). These QRPs have widespread ramifications in the instances where illusory effects inform further theory development (which informs future research, and so on). Replication can be a solution because successful replications indicate

that the original finding was unlikely to be the result of Type I error (Kerr, 1998; Shrout & Rodgers, 2018).

More generally, a lack of transparency in data collection and analysis methods makes it difficult for researchers to evaluate and replicate others' work. Published articles frequently lack the necessary detail for researchers to replicate each of the methodological steps (e.g., transformation of variables, outlier removal, etc.) and such a lack of transparency makes it difficult to evaluate whether the researcher might have engaged in QRPs. Thus, the reader is typically unable to identify whether a given effect is 'real'. On a wide scale, this clearly harms the credibility of scientific psychology as a discipline (Chambers, 2017; Simmons et al., 2011).

**Partial solutions to the replication crisis.** One proposed solution to the replication crisis is for researchers to utilise *preregistration* when conducting confirmatory (hypothesis-testing) research (Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012). Preregistration involves researchers describing their hypotheses and data analysis plans publicly *before* collecting data (Chambers, 2017; Nosek et al., 2018). The prespecified plan is typically uploaded online, and time stamped in a depository where the original version cannot be edited. Henceforth, the researcher is to proceed with their research in line with their prespecified plan. The point of preregistration, thus, is to limit researcher flexibility and engagement in at least some QRPs. Preregistration is one means of increasing transparency and reproducibility in psychological science.

Unfortunately, preregistration is only a partial solution to the replication crisis because the prespecified plans themselves do not undergo peer-review, which not only makes it possible that the preregistration neglects important details or contains inconsistencies, but also leaves open opportunities for the researcher to engage in QRPs (Chambers, 2017). For example, vague preregistrations still afford opportunities for undisclosed researcher degrees

of freedom. Peer-reviewed preregistration, on the other hand, can more definitively reduce researcher flexibility because the preregistration is evaluated by other academics who critique the plan and provide feedback. This publishing format is known as a *Registered Report* and includes two review stages (Chambers, Feredoes, Muthukumaraswamy, & Etchells, 2014; Nosek & Lakens, 2014). The *Stage 1 Submission* involves authors submitting an introduction and a proposed methods and analysis plan to a journal for peer-review before collecting data. If this passes in-depth peer-review, the manuscript is granted *in principle acceptance* (IPA) (and only once IPA is granted may data collection begin). This means that the final report is guaranteed to be published regardless of the results, if the protocols are satisfactorily followed. The assessment of whether the protocols were followed forms part of the *Stage 2 Submission* in which the entirety of the introduction, methods, results, and discussion are submitted for peer-review. Importantly, null results (which are equally as informative as positive findings) are published and meaningful replication attempts are encouraged. Thus, the Registered Report format – whereby studies are evaluated and accepted for publication based on their methodological rigor, as opposed to their results – improves the transparency and reproducibility of scientific psychology and is one step towards resolving the replication crisis. As of December 2019, 217 journals from a wide range of disciplines offer the Registered Report publishing format (see <https://cos.io/rr/>).

**Construct validation.** Another important aspect of a healthy science involves ensuring that measures of psychological phenomena are valid; that is, that they measure what they intend to measure. However, the role of questionable *measurement* practices (QMPs) in the replication crisis has received relatively less attention than QRPs (Flake & Fried, 2019). Many of the constructs of interest in psychological research are *latent*, in that they are not directly observable (e.g., happiness, self-esteem, dignity). These are unlike variables that are observable and can thus be measured directly, such as height or weight. In order to test

psychological theories that relate theoretical constructs to one another it is first necessary to ensure that we have valid and reliable measures of these constructs (Hussey & Hughes, 2019). To test whether there is a relationship between perceived inequality and workplace dignity, for instance, it is important that the tools used to measure these variables are actually doing so (and not measuring some other construct, or nothing at all). Failing to use valid measures presents a risk to the fidelity of subsequent findings and the conclusions drawn from those findings, because a score on a given measure is taken for granted to represent the associated construct during data analysis and interpretation (Hussey & Hughes, 2019). Invalid measures, then, can lead to invalid results (Flake, Pek, & Hehman, 2017). Thus, it is only after thorough assessment of the validity of a measure that it should be used (although I note later that validity assessment should be ongoing).

The assessment of the validity of a given measure of a latent variable is referred to as *construct validation*, which has been defined as “the process of integrating evidence to support the meaning of a number assumed to represent a psychological construct” (Flake et al., 2017, p. 371). One conceptualisation of construct validity pertains to the specific *use* of a measure (or scale) in that it can be population- or context-dependent; in other words, a scale may be appropriate for use and interpretation only in certain contexts (Cronbach & Meehl, 1955; Messick, 1989). For example, the WDS might only have construct validity among participants who live in the USA, as the questions within the WDS might mean something different, or have no meaning, for those living in, say, China. However, Borsboom, Mellenbergh, and van Heerden (2004) provide a different theory of construct validity in which a test can be said to be a valid measure of a psychological attribute if the attribute exists, and variation in the attribute causes variation in test scores. In either case, it is important for validity evidence to be continually gathered (as with replication of significant effects), to the extent that any time a scale is used its validity should be tested (Hussey &

Hughes, 2019). However, it is not common practice for this to occur, with many scales created ‘on-the-fly’ or modified without validation (Flake et al., 2017).

Construct validation is broken down into three sequential phases: substantive, structural, and external validation (Cronbach & Meehl, 1955). The substantive phase refers to the process of conceptualising a construct based on a literature review and the development of scale items with representative content. The structural phase refers to the process of developing theory around the structure of a scale, which involves analysing the scale’s psychometric properties (e.g., via item and factor analysis). Finally, the external phase typically involves assessing whether the scale is associated with or predicts other similar scales (*convergent validity*) and not associated or associated to a lesser degree with scales known to represent unrelated constructs (*discriminant validity*) (Loevinger, 1957).

Much of the theoretical work in psychology attends to the first phase of construct validation, substantive validation, whereas much of the empirical work attends to the third phase, external validation. However, the structural phase receives relatively less attention despite being of equal importance. Sampling 30% of published papers in the *Journal of Personality and Social Psychology* from 2014, Flake et al. (2017) noted that the field is far from being engaged in best practice in relation to structural validation: few studies reported ongoing assessment of the psychometric properties of their scales, 19% of the studies sampled reported modifying a scale but did not provide new validity evidence, and 46% of measures of latent variables seemed to be created on-the-fly, with no reference to validity evidence. Furthermore, Cronbach’s  $\alpha$ , which is commonly misunderstood and used incorrectly (see Schmitt, 1996; Sijtsma, 2009), was usually reported as the only structural validity metric (even though it is intended as a measure of internal consistency *reliability*)<sup>1</sup>.

---

<sup>1</sup>Two alternative estimates of reliability, McDonald’s (1999) Omega total ( $\omega_t$ ) and Omega hierarchical ( $\omega_h$ ), have been proposed as being more appropriate than  $\alpha$  (Dunn, Baguley, & Brunnsden, 2014).  $\omega_t$  measures the total

Indeed, claims about external validity may not hold where a construct has not been shown to have a strong theoretical (substantive) and psychometric (structural) basis. Furthermore, the use of inappropriately validated scales reduces the fidelity of findings involving those scales, and conceivably contributes to the replication crisis. Thus, it is important that ongoing validity evidence is collected, and that it does not neglect any of the three phases of construct validity.

Following Flake et al.'s (2017) recommendations for best practice during construct validation, Hussey and Hughes (2019) sought to assess the psychometric properties of 15 popular self-report questionnaires in personality and social psychology (composed of 26 subscales in total) using a sample of over 80,000 participants. Specifically, they explored the extent to which the scales' validity would survive a thorough suite of validity assessments. Despite 89% of scales demonstrating good internal consistency (meeting  $\omega_t \geq 0.7$ ) and 100% demonstrating good test-retest reliability (meeting  $r \geq 0.7$  for both a 1-hour test-retest correlation and for a correlation with a longer follow-up, varying between 1 day and 1 year), only 73% demonstrated good factor structure (meeting global fit criteria for all of  $CFI \geq .95$ ,  $TLI \geq .95$ ,  $RMSEA \leq .06$ , and  $SRMR \leq .09$ ), and only 4% showed good measurement invariance (meeting configural invariance criteria of  $SRMR \leq .09$  as well as at least one of the other global fit criteria used to assess factor structure, and meeting metric and scalar invariance criteria of  $\Delta CFI \geq -.15$  and  $\Delta RMSEA \leq .01$ ). Overall, this suggests that only 4% of scales could be thought of as demonstrating good structural validity (i.e., only 4% 'passed' all four metrics). Furthermore, the less often a validity metric was reported in the literature for a given scale, the less likely it was that the scale would pass the authors' assessment on that metric, whereas scales with metrics that were more frequently reported were more likely to

---

reliability of a scale by calculating the proportion of variance that is not attributed to measurement error, whereas  $\omega_h$  measures factor saturation by calculating variance that is attributed to a scale's primary factor.

pass. This suggests that many of the measures used in the psychological literature demonstrate ‘hidden invalidity’ (Hussey & Hughes, 2019).

Both Flake et al. (2017) and Hussey and Hughes (2019) have, thus, called for changes to the current practice of under-reporting thorough validity assessment. Indeed, there is a need for evidence of comprehensive validation work to increase confidence that a given scale truly measures what it is purported to measure. However, Hussey and Hughes highlight two key obstacles to achieving valid measures. Firstly, those that administer psychometric scales are usually motivated to conclude that their scales are valid. In an academic setting, demonstrating validity is a prerequisite to testing one’s hypotheses so that having valid measures increases the chance of publication. In an applied setting, the development and application of valid measures can lead to financial reward, in consulting arrangements, for instance. These motivations might lead to a resistance to change the current validation practice. Secondly, researchers can exploit numerous degrees of freedom during reliability and validity assessments – for instance, by selectively implementing reliability metrics (e.g.,  $\alpha$  versus  $\omega_i$ ) and reporting only model fit statistics that indicate favourable model fit – to demonstrate the ostensive validity of a scale. Analogous to *p*-hacking, Hussey and Hughes label this behaviour ‘*v*-hacking’.

Hussey and Hughes (2019) suggest three practical solutions to resolve the current inadequate practice with validity assessment: (a) journal publication should require thorough evidence of psychometric assessment; (b) choices for measures and their validity assessment should be preregistered, and; (c) researchers should make their raw data available so that multiple structural validity assessments can be conducted by other researchers. More broadly, Flake and Fried (2019) have developed a set of questions to help researchers avoid QMPs in their own work, and for others to evaluate whether a study might have used QMPs. These questions specifically address the substantive and structural validation stages of the construct

validation process. Adopting the solutions above will improve measurement transparency and rigor in the field of scientific psychology and, hence, move it another step forward in addressing the replication crisis.

### **The Current Study**

Workplace dignity is a phenomenon of growing interest, though it has received very little quantitative enquiry, particularly into what it is as a construct. Only recently has a direct measure of perceptions of workplace dignity (the WDS) been created and subjected to tests of validity and reliability. However, in the field of psychology there has been a lack of *ongoing* assessment (and sometimes any assessment) of the validity of scales that purportedly measure psychological constructs, and with the additional influence of QMPs when scales have been tested, this may have played a role in producing a replication crisis. Specifically, there is a need for preregistered replications of validation studies on psychometric measures. Thus, the purpose of the current study is to further probe workplace dignity by conducting a replication of Thomas and Lucas' (2019) initial validation study of the WDS (it is important to note that I am *not* suggesting that there were QMPs during the original validation of the WDS by Thomas and Lucas). In doing so, I will be following some of the guidelines for open and rigorous scientific psychology aforementioned by preregistering my method protocols and analysis plans prior to data collection, and sharing the de-identified raw dataset on a public-access online depository such that other researchers can reproduce my analyses and conduct further analyses of their own.

**A note on the following manuscript.** In the following section, I present an empirical article detailing the study conducted for this thesis. Initially, I submitted a different version of the following introduction and method protocols as a Stage 1 Submission Registered Report to the journal *Collabra*, which received peer-reviews and required a resubmit with major revisions. However, given that the first round of peer review took several months (and the

fact that a second round of reviews was required prior to consideration for In Principle Acceptance) the decision was made to abandon the Registered Report format and proceed with data collection, with the aim of submitting the preregistered study to *Collabra* using the traditional publication route after submission of this thesis (i.e., as one completed document including introduction through to discussion). Moreover, within the following manuscript, the reader can find the hypotheses and rationale for the current study and can expect a small amount of repetition of the content previously covered in the thesis introduction.

The following manuscript incorporates changes to the introduction and methods sections based on reviewers' comments on the original Stage 1 Submission and includes the results and a discussion of the findings. For a summary of the reviewers' comments on the original submission and the changes I made in response, see Appendix A. Moreover, the method section differs very slightly to the method protocols that were preregistered *only* in that it is re-written in past tense and now includes sample characteristics (i.e., information that was only available after data collection), whereas the preregistered version was written in future tense because data collection had not yet begun. Preregistered along with the full method protocols were the rationale and hypotheses of the current study, as well as the code for data analysis. This preregistration may be found at <https://osf.io/svpgf/>.

Finally, in the manuscript to be submitted below, I am the first author with Dr Matt Williams as the second author. My contribution was designing the study, selecting data analysis methods and creating the code for data analysis, collecting data, conducting data analysis, and writing the manuscript. Dr Matt Williams provided feedback throughout but did not write any of the manuscript.

## **Validating the Workplace Dignity Scale: Manuscript**

### **Purpose of the Study**

Until recently, a scale that measured workplace dignity had not been developed, despite many years of research that employed this concept. This has slowed extending knowledge on workplace dignity because valuable quantitative enquiry has been stilted. Nonetheless, Thomas and Lucas (2019) have recently developed the Workplace Dignity Scale (WDS) which seeks to address this gap. The purpose of the current study is to provide a preregistered replication of the validation work of Thomas and Lucas to ensure that the instrument is valid and reliable. This is especially important given the value of measuring workplace dignity, the potential for the WDS to be adopted in practice, and in the context of contributing to the ongoing methodological reform of psychology towards an open and reproducible science.

### **Introduction**

It has long been thought that dignity is a core human characteristic that distinguishes human beings from other animals (Bolton, 2007). Broadly defined as the sense of worth and respect deserved by all people (Hodson, 2001), the concept of dignity is referenced across a wide range of disciplines – from medical ethics to the law – and is considered the ultimate human value, to the extent that it forms a foundation of the United Nations universal declaration of human rights (United Nations, 1948). As such, it is important to explore how dignity manifests and is influenced in the many facets of human life. In particular, the workplace is suggested as an environment in which dignity can be both realised and violated.

Research undertaken on workplace dignity thus far has primarily relied on ethnographic methods, exploring, for example: the lived experience of nurses facing dignity violations (Khademi, Mohammadi, & Vanaki, 2012), discursive analyses of neoliberal discourse of dirty work (Purser, 2009), and the protection strategies that minority groups use

when facing workplace dignity violations (Baker & Lucas, 2017). These idiographic approaches demonstrate the breadth of manifestations of workplace dignity. However, a body of quantitative work focussing explicitly on workplace dignity has yet to be developed to complement these findings. The small number of studies throughout the organisational literature (e.g., Lucas et al., 2017) that have sought to quantitatively investigate workplace dignity have relied on variables from a single dataset which are theorised to relate to dignity (e.g., meaningful work, absence of supervisor conflict) but which do not directly assess workplace dignity itself. Thus, a robust measure of workplace dignity has been lacking until recently.

A nomothetic approach to investigating workplace dignity alongside idiographic approaches (e.g., ethnographic approaches) would lead to more comprehensive and integrated understanding. For example, a quantitative approach could assess the extent to which other factors influence and are influenced by workplace dignity (e.g., job enlargement, turnover intentions). In turn, new insights may be used in applied settings. For instance, a quantitative measure of workplace dignity could form part of a baseline assessment of an organisation's climate, from which the effects of an intervention (such as a diversity training programme) could be measured.

### **What is Workplace Dignity?**

It has been suggested that workplace dignity may involve four key principles (Thomas & Lucas, 2019) which can, to some extent, be traced back to religious teachings on human dignity (Sison et al., 2016). Firstly, workplace dignity is thought to be 'communicatively bound' whereby it depends on one's own assessment of their worth and the extent to which others display recognition and acknowledgement of this worth through social interaction (Lucas, 2015). Indeed, evidence from focus groups indicates that the violation or promotion of dignity occurs largely through communication, to the extent that demonstrating

competence or making a contribution was not itself enough to affect the experience of dignity (Lucas, 2015). Secondly, workplace dignity is self-constructed: While dignity judgements are influenced by discourses surrounding work, “it is the individual who is the ultimate arbiter of her or his experience of workplace dignity” (Thomas & Lucas, 2019, p. 76). Thirdly, there are two components to workplace dignity. There is dignity that is unconditional and is deserved by all (*inherent* dignity) and that which is earned through instrumental contributions in the workplace (*earned* dignity) (Hodson, 2001). The earned dignity component is central to distinguishing the construct of workplace dignity, from (inherent) dignity more broadly. Because one can earn dignity through their behaviours in the workplace, workplace dignity might be thought of as distinct from simply inherent dignity operating in a workplace context (Lucas, 2015). Finally, workplace dignity is considered bivalent in that it has positive elements, from which dignity can be promoted, and negative elements, from which dignity must be protected, analogous to Herzberg's (1968/2003) model of hygiene and motivating factors (Lucas, 2017; Thomas & Lucas, 2019).

Therefore, Thomas and Lucas (2019) suggest that the psychometric measure of workplace dignity should attend to these principles by using self-report, assessing social interactions, including positive and negative items, and enquiring about both inherent and earned dignity. It is through these principles that Thomas and Lucas developed the first psychometric measure of workplace dignity: The Workplace Dignity Scale (WDS). As for any psychological construct, it is important that the supposed measurement of workplace dignity is valid and reliable. For clarity, we are not relying on the definition of construct validity in the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). Instead, our analyses are influenced by Borsboom, Mellenbergh, and van Heerden's (2004) theory of validity, which posits that a test is a valid

measure of an attribute if the attribute exists and the attribute causes variation in the test scores (also see Borsboom, Cramer, Kievit, Scholten, & Franić, 2009), as well as Cronbach and Meehl's (1955) description of construct validity.

### **The Workplace Dignity Scale**

In the following paragraphs, we present the results of the initial validation of the WDS by Thomas and Lucas (2019). The WDS is the first psychometric measure of employee perceptions of dignity in the workplace. It is based on Lucas' (2017) broad definition of workplace dignity: "the self-recognised and other-recognised worth acquired from (or injured by) engaging in work activity" (p. 2549). The 18-item scale consists of self-report items such as "I feel respected when I interact with people at work", and "I am treated as less valuable than objects or pieces of equipment" which are responded to on a 7-point Likert-type scale (1 = *Strongly Disagree* to 7 = *Strongly Agree*).

**Initial scale development.** Initially, four general themes of workplace dignity (respectful interaction, recognition of competence and contribution, equality, and inherent value) emerged following focus group interviews with 62 working adults (Lucas, 2015). These participants were asked about their personal definitions of 'dignity at work' and any examples of where their dignity was either affirmed or denied. Conceptual definitions of the four themes and an additional general workplace dignity definition were generated prior to the development of a bank of positively- and negatively-valenced items (in accordance with the theorised bivalent nature of workplace dignity). The content validity of 97 items was then evaluated by 11 experts who rated the extent to which each item was essential for measuring its intended theme and gave open-ended feedback. Items yielding a consistent response from experts were retained (Ayre & Scally, 2014; Lawshe, 1975).

**Construct and nomological validity.** Thomas and Lucas (2019) analysed the subsequent 61-item scale based on the responses of 401 participants who passed attention

checks and met the inclusion criteria of working at least 30 hours per week, being over 21-years-old, and having at least 2 years paid work experience.

Thomas and Lucas settled on a six-factor measurement model indicated by 18 items using exploratory structural equation modelling (E-SEM) whereby models with an increasing number of factors were tested successively. The six-factor model was yielded because, relative to other models, this model was more parsimonious, had stronger factor loadings, fewer cross-loaded items, better global fit (based on the comparative fit index [CFI], Tucker–Lewis index [TLI], and root mean square error approximation [RMSEA]), and was supported by modification indices and background theory. Four factors reflected the previously identified themes, one factor reflected general workplace dignity, and the remaining factor consisted of negatively-valenced items, subsequently labelled workplace indignity.

Based on previous literature indicating that workplace dignity is multidimensional, and the high correlations between factors involving positively-valenced items, Thomas and Lucas tested a second-order model using CFA whereby Dignity was reflected by five first-order factors and could covary with the separate Indignity factor. This model was argued to provide the best fit to the data,  $\chi^2(128) = 392.97$ , CFI = .955, TLI = .946, RMSEA = .072 [.064, .080], compared with first-order, one-, and two-factor solutions, based on CFI, TLI, and RMSEA values, and modification indices. Internal consistency values ( $\alpha$ ) for the Dignity and Indignity factors, were .96 and .88, respectively. We note here, however, that a second-order model whereby Dignity is reflected by six factors (i.e., Indignity is a sub-factor loading on the Dignity higher-order factor) would have had identical fit to Thomas and Lucas' second-order model. In Study 3, Thomas and Lucas (2019) used the E-SEM procedure with a new sample ( $N = 532$ ) to assess whether they could replicate the second-order measurement model established in Study 2. The identical model was, indeed, supported, and a separate CFA on the new sample also supported the

model as having good fit:  $\chi^2(128) = 493.01$ ,  $p < .001$ , CFI = .958, TLI = .951, RMSEA = .073 [.066, .080]. Internal consistency estimates here for Dignity and Indignity were, again, .96 and .88, respectively.

Thomas and Lucas (2019) then evaluated the nomological validity of the WDS by exploring relationships between the WDS and theoretically-related variables using SEM. Dignity was significantly and positively related to Workplace Status, Need for Competence, and Interpersonal Justice, and negatively related to Workplace Alienation and Workplace Incivility (each of which yielded single-factor solutions). Dignity was also negatively related to the two factors yielded from Workplace Objectification: Objectification and Humanization (reverse coded to align with Objectification). Indignity was negatively related to Competence and Interpersonal Justice, and positively related to Incivility, Alienation, and Objectification, but not related to Status or Humanization. These relationships were largely the same when they calculated correlations using observed scale scores, with the exception of Dignity and Humanization which were not significantly correlated with one another. The authors took the different pattern of relationships that Dignity and Indignity had with other variables as further evidence for treating them as qualitatively different factors.

Finally, discriminant validity was assessed (and considered established) with Fornell and Larcker's (1981) criterion whereby the average variance extracted in WDS items (.72) was higher than the highest squared correlation with the other latent variables ( $r^2_{\text{dignity-justice}} = .59$ ).

### **Psychometric Research and Reproducibility**

Here, we turn away from Thomas and Lucas' (2019) study and discuss the importance of structural validation more broadly. The accuracy of conclusions that are drawn about latent constructs are highly dependent on the structural validity of the corresponding psychometric

measures and accordingly, the theory of a construct's *structure* needs to be subject to ongoing assessment. However, rigorous structural validation is rarely reported, leaving the possibility that conclusions are drawn on the basis of questionable measurement practices (Flake & Fried, 2019; Hussey & Hughes, 2019). Indeed, it has become common practice for Cronbach's  $\alpha$  to be the only reported assessment of a measure's structural fidelity, despite concerns around the usefulness of this statistic (Sijtsma, 2009).

Concerned about this tendency to underreport, Hussey and Hughes (2019) applied several tests of structural validity to 15 widely used scales in social and personality psychology with a large dataset ( $N = 81,986$ ). Subjected to more comprehensive assessment of structural validity, many of these scales appeared to perform questionably or poorly. Only 4% of the measures demonstrated 'good' structural validity: that is, they passed recommended thresholds for internal consistency, test-retest reliability, confirmatory factor structure *and* measurement invariance tests. Moreover, the less commonly a test of structural validity was reported in the literature for a given measure, the more likely it was to fail in Hussey and Hughes' analyses. This suggests potentially 'widespread hidden invalidity' among many frequently used measures and, thus, the robustness of many findings in social sciences that rely on the validity of these measures may be called into question.

In addition to calling for more comprehensive, multi-method approaches to validation, Hussey and Hughes (2019) highlight the scope for researcher degrees of freedom to threaten the validity of psychometric scales. That is, for each test of structural validity there are many available metrics at the researcher's disposal, a variety of different recommendations for cut-off values, and many other degrees of freedom that can differentially influence the apparent validity of a measure. Thus, the notion of v-hacking (*validity-hacking*) arises where researchers selectively choose and report tests that produce the most favorable outcomes which can lead to an over-inflation of the validity of a measure (largely incentivised by

publication pressure). In addition to selective reporting of favorable statistics, post-hoc tweaks (e.g., through modification indices) can make a model look ‘good’ at the expense of accurately identifying the truth because the model could be overfitted to the sample data.

The first step in combatting issues with undisclosed flexibility is to preregister the decision-making pathway prior to data collection, and to commit to reporting the results of this prespecified plan. This helps to solve the issue of hidden invalidity because researchers are unable to make decisions that make a scale *appear* valid. More generally, preregistration provides a means for conducting research according to an open science philosophy, whereby research is conducted in a transparent fashion in order to discover psychological truths (Chambers, 2017).

Lastly, like research involving non-significant results, replications are also notoriously difficult to publish, despite replication being critical for confirming observed phenomena (Schmidt, 2009). In the context of scale development, it is important that validity assessments are replicated – to the extent that Hussey and Hughes (2019) suggest that the internal factor structure of any given scale should be examined each time the scale is used.

### **The Current Study<sup>2</sup>**

In this study, we aim to examine the internal factor structure, reliability and nomological validity of the WDS, thus conducting a replication of Thomas and Lucas’ (2019) Study 2. The main differences between the current study and that of Thomas and Lucas are

---

<sup>2</sup> The Current Study section, as well as a Method Protocols section indicating what we planned to do, were uploaded as is presented here to the Open Science Framework and frozen prior to data collection, as part of the preregistration process (although please note that the tense has been changed and sample characteristics included now that the study has been completed). Furthermore, an R script detailing the data preprocessing and analysis, as well as a figure of the model to be tested were also uploaded a priori. See [osf.io/svpgf](https://osf.io/svpgf).

that: (a) we will be conducting a CFA to assess the fit of their model on new data, given that we are seeking to confirm their proposed factor structure, whereas they used E-SEM prior to CFA (our single CFA will be the only SEM-based analysis in our study); (b) in addition to reporting the global fit statistics they reported, we will report the standardised root mean square residual (SRMR); (c) we will not be including the Workplace Objectification Scale, given that it has not been subject to thorough validation work prior to in Thomas and Lucas' study; (d) we will only be replicating their observed variable/correlational approach to assessing nomological validity, whereas they additionally calculated SEM relationships with latent factors (for correlations calculated using sum scores, see above-diagonal area on Thomas and Lucas' Table 3, for nomological relationships estimated using SEM see below-diagonal area on Thomas and Lucas' Table 3 and Figure 1 of Thomas and Lucas) – this is in order to simplify our analyses and address the limitation of amplified SEM-based correlations; (e) in addition to reporting  $\alpha$  as in Thomas and Lucas, we will report McDonald's Omega hierarchical ( $\omega_h$ ) and Omega total ( $\omega_t$ ) reliability estimates, and; (f) we will be preregistering our data collection and analysis plans prior to collecting data, whereas Thomas and Lucas did not.

**Hypotheses.** We will be fitting a second-order CFA model that includes a Dignity factor reflected by five sub-factors and an Indignity factor that covaries with Dignity, which was the best fitting model in Thomas and Lucas (2019). We hypothesise that this model will provide good fit (see inference criteria in Method section) to the data (*hypothesis 1a*).

Thomas and Lucas suggested that because Dignity correlated with Indignity at  $-.59$  using observed scale scores and  $-.64$  using SEM estimation, and because the relationships between Dignity and other variables were not simply mirrored in the relationships between Indignity and those variables, Dignity and Indignity were related but operated on separate continuums. Therefore, we hypothesise that Dignity will negatively correlate (using observed

scale scores) with Indignity at  $-.7 < \rho < -.3$  (*hypothesis 1b*). That is, Dignity and Indignity will correlate negatively, but the magnitude will not be so large as to indicate that they measure the same thing.

To assess the nomological validity of the WDS we will be computing correlations (treating each of the scales as observed variables) between Dignity and Indignity and theoretically related variables. Specifically, based on Thomas and Lucas (2019), we hypothesise that Dignity will be positively correlated at  $\rho > .3$  with Need for Competence (*hypothesis 2a*), Interpersonal Justice (*2b*), and Workplace Status (*2c*). Need for Competence refers to individuals' desire to interact effectively with their workplace environment (Van den Broeck, Vansteenkiste, De Witte, Soenens, & Lens, 2010), and should, thus, be theoretically related to the *earned* component of workplace dignity (for more detail on the theoretical rationale for each nomological hypothesis, see Thomas and Lucas, 2019). Interpersonal Justice relates to the quality of interpersonal interactions an individual has, with respect being a key factor (Bies & Moag, 1986), and should be related to workplace dignity given the importance of respectful interaction for upholding inherent dignity (Sayer, 2007). It has been argued that an individual's status within their group is closely linked with the level of respect that they perceive to be treated with (Blader & Yu, 2017); one's workplace status, then, should be similarly related to workplace dignity given the importance of respect to dignity.

Similarly, as per Thomas and Lucas, we hypothesise that Dignity will be negatively correlated at  $\rho < -.3$  with Alienation (*2d*) and Incivility (*2e*). Alienation and Dignity are closely related with one another: to be psychologically damaged by or disconnected from one's work is often thought of as an automatic violation to one's dignity (Hodson, 2001). Incivility should be inversely related to Dignity because uncivil behaviors typically involve displaying a lack of respect (Andersson & Pearson, 1999).

The null hypotheses of  $\rho = .3$  and  $\rho = -.3$  are chosen because a correlation of .3 would indicate a medium effect size, and because such correlations had a magnitude of approximately .5 in Thomas and Lucas (2019). This allows us to directly test whether the true correlations have magnitudes greater than .3 in each case above.

If workplace dignity is to be considered bivalent with Indignity and Dignity being related but “operating somewhat independently” (Thomas & Lucas, 2019, p. 100), there needs to be *a priori* theory about Indignity’s relationship with nomological variables. However, Thomas and Lucas did not hypothesise about how Indignity would relate to nomological variables, nor outline a rationale *a priori* as to why it would or would not. Thus, in line with Thomas and Lucas, we make no formal predictions about Indignity’s relationship with nomological variables. Finally, we hypothesise that Dignity will demonstrate acceptable reliability (*hypothesis 3a*), as will Indignity separately (*3b*), with the lower bound of the 95% confidence interval for  $\omega_t > 0.7$ .

## **Method**

### **Sample Characteristics**

Participants were sought from the Prolific Academic database that were aged 21 years or over, worked at least 31 hours per week, worked in the USA, and had a Prolific approval rating of 95% or higher. Thus, for the purposes of replication, participants were recruited from the same population as in Thomas and Lucas (2019). We employed a stopping rule that data collection would cease when we reached 850 participant responses. This is approximately twice the sample size of Thomas and Lucas’ (2019) Study 2. Furthermore, based on a power analysis for computing a correlation between Dignity/Indignity and a theoretically related variable, for 90% power, an  $N$  of 809 would be required if  $\rho = .4$  when the null hypothesis is that  $\rho = .3$ . The smallest significant correlation for Thomas and Lucas (2019) (calculated using observed scale scores) between either Dignity or Indignity and the

nomological variables included in the current study was  $r = .48$  and was between Dignity and Status (also see inference criteria in the Nomological Validity subsection). Moreover, when using the maximum likelihood (ML) estimation method with SEM for the proposed model, which includes 60 statistical estimates, a sample size of 850 gives a ratio of approximately 14:1 participants per parameter. This ratio exceeds the minimum recommendation of 10:1 provided by Jackson (2003).

In total, there were 853 participants who completed the study (three more than our preregistered stopping rule), owing to an unanticipated feature of the Prolific data collection process. Three extra participants were included to fill in for those who Prolific determined to have taken too long to complete the survey but whose data were nonetheless suitable. Thus, we obtained slightly more data than our set target.

## Measures

The demographic variables measured included age, gender, ethnicity, country of work, highest educational level completed, occupation type, hours worked per week, years in current job, and years of prior work experience. These demographic variables, however, were not used in the main set of analyses.

**Scales.** The scales used in this study included the 18-item WDS (Thomas & Lucas 2019), the 4-item Interpersonal Justice Scale (Colquitt, 2001), the 6-item Need for Competence Scale (Van den Broeck, Vansteenkiste, Witte, Soenens, & Lens, 2010), the 5-item Workplace Status scale (Djurdjevic et al., 2017), the 7-item Workplace Incivility scale (Cortina, Magley, Williams, & Langhout, 2001), and 5 items (as per Thomas & Lucas, 2019) from the 8-item Work Alienation scale (Nair & Vohra, 2009). These scales were presented to participants using the same stems as in Thomas and Lucas, some of which were modified from their original papers; for instance, participants were asked “During the past FIVE years of employment, how often have you been in a situation where any of your supervisors or

coworkers...” for the Workplace Incivility Scale. These were the only measured variables used in the main analyses.

**Scale coding.** Each of the scales had the same number of response options as presented in Thomas and Lucas (2019). Scales were coded also according to Thomas and Lucas (e.g., Strongly Disagree and Strongly Agree on an item of the WDS were coded as 1 and 7, respectively). Items 1 and 4 of the Need for Competence Subscale were reverse coded. Greater scores on each scale indicate higher levels of the construct they intend to measure.

**Scale summation.** Mirroring Thomas and Lucas (2019), responses to each of the scales were summed into single scale scores, aside from the WDS, which was split into two subscales: Dignity (which included all items aside from I1, I2, I3, and I4) and Indignity (I1, I2, I3, I4) (see Figure 1). Scale scores for Dignity, Indignity, Interpersonal Justice, Need for Competence, Status, Incivility, and Alienation were created by summing the responses to the items within each scale for each participant.

### **Data Exclusions**

Only participants who answered “Yes” to the study consent question were permitted to participate. Respondents who answered “No” were directed out of the survey using survey flow settings, and their responses discarded. Prolific Academic only advertised this study to those who had specified that they were 21 years or over, work 31 or more hours per week, work in the USA, and had a Prolific approval rating of 95% or higher (as per our prescreening request). However, if a participant indicated that they did not meet these demographic criteria in response to the demographic questions early in the survey, they were directed out of the survey and their response discarded.

As a quality check, an attention check was included at the end of the Workplace Incivility Scale reading: “Please demonstrate that you are paying attention by ticking Often”. Participants who gave any other response or did not respond to this item were excluded

during data processing. Our preregistered exclusion criteria specified that in the event that there were any duplicate responses from the same Prolific worker (as indicated by a matching Prolific ID), and these duplicate responses were still present after applying the exclusion criteria specified above, then only the most recent response from each Prolific worker would be retained and the remainder excluded. Furthermore, participants were to be excluded if the computer returned a response outside the available range of responses on any item.

We did not specify any exclusion criteria at the level of data within participants because our use of an online survey with items in a rating scale format limited the possibility of extreme outliers. We prespecified that participants who had 11 or more of the 45 items in the main scales missing were to be excluded. After participants were excluded based on these and the earlier criteria, we conducted a single imputation using the expectation maximisation method to impute missing values. Only the main study variables in the dataset were included in the imputation model. Single imputation was chosen (rather than a more complex imputation method) because we anticipated that there would be very few missing data points (which is typical when recruiting through Prolific Academic; e.g., Margolis, Schwitzgebel, Ozer, & Lyubomirsky, 2019). Although the full information maximum likelihood procedure can handle missing data for the CFA, the remaining analyses required imputation. Thus, for the sake of simplicity, we used the imputed dataset for all the analyses, including the CFA. The above rule for missing data was specified to cover all instances where participants failed to complete the survey including those in which technical issues arose.

Of the 853 original participants, two participants did not work in the US, 32 did not work at least 31 hours per week, and 10 failed the attention check. There were no duplicates nor responses outside of the range of possible responses. The resulting number of participants included for analysis was 812 (as opposed to 809), as three of the excluded participants failed

more than one of the exclusion criteria. One item was missing for one participant, and, thus, no one was excluded based on the exclusion criteria for missing data.

## **Procedure**

We conducted an observational (non-experimental) study in an online survey format to assess the validity of the WDS. As such, there were no randomisation or blinding procedures. Participants completed the full questionnaire using the Qualtrics survey platform. Demographic information was collected first followed by the six main scales, the order of which was randomised.

## **Data Analysis**

**Confirmatory factor analysis.** We conducted a CFA with the maximum likelihood (ML) estimation method using the lavaan package (version 0.6-3; Rosseel, 2012) within the R software environment (version 3.6.0; R Core Team, 2019) to assess the factor structure of the 18-item WDS. The first item that each latent variable loaded on to was set to 1 (the ‘marker item’ approach), and unstandardised parameter estimates are reported.

*Assumptions.* Items were treated as continuous in line with the ML estimation method. This has been suggested as appropriate when items have at least five response options (Rhemtulla, Brosseau-Liard, & Savalei, 2012). ML has the following assumptions:

1. Independence of errors; this was not directly tested as it is captured in the model output.
2. Linear relationships between constructs; this is not feasibly testable.
3. Multivariate normality; we tested for multivariate skew and kurtosis using Mardia’s tests in order to demonstrate how well the data approximated a multivariate normal distribution. However, our preregistration specified that the analysis method was not to be changed based on the results of these diagnostics.

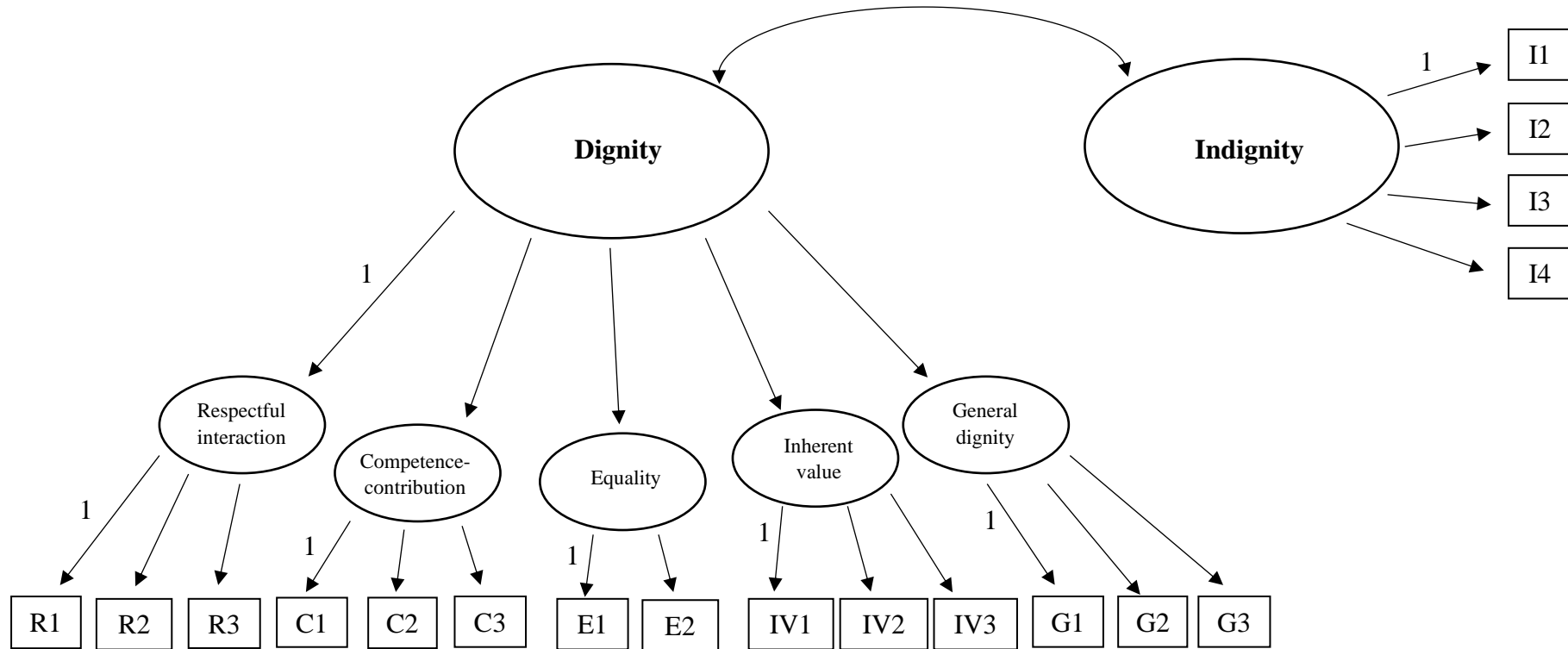
**Model specification and identification.** The best fitting model identified by Thomas and Lucas (2019) was fitted to our data whereby five dignity factors: Respectful Interaction, Competence-Contribution, Equality, Inherent Value, and General Dignity (and their respective indicators) reflect a higher-order Dignity factor, and with Indignity (and its respective indicators) as a second factor allowed to covary with Dignity.

The WDS includes 18 items (observed variables) and, hence, has 171 distinct sample moments. This figure was calculated using the formula:  $[v(v+1)]/2$ , where  $v$  is the number of observed variables (T. Brown, 2015). There were 42 parameters to be estimated, thus, giving 129 degrees of freedom. One factor loading for each of the seven factors (i.e., both the first- and second-order factors) was fixed (known), with the remaining (unknown) parameters estimated including: 18 error terms (one per item), 7 factor variances, 16 factor loadings, and one factor covariance (see Figure 1). We only identified one CFA model as per our preregistered plan.

**Inference criteria.** To assess the global fit of the CFA model we planned to report the values of each of the following fit indices (values indicating ‘good’ fit are given in parentheses, based on recommendations by Hu & Bentler, 1999): Root Mean Square Error of Approximation (RMSEA < 0.06), Standardised Root Mean Square of the Residual (SRMR < 0.09), Tucker Lewis Index (TLI > 0.95), and Comparative Fit Index (CFI > 0.95). A confidence interval for the RMSEA value was also reported. The Chi-squared test statistic was also reported accompanied by a  $p$  value with a significance threshold of  $p < 0.05$ . However, *hypothesis 1a* was to be considered supported if the SRMR, CFI, TLI, and RMSEA statistics *all* demonstrate good fit, in line with the criteria used by Hussey and Hughes (2019). In the case that the single CFA model being tested returned a negative result (i.e., hypothesis 1a was not supported), this result is still meaningful in the context of replication and, thus, we did not have contingency plans for probing unexpected CFA results (any analyses we did

conduct in this regard, such as investigating modification indices, were to be explicitly labelled as exploratory).

*Hypothesis 1b* was to be considered supported if the lower bound of the 95% confidence interval for the Spearman's correlation between Dignity and Indignity was greater than  $-.7$  and the upper bound less than  $-.3$ . Regardless of whether hypotheses 1a and 1b were supported, in the spirit of replication we planned to and did proceed with subsequent nomological (and then reliability) analyses, using sum scores to test correlations.



*Figure 1.* Thomas and Lucas' (2019) best fitting model that was preregistered to be identified in the current study using CFA. Note that each item will have its own measurement error term, with no covariances between error terms. Error terms are not displayed for the sake of brevity.

**Nomological validity.** Each of the scales were to be treated as observed variables and nomological validity was assessed by computing Spearman's correlations with 95% confidence intervals between Dignity (and Indignity) with each of: Need for Competence, Interpersonal Justice, Status, Alienation, and Incivility.

*Inference criteria.* Each of the sub-hypotheses of *hypothesis 2* that involved testing  $\rho > .3$  (i.e., Dignity positively correlating with Need for Competence, Interpersonal Justice, and Status) were to be considered supported if the lower bound of the confidence interval was greater than .3, for the null hypothesis that  $\rho = .3$ . Similarly, each of the sub-hypotheses of *hypothesis 2* that involved testing  $\rho < -.3$  (i.e., Dignity negatively correlating with Alienation and Incivility) were to be considered supported if the upper bound of the confidence interval was less than -.3, for the null hypothesis that  $\rho = -.3$ . The null hypotheses of  $\rho = .3$  and  $\rho = -.3$  were chosen because a correlation of .3 would indicate a medium effect size, and because such correlations had a magnitude of approximately .5 in Thomas and Lucas (2019). This allowed us to directly test whether the true correlations had magnitudes greater than .3 in each case above.

Familywise corrections for Type 1 error were not calculated because we followed a specific preregistered set of analyses and applying such a correction would have increased the risk of incorrectly finding that particular claims in the original study had not been replicated.

**Reliability.** We calculated Cronbach's  $\alpha$  and McDonald's (1999)  $\omega_h$  and  $\omega_t$  as estimates of reliability for the Dignity, Indignity, Interpersonal Justice, Workplace Status, Need for Competence, Incivility, and Alienation scales. These estimates were calculated using a 'common' EFA, with a three-factor solution estimated using the minimum residual method, a maximum of 100 iterations, and direct oblimin rotation. For each reliability statistic, the lower bound of the 95% confidence interval should be greater than 0.7 to demonstrate acceptable reliability (Hussey & Hughes, 2019).

**Inference criteria.** *Hypothesis 3a* was to be considered supported if the lower bound of the 95% confidence interval for  $\omega_t$  was greater than 0.7 for the *Dignity* scale. *Hypothesis 3b* was to be considered supported if the lower bound of the 95% confidence interval for  $\omega_t$  was greater than 0.7 for the *Indignity* scale. We chose  $\omega_t$  as our main estimate of reliability, while also reporting  $\alpha$  and  $\omega_h$  in line with modern recommendations (Dunn, Baguley, & Brunsdon, 2014; Hussey & Hughes, 2019; Revelle & Condon, 2018).

## Results

### Descriptive Statistics and Demographic Information

Table 1 displays the means and standard deviations of the scale scores on each of the measured variables. Table 2 displays the demographic characteristics of the participants included for analysis. Moreover, respondents reported that they had worked in their current jobs for an average of 6.5 years ( $SD = 5.2$ ) and had an average total work experience of 15.9 years ( $SD = 10.0$ ).

Table 1

*Descriptive Statistics of Scale Scores for Dignity, Indignity, and Nomological Variables*

Variables	<i>M</i> (Range of possible scores)	<i>SD</i>
Dignity	78.87 (14-98)	14.73
Indignity	8.77 (4-28)	5.28
Need for Competence	27.84 (6-42)	2.55
Interpersonal Justice	17.63 (4-20)	3.07
Workplace Incivility	14.00 (7-35)	5.50
Workplace Status	15.28 (5-25)	5.21
Workplace Alienation	16.61 (5-35)	8.44

Table 2

*Demographic Information*

Demographic variable	Category	% respondents
Gender	Male	57.8
	Female	41.3
	Non-binary	0.7
	Prefer not to say	0.2
Age	21-30	34.0
	31-40	40.1
	41-50	17.1
	Over 50	8.7
Ethnicity	White	74.0
	Black or African American	6.6
	Asian	6.0
	Other ethnicity <sup>a</sup>	13.4
Educational attainment	Some or all high school	5.9
	Some college no degree	15.3
	Bachelor's, Associate or Technical degree	57.0
	Master's, Professional, or Doctorate degree	21.8
Occupation	Computer/Mathematical	16.6
	Management	12.4
	Business and Financial Operations	9.5
	Other occupation <sup>b</sup>	61.5

*Note.* <sup>a</sup>Other ethnicity includes those of mixed ethnicity, American Indian or Alaska Native, Latino, and Hispanic. <sup>b</sup>Other occupation includes those from a wide range of Department of Labor classifications; the most common three are presented.

**Hypothesis 1a**

To assess the global fit of the CFA model, four fit indices were calculated: RMSEA = .073 [.068, .078], SRMR = .03, TLI = .958, CFI = .965. Each of these fit indices apart from the RMSEA met the prespecified inference criteria for the model to demonstrate good fit.

Thus, the hypothesis that the model will provide good fit to the data was not supported as the

inference criteria required every statistic to meet their specified thresholds. In addition, we also calculated  $\chi^2$ , although this did not form part of our inference criteria:  $\chi^2(129) = 687.98$ ,  $p < .001$ . Mardia's test statistics for multivariate skew and kurtosis were 10848.03 ( $p < .001$ ) and 162.68 ( $p < .001$ ), respectively, as such there was evidence of multivariate non-normality. Path estimates for the measurement model are presented in Figure 2.

### **Hypothesis 1b**

The Spearman's correlation calculated between Dignity and Indignity (using observed scale scores) was  $-.75$   $[-.78, -.70]$ . While the upper bound of the confidence interval was less than  $-.3$ , the lower bound was less than  $-.7$ . Thus, the hypothesis that Dignity would negatively correlate with Indignity at  $-.7 < \rho < -.3$  was not supported (see Table 3).

### **Hypothesis 2**

Spearman's correlations between Dignity and nomological variables, as well as between Indignity and nomological variables are presented in Table 3. Correlations calculated between Dignity and the nomological variables indicated positive relationships between Dignity and Need for Competence ( $r_s = .43$   $[.36, .49]$ ), Interpersonal Justice ( $r_s = .63$   $[.58, .67]$ ), and Workplace Status ( $r_s = .51$   $[.45, .56]$ ). The lower bounds of the confidence intervals for each of these correlations were greater than  $.3$ , which supported the hypotheses 2a, 2b, and 2c. Furthermore, Spearman's correlations indicated negative relationships between Dignity and Workplace Alienation ( $r_s = -.64$   $[-.68, -.59]$ ), and Workplace Incivility ( $r_s = -.54$   $[-.59, -.48]$ ). The upper bounds of the confidence intervals for these correlations were less than  $-.3$ , which supported hypotheses 2d and 2e.

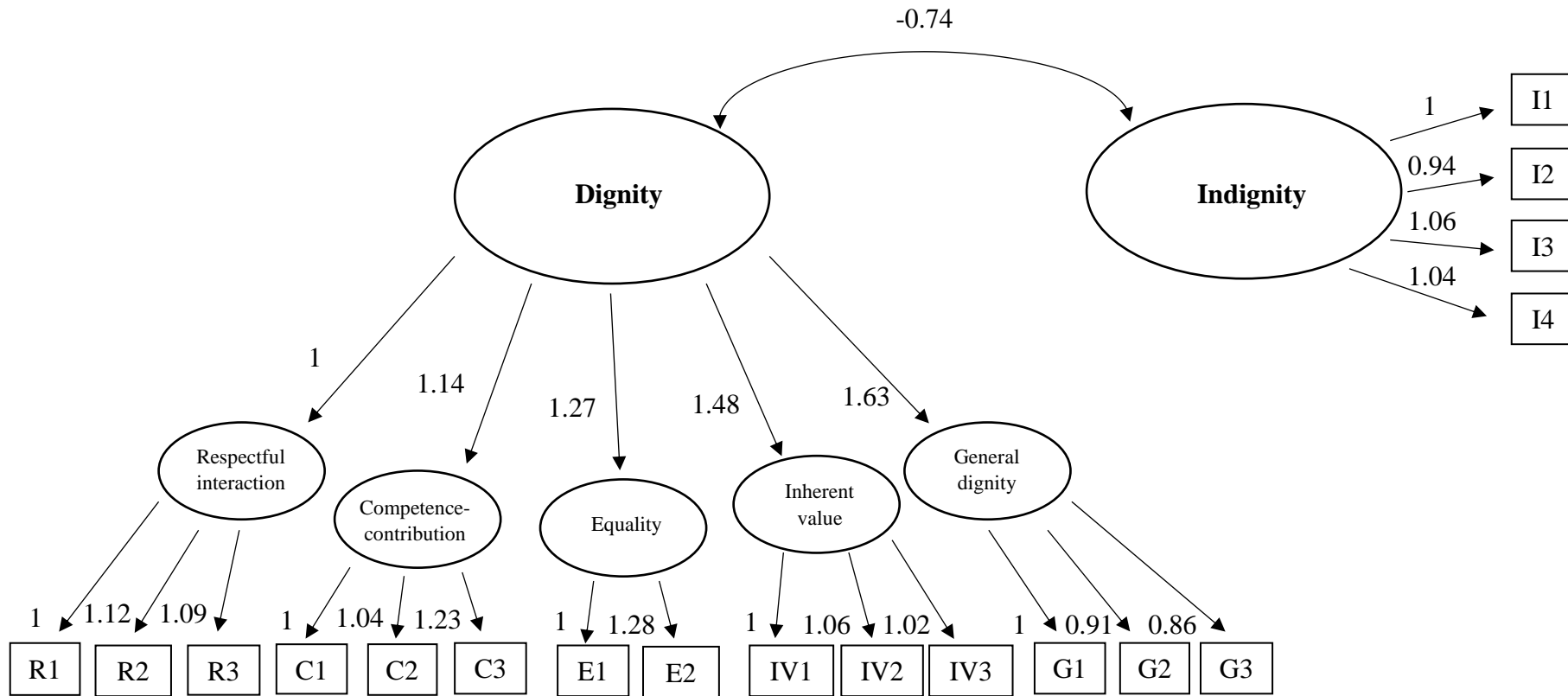


Figure 2. Measurement model including unstandardised path estimates. Error variances are omitted for brevity.

Table 3

*Correlations between Dignity, Indignity, and Nomological Variables*

Variables	Dignity	95% CI		Indignity	95% CI	
		Lower	Upper		Lower	Upper
Dignity	-	-	-	-	-	-
Indignity	-.75	-.78	-.70	-	-	-
Need for Competence	.43	.36	.49	-.42	-.48	-.36
Interpersonal Justice	.63	.58	.67	-.61	-.65	-.56
Workplace Status	.51	.45	.56	-.33	-.39	-.26
Workplace Alienation	-.64	-.68	-.59	.54	.48	.59
Workplace Incivility	-.54	-.59	-.48	.52	.46	.58

**Hypothesis 3**

The reliability of the WDS was assessed by estimating  $\omega_t$  and 95% confidence intervals for the Dignity and Indignity factors, separately. These estimates and those of  $\alpha$  and  $\omega_h$  are presented for each factor, as well as for the nomological variables in Table 4. The  $\omega_t$  reliability estimates and confidence intervals for Dignity and Indignity, respectively, were .98 [.97, .99] and .95 [.92, .98]. As the lower bound of the 95% confidence intervals were greater than .7, hypotheses 3a and 3b were supported.

Table 4

*Reliability Estimates for the WDS and Nomological Variables*

Variables	$\omega_t$	$\omega_h$	$\alpha$
Dignity	.98	.91	.97
Indignity	.95	.91	.94
Need for Competence	.94	.86	.91
Interpersonal Justice	.93	.92	.92
Workplace Status	.96	.95	.95
Workplace Alienation	.96	.90	.95
Workplace Incivility	.93	.85	.92

## Discussion

The purpose of this study was to conduct a close replication of Study 2 of Thomas and Lucas (2019). Thus, we investigated the validity and the reliability of the WDS by conducting a CFA to assess whether the best-fitting model suggested by Thomas and Lucas could be replicated in a larger sample, by calculating correlations between the Dignity and Indignity components of the WDS, as well as between Dignity and theoretically related variables (using observed scale scores), and by estimating reliability using a psychometrically sound measure of internal consistency,  $\omega_t$ . Furthermore, by preregistering our method protocols, we sought to contribute to the ongoing methodological reform of scientific psychology, particularly in relation to the need to ensure that important scales are reliable and valid prior to their widespread use (Hussey & Hughes, 2019).

### Confirmatory Findings

With respect to the fit of the CFA model, three out of four of the model fit statistics (SRMR, CFI, and TLI) met our prespecified criteria for indicating good fit as recommended by Hu and Bentler (1999) with only the RMSEA (= .073) failing to meet the recommended threshold (of less than .06; although we acknowledge that the RMSEA value was close to this threshold). Thus, *hypothesis 1a*, which required all four of the fit statistics to meet the prespecified criteria (Hussey & Hughes, 2019), was not supported. Nonetheless, the RMSEA, CFI, and TLI values described in this study were comparable to those identified by Thomas and Lucas (who did not calculate SRMR). However, in neither study did the RMSEA value (nor the lower bound of its 90% confidence interval) meet the recommended threshold suggesting acceptable fit (Hu & Bentler, 1999).

The next hypothesis proposed that the Dignity and Indignity components of the WDS would be negatively correlated with one another, but that the magnitude of the correlation would not be so large as to indicate that they, in fact, measure the same thing. This hypothesis

(*Ib*) was not supported: we found that the estimated correlation had a magnitude (-.75) that was greater than predicted. This result is perhaps inconsistent with the theoretical notion that Dignity and Indignity are qualitatively distinct constructs. Interestingly, this estimate was somewhat larger than that found by Thomas and Lucas (2019) (-.59), albeit in the same direction. Nonetheless, we did not plan *a priori* to test whether our correlations significantly differed from those of Thomas and Lucas (2019). However, given that our de-identified data will be publicly available on an online depository, future researchers may wish to explore this.

Each of the sub-hypotheses comprising *hypothesis 2* regarding correlations between the Dignity component of the WDS and theoretically related variables were supported. Dignity was positively correlated with Need for Competence, Interpersonal Justice, and Workplace Status, and negatively correlated with Workplace Alienation and Incivility. Furthermore, the magnitudes of the positive correlations were all significantly larger than .3, and the negative correlations larger than -.3. Here, our findings were very similar to the findings of Thomas and Lucas (2019) in terms of the size of these correlations, with the greatest difference in size between studies being .10 (for the correlation between Dignity and Interpersonal Justice). Again, a preregistered objective assessment of whether the correlations differ between studies may help to establish the theoretical position of workplace dignity within its proposed nomological network. Nonetheless, our results seem to support the nomological validity of the WDS.

Finally, we found that the WDS demonstrated strong internal consistency, as evidenced by particularly high  $\omega_t$  estimates for both Dignity and Indignity (of .98, and .95, respectively). Thus, *hypothesis 3* was supported. Unlike Thomas and Lucas (2019) who relied on  $\alpha$  as their estimate of reliability, we used  $\omega_t$  – which measures the total reliability of a scale by calculating the proportion of variance that is not attributed to measurement error – as

our main estimate in line with recent recommendations in the literature (e.g., Dunn et al., 2014). Furthermore, we also reported  $\alpha$  values that are consistent with those estimated by Thomas and Lucas (2019).

### **Are Dignity and Indignity Distinct Phenomena?**

Overall, in this study, our results were similar to most of the findings of Thomas and Lucas' (2019) Study 2. One question, however, that neither the initial validation work, nor the current replication attempt addressed is whether Dignity and Indignity should be considered as separate factors. The items that make up the Indignity factor are very similar to those that make up the Dignity factor, in that they seem simply to be negatively worded Dignity items. For example, the Indignity item: "I am treated in undignifying ways at work" seems to be a reversal of the Dignity item: "I am treated with dignity at work". Similarly, it is unclear whether the proposed factor structure of the WDS, in which Dignity has five first order factors and covaries with Indignity, is empirically appropriate over a factor structure where Indignity forms a sixth first-order factor beneath Dignity. As aforementioned, the global fit statistics would be no different in the single factor model than in the dual-factor model; the only difference between the models is that the relationship between Dignity and Indignity is expressed as a covariance term in the dual-factor model and a regression path in the single factor model.

Furthermore, Thomas and Lucas (2019) claimed that the relationships between Dignity and nomological variables were not simply mirrored in the relationships between Indignity and nomological variables (but with opposite signs), thus suggesting that Dignity and Indignity were qualitatively different from one another. However, we found that the pattern of nomological relationships with Dignity compared to that of Indignity *were* practically mirrored, showing mostly similar magnitudes though in opposite directions. Additionally, the correlation between Dignity and Indignity was particularly large (-.75),

indicating that perhaps Indignity, for the most part, is simply the reverse of Dignity, rather than being qualitatively unique. Moreover, it may be the case that in the original EFA by Thomas and Lucas (2019), the Indignity items emerged as a separate factor simply because they were all negatively worded. Indeed, factor analysis can sometimes cause spurious factors to emerge that reflect negatively worded items (Schmitt & Stuits, 1985; Spector, Van Katwyk, Brannick, & Chen, 1997; Woods, 2006). As alluded to above, inspection of the four Indignity items indicates that the item-wording is highly similar to that of items within other sub-factors of Dignity, suggesting that the Indignity items may not necessarily reflect a qualitatively different phenomenon than Dignity.

Whether Dignity and Indignity are distinct phenomena could not be readily examined in the current study because the two models – one where Indignity covaried with Dignity and one where it is a first-order factor caused by a higher-order Dignity factor – are empirically indistinguishable. We argue that further theoretical work is necessary to outline clearly how the two components of workplace dignity are different (if indeed they are). While it has been posited that they are separate phenomena, the empirical implications of this theoretical statement have not been unambiguously described. Once the empirical implications have been adequately laid out, quantitative tests can be designed to falsify the claim that Dignity and Indignity are indeed qualitatively distinct phenomena.

### **Limitations and Future Directions**

A limitation of the current study is that we did not attempt to replicate Study 3 of Thomas and Lucas (2019) which explored potential antecedents and consequences of workplace dignity, thus, further testing the external validity of the WDS. As such, we did not assess the validity of the WDS *comprehensively*. We also acknowledge that by replicating Study 3 we would have been able to further test the claim that workplace dignity is bivalent, given that Thomas and Lucas (2019) used the pattern of external relationships with Dignity

and with Indignity as evidence. Nonetheless, we maintain that further theoretical refinement of the nature of the two components should be prioritised in future work.

A second limitation is that the age of our sample was skewed towards younger workers (i.e., approximately 87% of the sample was aged between 21 and 50), whereas a substantial proportion (approximately 30%) of the US workforce is composed of older workers (U.S. Bureau of Labor Statistics, 2018). Moreover, with an aging workforce this proportion continues to grow, meaning that as it currently stands, the WDS may lack generalisability to older workers. Future research should investigate whether the current structural validity of the WDS extends to older workers by testing for measurement invariance between age groups; it may be the case that there are generational differences in how people interpret the items of the WDS.

Finally, we did not assess the factor structure of each nomological variable given that we wanted to keep a narrow focus for this replication attempt and the primary focus was to assess the factor structure of the WDS. However, we acknowledge the need to continually assess the factor structure of scales each time they are used with a new sample and encourage future researchers not only to reassess the validity of the WDS but also to reassess other measures used in conjunction with the WDS. Moreover, by using SEM to test nomological validity we could have accounted for the effects of measurement error on estimates of the relationships between WDS factors and nomological variables (Westfall & Yarkoni, 2016).

Future work should continue to probe the validity of the WDS, not only by conducting studies similar to this one, but also by validating the scale with populations from different countries to the US to assess measurement invariance. Indeed, it may be case that different cultures have different conceptualisations of dignity in the workplace (or simply dignity) (e.g., in non-Western countries) and that the WDS is not interpreted by different populations in a conceptually similar manner.

## **Conclusion**

Overall, we found evidence in support of the nomological validity and reliability of the WDS, although the proposed measurement model was not supported, suggesting that more work is needed to determine whether it requires refinement. The results of the present study in conjunction with those of Thomas and Lucas (2019) merit continued investigation of the factor structure of the WDS, replication and extension of the external validity work conducted in Study 3 of Thomas and Lucas (2019), and, depending on the results of this future research, the applied use of the WDS in organisations. It is our hope that such investigations are conducted according to Open Science principles to minimise potential QMPs during construct validation and, more broadly, to contribute to the ongoing methodological reform of scientific psychology.

## **Reflections on Completing a Registered Report-Based Master's Thesis**

In this final section, I reflect on my personal experience of completing what is a relatively unique thesis in that I initially followed the Registered Report format for publication, sought to conduct a systematic replication of a psychometric validation study, and preregistered my analysis script (in addition to the method protocols, data analysis plan, and hypotheses). Although I change to a personal style of writing, it is my hope that others might find reading my experience and opinions helpful for undertaking similar research.

### **Registered Reports and Psychometric Validation**

The use of the Registered Report format for conducting psychometric validation studies, such as the one presented in this thesis, is in its infancy. This is likely because the Registered Report format itself is new, paired with the fact that preregistered psychometric validation studies are rare. Nonetheless, based on my experience, I believe that this format is particularly suited to psychometric development and validation studies for two main reasons.

Firstly, the Registered Report format eliminates the incentive and the opportunity for researchers to engage in QMPs such as *v*-hacking (i.e., exploiting researcher degrees of freedom during reliability and validity assessments to produce favourable results). Because publication is *not* based on whether the measure of interest was shown to be valid, but rather the quality of the validation methods, researchers are not incentivised to, for example, selectively report global fit statistics that suggest good model fit for their scales. Furthermore, researchers have fewer opportunities to engage in QMPs because the validation methods (e.g., a description of which model fit statistics will be used) are specified prior to data collection, and if the methods are deemed to be scientifically sound, the study will be published contingent on those methods being closely followed.

Secondly, psychometric validation is complex and difficult, albeit critical because it ensures that the proposed measures of psychological constructs are sound. Indeed, it is

important that a measure's validity and reliability is investigated extensively and accurately before its widespread use as early validation studies are often relied on as the primary source of validity evidence for researchers employing such measures (Flake et al., 2017). To emphasise, many of the constructs studied in psychology, such as self-esteem, happiness, or depression, are taken for granted as being measured when their associated scales are employed, but it is the validation of those scales that supports that assumption. The Registered Report process, I argue, can enhance the quality of construct validation, as the critical predecessor to substantive research that *uses* measurement tools, by ensuring that it is conducted more transparently (through preregistration) and accurately (through emphasis on methodological rigour during peer-review) than current practice.

**The utility of preregistering analysis code.** The Registered Report format was particularly valuable because it forced me to extensively consider and articulate my validation analysis plans prior to data collection. While it was not required that a detailed code of these analyses be submitted to *Collabra*, I felt that creating and submitting this code *a priori* was particularly useful, to the extent that I would recommend journal editors formally request this of authors planning or submitting validation studies, especially given the numerous researcher decisions involved in scale development and validation. One reason it was useful is because it helped me to identify additional decisions that had to be made during analysis that I might not have otherwise considered before data collection; decisions which may have substantially affected my results. In turn, this allowed me to provide an even more comprehensive and, thus, transparent method protocols section. For example, although I had decided which statistics and inference criteria I would use to estimate internal consistency, in writing the code I found that I also had to decide *how* the estimates (and their confidence intervals) would be calculated (e.g., the type of EFA to specify in the calculation of Omega, number of bootstrapping iterations, rotation method, etc.).

Another reason that providing code of analysis plans is useful is because it allows authors to clarify exactly what their analyses are and detail specific information that might otherwise be excluded due to manuscript word limits. For example, it may not be obvious whether or what assumption checks have been undertaken by a researcher to ensure that their statistical methods are suitable for the data. Thomas and Lucas (2019) did not describe the results of any assumption checks (nor was it clear, for example, what estimation method they used, or whether Pearson's or Spearman's correlations were calculated). In fact, in the current study there was evidence to suggest that the assumption of multivariate normality, which is required for the ML estimation method, was not met. Providing the code for their analyses would have made it clear to readers whether Thomas and Lucas (2019) had completed any assumption checks. Furthermore, by making this code available alongside their full data set, this would have allowed other researchers to 'double-check' the results described in their manuscript and to conduct any further analyses on their data set that were not included in their code. Thus, by requesting code from authors at the point of Stage 1 Submission in the Registered Report format, journal editors would be promoting further transparency in the field. Moreover, for those planning replication studies, having open access to detailed codes of the analyses conducted in the to-be-replicated study will support researchers to undertake a more accurate replication attempt.

### **Systematic Replications of Psychometric Validation Studies**

Systematic replications of psychometric validation studies, like the one presented in this thesis, are rare. However, in my estimation, such replications are of critical importance to the field of psychology and it is my recommendation that these should become a normal part of the scale development and validation process (and should be preregistered). I believe that by pushing for systematic replications of validation studies, authors and journal editors can explicitly signal the need for construct validation to be treated as an ongoing process (Hussey

& Hughes, 2019). While it unrealistic to expect that researchers replicate the original scale development/validation study every time they use a given scale, it is important that they, nonetheless, conduct some validity assessment (e.g., of factor structure, internal consistency reliability) or provide *relevant* validity evidence for the scale they are employing. However, there should be at least *some* studies whose main aim is to attempt to systematically replicate original validation studies to have the effect of increasing the validity evidence base for other researchers to rely on when using a scale.

**Replication value and psychometric validation.** While psychometric validation replications should be conducted more often, it would be an inefficient use of resources for every scale validation study to be replicated (just like it would be inefficient to replicate every other study). Indeed, there should be some standard for assessing the *replication value* of such studies. In this thesis, I decided that replicating the WDS was a worthy pursuit because workplace dignity has received a lot of attention across a range of disciplines and the scale has the potential for widespread use in both academic and organisational contexts. Moreover, because the initial psychometric validation occurred recently (advanced access of Thomas and Lucas' study was available from October 2018), it was important to replicate this work to substantiate the validity and reliability of the WDS prior to its use. In this case, the replication value of Thomas and Lucas (2019) was *subjectively* determined.

However, an objective method for determining the replication value of any given study is necessary. Indeed, as the importance of replication has been pushed to the forefront of academic discourse due to the replication crisis, so have discussions of replication value (Brandt et al., 2014; Coles, Tiokhin, Scheel, Isager, & Lakens, 2018). Some have put forward methods for selecting which studies to replicate (i.e., for assessing replication value). For instance, Field et al. (2018) proposed using Bayesian methods to re-analyse existing research findings before making a qualitative assessment of replication value, which includes criteria

such as theoretical importance, insufficient prior investigation, and replication feasibility. Similarly, Isager (2019) described being in the process of developing a formula to quantify replication value. It is my recommendation that such assessments be applied, or new assessments developed, to determine the replication value of psychometric validation studies to help researchers prioritise psychometric replication efforts.

**Assessing replication outcomes.** One of the challenges I faced in conducting a systematic replication of a psychometric validation study, was in determining what set of results would constitute a successful replication of Thomas and Lucas (2019). Moreover, it was challenging to decide when and whether it was appropriate to include further tests to those included in the original validation study in an attempt to *improve* the validity assessment (and, thus, deviate from the exact methods of Thomas and Lucas). In other words, I attempted to strike a balance between carefully reproducing Thomas and Lucas' procedure and providing a more thorough assessment to determine the validity of the WDS (in part, this distinction reflects the difference between *direct* and *conceptual* replications; see Chambers, 2017; Simons, 2014). For example, while I included all the global fit statistics as in the original study, I also included the SRMR statistic and created hypotheses with inference criteria based on Hussey and Hughes (2019) to determine what values and combination of global fit statistics indicated 'good' model fit.

However, the question of whether my hypotheses were supported is different to the question of whether I replicated the findings of Thomas and Lucas. Indeed, I did not include an objective assessment of the replication *outcome* (i.e., whether or not the original findings were successfully replicated), as I was not aware of these methods at the time of Stage 1 Submission, though others that wish to use these methods on my data may do so. Such assessments typically involve the use of Bayesian tests and are becoming more prominent as the field of psychology places greater emphasis on the importance of replication (e.g.,

Simonsohn, 2015; Verhagen & Wagenmakers, 2014; Zondervan-Zwijnenburg et al., 2019). I suggest that journal editors require authors who undertake replications of validation studies to conduct and provide the results of tests that directly assess replication outcomes, to minimise the potential for replications to be inconclusive. Moreover, although preregistered systematic replications of psychometric validation studies are currently rare, I believe that such studies would benefit from a having standard set of tools to assess the replication outcome, which attend to the many stages of validity and reliability assessment (i.e., were internal consistency, factor structure, and measurement invariance findings replicated?).

### **Personal Reflections**

At the outset of this research project, the intention was to complete a Registered Report, and for the completed manuscript to form the core part of my Master's thesis. However, due to an abnormally long Stage 1 review process (which the editor acknowledged and apologised for) I was unable to continue with the Registered Report process. Nonetheless, I found that even getting partially through the Registered Report process was valuable for several reasons. Firstly, the requirement to provide method and analysis plans to a journal prior to collecting data, meant that the first thing I did during this project was think about and write down these plans. This was valuable because I was able to think about and justify the many decisions involved under less time pressure than if I had done this later in the year. Moreover, this process provided me with first-hand insight into just how many researcher degrees of freedom there are that typically go undisclosed in research and, as such, helped me to develop more of a critical eye for methodological details and their implications. Secondly, I found the peer-review process to be invaluable in improving the quality of my own work through receiving feedback from multiple expert sources. Finally, by undertaking this style of thesis I was exposed to much of the commentary on the replication crisis and the need for the discipline to be open and transparent. This has been an eye-opening journey and

has shown me the extent to which research findings are taken for granted. As such, I consider myself a stronger researcher for this challenging experience and I consider it essential to incorporate open science practices (e.g., by utilising the Registered Report format) in my future research endeavours.

## References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Andersson, L. M., & Pearson, C. M. (1999). Tit for tat? The spiraling effect of incivility in the workplace. *Academy of Management Review*, *24*(3), 452–471.  
<https://doi.org/10.5465/amr.1999.2202131>
- Ashforth, B. E., & Kreiner, G. E. (1999). “How Can You Do It?”: Dirty work and the challenge of constructing a positive identity. *Academy of Management Review*, *24*(3), 413–434. <https://doi.org/10.5465/amr.1999.2202129>
- Ashforth, B. E., & Kreiner, G. E. (2014). Dirty work and dirtier work: differences in countering physical, social, and moral stigma. *Management and Organization Review*, *10*(1), 81–108. <https://doi.org/10.1111/more.12044>
- Ayre, C., & Scally, A. J. (2014). Critical values for Lawshe’s content validity ratio: revisiting the original methods of calculation. *Measurement and Evaluation in Counseling and Development*, *47*(1), 79–86. <https://doi.org/10.1177/0748175613513808>
- Baker, S. (2014). You bring yourself to work: an exploration of LGB/TQ experiences of (in)dignity and identity. *Communication Studies Theses, Dissertations, and Student Research*. Retrieved from <https://digitalcommons.unl.edu/commstuddiss/27>
- Baker, S. J., & Lucas, K. (2017). Is it safe to bring myself to work? Understanding LGBTQ experiences of workplace dignity: LGBTQ experiences of workplace dignity. *Canadian Journal of Administrative Sciences / Revue Canadienne Des Sciences de l’Administration*, *34*(2), 133–148. <https://doi.org/10.1002/cjas.1439>
- Bal, M. (2017). *Dignity in the workplace*. New York, NY: Springer Berlin Heidelberg.

- Bélanger, J., & Edwards, P. (2013). The nature of front-line service work: distinctive features and continuity in the employment relationship. *Work, Employment and Society*, 27(3), 433–450. <https://doi.org/10.1177/0950017013481877>
- Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, 100(3), 407–425. <https://doi.org/10.1037/a0021524>
- Blader, S. L., & Yu, S. (2017). Are status and respect different or two sides of the same coin? *Academy of Management Annals*, 11(2), 800–824. <https://doi.org/10.5465/annals.2015.0150>
- Bolton, S. C. (2007). Dignity in and at work: why it matters. In *Dimensions of Dignity at Work* (pp. 3–16). <https://doi.org/10.1016/B978-0-7506-8333-3.50006-8>
- Bolton, S. C., & Laaser, K. (2013). Work, employment and society through the lens of moral economy. *Work, Employment and Society*, 27(3), 508–525. <https://doi.org/10.1177/0950017013479828>
- Borsboom, D., Cramer, A. O. J., Kievit, R. A., Scholten, A. Z., & Franic, S. (2009). The end of construct validity. In *The concept of validity: Revisions, new directions, and applications*. (pp. 135–170). Charlotte, NC, US: IAP Information Age Publishing.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, 111(4), 1061–1071. <https://doi.org/10.1037/0033-295X.111.4.1061>
- Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., ... van 't Veer, A. (2014). The replication recipe: what makes for a convincing replication? *Journal of Experimental Social Psychology*, 50, 217–224. <https://doi.org/10.1016/j.jesp.2013.10.005>

- Brennan, A., & Lo, Y. S. (2007). Two conceptions of dignity: honour and self-determination. In J. Malpas & N. Lickiss (Eds.), *Perspectives on Human Dignity: A Conversation* (pp. 43–58). [https://doi.org/10.1007/978-1-4020-6281-0\\_5](https://doi.org/10.1007/978-1-4020-6281-0_5)
- Brown, A. D. (2015a). Identities and identity work in organizations: identities and identity work. *International Journal of Management Reviews*, *17*(1), 20–40. <https://doi.org/10.1111/ijmr.12035>
- Brown, T. A. (2015b). *Confirmatory factor analysis for applied research* (Second edition). New York ; London: The Guilford Press.
- Chambers, C. (2017). *The seven deadly sins of psychology: a manifesto for reforming the culture of scientific practice*. Princeton University Press.
- Chambers, C. D., Feredoes, E., Muthukumaraswamy, S. D., & Etchells, P. (2014). Instead of “playing the game” it is time to change the rules: Registered Reports at AIMS Neuroscience and beyond. *AIMS Neuroscience*, *1*, 4–17. <http://dx.doi.org/10.3934/Neuroscience2014.1.4>
- Chiappetta-Swanson, C. (2005). Dignity and dirty work: nurses’ experiences in managing genetic termination for fetal anomaly. *Qualitative Sociology*, *28*(1), 93–116. <https://doi.org/10.1007/s11133-005-2632-0>
- Coles, N. A., Tiokhin, L., Scheel, A. M., Isager, P. M., & Lakens, D. (2018). *The costs and benefits of replication studies* [Preprint]. <https://doi.org/10.31234/osf.io/c8akj>
- Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology*, *86*(3), 386–400. <https://doi.org/10.1037/0021-9010.86.3.386>
- Cortina, L. M., Magley, V. J., Williams, J. H., & Langhout, R. D. (2001). Incivility in the workplace: incidence and impact. *Journal of Occupational Health Psychology*, *6*(1), 64–80. <https://doi.org/10.1037/1076-8998.6.1.64>

- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302. <https://doi.org/10.1037/h0040957>
- Crowley, M. (2014). Class, control, and relational indignity: Labor process foundations for workplace humiliation, conflict, and shame. *American Behavioral Scientist*, 58(3), 416–434. <https://doi.org/10.1177/0002764213503335>
- Djurdjevic, E., Stoverink, A. C., Klotz, A. C., Koopman, J., da Motta Veiga, S. P., Yam, K. C., & Chiang, J. T.-J. (2017). Workplace status: The development and validation of a scale. *Journal of Applied Psychology*, 102(7), 1124–1147. <https://doi.org/10.1037/apl0000202>
- Dufur, M. J., & Feinberg, S. L. (2007). Artificially restricted labor markets and worker dignity in professional football. *Journal of Contemporary Ethnography*, 36(5), 505–536. <https://doi.org/10.1177/0891241606294120>
- Dunn, T. J., Baguley, T., & Brunsten, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105(3), 399–412. <https://doi.org/10.1111/bjop.12046>
- Field, S. M., Hoekstra, R., Bringmann, L. F., & van Ravenzwaaij, D. (2018). *When and why to replicate: as easy as 1, 2, 3?* [Preprint]. <https://doi.org/10.31234/osf.io/h7js8>
- Flake, J. K., & Fried, E. I. (2019). *Measurement schmeasurement: Questionable measurement practices and how to avoid them* [Preprint]. <https://doi.org/10.31234/osf.io/hs7wm>
- Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. *Social Psychological and Personality Science*, 8(4), 370–378. <https://doi.org/10.1177/1948550617693063>

- Fleming, P. (2005). 'Kindergarten cop': paternalism and resistance in a high-commitment workplace. *Journal of Management Studies*, 42(7), 1469–1489.  
<https://doi.org/10.1111/j.1467-6486.2005.00551.x>
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39. <https://doi.org/10.2307/3151312>
- Francis, G. (2012). Publication bias and the failure of replication in experimental psychology. *Psychonomic Bulletin & Review*, 19(6), 975–991. <https://doi.org/10.3758/s13423-012-0322-y>
- Gelman, A., & Loken, E. (2013). *The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time*. Retrieved from [http://www.stat.columbia.edu/~gelman/research/unpublished/p\\_hacking.pdf](http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf)
- Gunn, A. M. (2011). The discursive construction of care when there is no care to be found: Organizational life (re)framed by those on the socio-economic margins facing job loss. *Culture and Organization*, 17(1), 65–85.  
<https://doi.org/10.1080/14759551.2011.530745>
- Hamilton, P., Redman, T., & McMurray, R. (2019). 'Lower than a snake's belly': discursive constructions of dignity and heroism in low-status garbage work. *Journal of Business Ethics*, 156(4), 889–901. <https://doi.org/10.1007/s10551-017-3618-z>
- Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The extent and consequences of p-hacking in science. *PLOS Biology*, 13(3), e1002106.  
<https://doi.org/10.1371/journal.pbio.1002106>

- Healy, M., & Wilkowska, I. (2017). Marx, alienation and the denial of dignity of work. In M. Kostera & M. Pirson (Eds.), *Dignity and the Organization* (pp. 99–124).  
[https://doi.org/10.1057/978-1-137-55562-5\\_6](https://doi.org/10.1057/978-1-137-55562-5_6)
- Herzberg, F. (2003, January 1). One more time: how do you motivate employees? *Harvard Business Review*, (January 2003). Retrieved from <https://hbr.org/2003/01/one-more-time-how-do-you-motivate-employees>
- Hodgkiss, P. (2013). A moral vision: human dignity in the eyes of the founders of sociology. *The Sociological Review*, 61(3), 417–439. <https://doi.org/10.1111/1467-954X.12049>
- Hodson, R. (2001). *Dignity at work*. <https://doi.org/10.1017/CBO9780511499333>
- Hodson, R. (2004). A meta-analysis of workplace ethnographies: Race, gender, and employee attitudes and behaviors. *Journal of Contemporary Ethnography*, 33(1), 4–38.  
<https://doi.org/10.1177/0891241603259808>
- Hodson, R. (2007). The consequences of management competence for working with dignity: a cross-methods validation. In *Dimensions of Dignity at Work* (pp. 109–133).  
<https://doi.org/10.1016/B978-0-7506-8333-3.50012-3>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55.  
<https://doi.org/10.1080/10705519909540118>
- Hussey, I., & Hughes, S. (2019). *Hidden invalidity among fifteen commonly used measures in social and personality psychology* [Preprint]. <https://doi.org/10.31234/osf.io/7rbfp>
- International Labour Organization. (2017). *Decent work and the 2030 agenda for sustainable development*. Retrieved from [https://www.ilo.org/global/topics/sdg-2030/resources/WCMS\\_436923/lang--en/index.htm](https://www.ilo.org/global/topics/sdg-2030/resources/WCMS_436923/lang--en/index.htm)

- Isager, P. M. (2019). *Quantifying replication value: a formula-based approach to study selection in replication research*. <https://doi.org/10.23668/psycharchives.2392>
- Jackson, D. L. (2003). Revisiting sample size and number of parameter estimates: some support for the N:q hypothesis. *Structural Equation Modeling: A Multidisciplinary Journal*, 10(1), 128–141. [https://doi.org/10.1207/S15328007SEM1001\\_6](https://doi.org/10.1207/S15328007SEM1001_6)
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23(5), 524–532. <https://doi.org/10.1177/0956797611430953>
- Jones, K. P., Sabat, I. E., King, E. B., Ahmad, A., McCausland, T. C., & Chen, T. (2017). Isms and schisms: A meta-analysis of the prejudice-discrimination relationship across racism, sexism, and ageism. *Journal of Organizational Behavior*, 38(7), 1076–1110. <https://doi.org/10.1002/job.2187>
- Karlsson, J. (2011). *Organizational misbehaviour in the workplace: narratives of dignity and resistance*. <https://doi.org/10.1057/9780230354630>
- Kensbock, S., Jennings, G., Bailey, J., & Patiar, A. (2014). Distinction work and its consequences for women working as room attendants within five star hotels on the Gold Coast, Australia. *Hospitality & Society*, 4(1), 55–73. [https://doi.org/10.1386/hosp.4.1.55\\_1](https://doi.org/10.1386/hosp.4.1.55_1)
- Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review*, 2(3), 196–217. [https://doi.org/10.1207/s15327957pspr0203\\_4](https://doi.org/10.1207/s15327957pspr0203_4)
- Khademi, M., Mohammadi, E., & Vanaki, Z. (2012). Nurses' experiences of violation of their dignity. *Nursing Ethics*, 19(3), 328–340. <https://doi.org/10.1177/0969733011433926>

- Kline, R. B. (2015). *Principles and practice of structural equation modeling: Fourth edition*. Retrieved from <https://www.guilford.com/books/Principles-and-Practice-of-Structural-Equation-Modeling/Rex-Kline/9781462523344>
- Kruschke, J. K., & Liddell, T. M. (2018). The bayesian new statistics: hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, 25(1), 178–206. <https://doi.org/10.3758/s13423-016-1221-4>
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563–575. <https://doi.org/10.1111/j.1744-6570.1975.tb01393.x>
- Lee, M. Y. K. (2008). Universal human dignity: some reflections in the Asian context. *Asian Journal of Comparative Law*, 3, 1–33. <https://doi.org/10.1017/S2194607800000211>
- Levelt Committee. (2011). *Interim report regarding the breach of scientific integrity committed by Prof. D.A. Stapel*. Retrieved from [https://web.archive.org/web/20160627142859/https://www.tilburguniversity.edu/upload/547aa461-6cd1-48cd-801b-61c434a73f79\\_interim-report.pdf](https://web.archive.org/web/20160627142859/https://www.tilburguniversity.edu/upload/547aa461-6cd1-48cd-801b-61c434a73f79_interim-report.pdf)
- Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, 3(3), 635–694. <https://doi.org/10.2466/pr0.1957.3.3.635>
- Lucas, K. (2015). Workplace dignity: communicating inherent, earned, and remediated dignity. *Journal of Management Studies*, 52. <https://doi.org/10.1111/joms.12133>
- Lucas, K. (2017). *Workplace dignity*. <https://doi.org/10.1002/9781118955567.wbieoc222>
- Lucas, K., Manikas, A. S., Mattingly, E. S., & Crider, C. J. (2017). Engaging and misbehaving: How dignity affects employee work behaviors. *Organization Studies*, 38(11), 1505–1527. <https://doi.org/10.1177/0170840616677634>
- Macklin, R. (2003). Dignity is a useless concept. *BMJ: British Medical Journal*, 327(7429), 1419–1420.

- Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: how often do they really occur? *Perspectives on Psychological Science*, 7(6), 537–542.  
<https://doi.org/10.1177/1745691612460688>
- Margolis, S., Schwitzgebel, E., Ozer, D. J., & Lyubomirsky, S. (2019). A new measure of life satisfaction: the riverside life satisfaction scale. *Journal of Personality Assessment*, 101(6), 621–630. <https://doi.org/10.1080/00223891.2018.1464457>
- Martin, S. E. (1980). *Breaking and entering: policewomen on patrol*. Berkeley: University of California Press.
- McDonald, R. P. (1999). Test homogeneity, reliability, and generalizability. In *Test theory: A unified approach* (pp. 76–120). Mahwah, NJ: Lawrence Erlbaum Associates.
- Mears, A., & Finlay, W. (2005). Not just a paper doll: how models manage bodily capital and why they perform emotional labor. *Journal of Contemporary Ethnography*, 34(3), 317–343. <https://doi.org/10.1177/0891241605274559>
- Messick, S. (1989). Validity. In *Educational measurement* (3rd ed., pp. 13–104). New York, NY: American Council on education and Macmillan.
- Nair, N., & Vohra, N. (2009). Developing a new measure of work alienation. *Journal of Workplace Rights*, 14(3), 293–309. <https://doi.org/10.2190/WR.14.3.c>
- Newman, K. S. (2000). *No shame in my game: the working poor in the inner city*. Knopf Doubleday Publishing Group.
- Nielsen, M. B., Glasø, L., & Einarsen, S. (2017). Exposure to workplace harassment and the Five Factor Model of personality: A meta-analysis. *Personality and Individual Differences*, 104, 195–206. <https://doi.org/10.1016/j.paid.2016.08.015>
- Noronha, E., Chakraborty, S., & D’Cruz, P. (2018). ‘Doing Dignity Work’: Indian security guards’ interface with precariousness. *Journal of Business Ethics*.  
<https://doi.org/10.1007/s10551-018-3996-x>

- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., ...  
Yarkoni, T. (2015). Promoting an open research culture. *Science*, 348(6242), 1422–1425. <https://doi.org/10.1126/science.aab2374>
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115(11), 2600–2606. <https://doi.org/10.1073/pnas.1708274114>
- Nosek, B. A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology*, 45(3), 137–141. <https://doi.org/10.1027/1864-9335/a000192>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- Paloniemi, S. (2006). Experience, competence and workplace learning. *Journal of Workplace Learning*. <https://doi.org/10.1108/13665620610693006>
- Pirson, M. (2017). *Humanistic management: protecting dignity and promoting well-being*. Retrieved from <http://public.eblib.com/choice/publicfullrecord.aspx?p=5101916>
- Pontifex. (2015, June 11). Where there is no work, there is no dignity. Retrieved from <https://twitter.com/Pontifex/status/608909299704709120>
- Purser, G. (2009). The dignity of job-seeking men: Boundary work among immigrant day laborers. *Journal of Contemporary Ethnography*. <https://doi.org/10.1177/0891241607311867>
- R Core Team. (2019). R: a language and environment for statistical computing. (Version 3.6.0). Retrieved from <https://www.R-project.org/>
- Radder, H. (1996). *In and about the world: philosophical studies of science and technology*. Albany, NY: State University of New York Press.

- Revelle, W., & Condon, D. M. (2018). *Reliability from alpha to omega: a tutorial* [Preprint].  
<https://doi.org/10.31234/osf.io/2y3w9>
- Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*(3), 354–373. <https://doi.org/10.1037/a0029315>
- Ritchie, S. J., Wiseman, R., & French, C. C. (2012). Failing the future: Three unsuccessful attempts to replicate Bem's 'Retroactive Facilitation of Recall' effect. *PLoS ONE, 7*(3), e33423. <https://doi.org/10.1371/journal.pone.0033423>
- Rivera, K. D. (2015). Emotional taint: making sense of emotional dirty work at the U.S. border patrol. *Management Communication Quarterly, 29*(2), 198–228.  
<https://doi.org/10.1177/0893318914554090>
- Roscigno, V. J., & Hodson, R. (2004). The organizational and social foundations of worker resistance. *American Sociological Review, 69*(1), 14–39. Retrieved from JSTOR.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(1), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Sayer, A. (2007). Dignity at work: broadening the agenda. *Organization, 14*(4), 565–581.  
<https://doi.org/10.1177/1350508407078053>
- Sayer, A. (2011). *Why things matter to people: social science, values and ethical life*. Cambridge University Press.
- Schilpzand, P., Pater, I. E. D., & Erez, A. (2016). Workplace incivility: A review of the literature and agenda for future research. *Journal of Organizational Behavior, 37*(S1), S57–S88. <https://doi.org/10.1002/job.1976>

- Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology, 13*(2), 90–100.  
<https://doi.org/10.1037/a0015108>
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment, 8*(4), 350–353. <https://doi.org/10.1037/1040-3590.8.4.350>
- Schmitt, N., & Stuits, D. M. (1985). Factors defined by negatively keyed items: the result of careless respondents? *Applied Psychological Measurement, 9*(4), 367–373.  
<https://doi.org/10.1177/014662168500900405>
- Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual Review of Psychology, 69*(1), 487–510. <https://doi.org/10.1146/annurev-psych-122216-011845>
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*(1), 107–120. <https://doi.org/10.1007/s11336-008-9101-0>
- Silva, M. R., & Caetano, A. (2016). Organizational justice across cultures: A systematic review of four decades of research and some directions for the future. *Social Justice Research, 29*(3), 257–287. <https://doi.org/10.1007/s11211-016-0263-0>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359–1366.  
<https://doi.org/10.1177/0956797611417632>
- Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science, 9*(1), 76–80. <https://doi.org/10.1177/1745691613514755>
- Simonsohn, U. (2015). Small telescopes: detectability and the evaluation of replication results. *Psychological Science, 26*(5), 559–569.  
<https://doi.org/10.1177/0956797614567341>

- Sison, A. J. G., Ferrero, I., & Guitián, G. (2016). Human dignity and the dignity of work: Insights from Catholic Social Teaching. *Business Ethics Quarterly*, 26(4), 503–528. <https://doi.org/10.1017/beq.2016.18>
- Spector, P. E., Van Katwyk, P. T., Brannick, M. T., & Chen, P. Y. (1997). When two factors don't reflect two constructs: how item characteristics can produce artifactual factors. *Journal of Management*, 23(5), 659–677. <https://doi.org/10.1177/014920639702300503>
- Steege, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, 11(5), 702–712. <https://doi.org/10.1177/1745691616658637>
- Stuesse, A. (2010). What's "Justice and Dignity" got to do with It?: migrant vulnerability, corporate complicity, and the state. *Human Organization*, 69(1), 19–30. <https://doi.org/10.17730/humo.69.1.886106412v421152>
- Thomas, B., & Lucas, K. (2019). Development and validation of the Workplace Dignity Scale. *Group & Organization Management*, 44(1), 72–111. <https://doi.org/10.1177/1059601118807784>
- Titunik, R. F. (1995). Status, vanity and equal dignity in Max Weber's political thought. *Economy and Society*, 24(1), 101–121. <https://doi.org/10.1080/03085149500000004>
- United Nations. (1948). *Universal declaration of human rights*. Geneva.
- U.S. Bureau of Labor Statistics. (2018). *Employed persons by detailed occupation and age*. 15.
- Van den Broeck, A., Vansteenkiste, M., Witte, H., Soenens, B., & Lens, W. (2010). Capturing autonomy, competence, and relatedness at work: construction and initial validation of the Work-related Basic Need Satisfaction scale. *Journal of Occupational*

and *Organizational Psychology*, 83(4), 981–1002.

<https://doi.org/10.1348/096317909X481382>

- Verhagen, J., & Wagenmakers, E.-J. (2014). Bayesian tests to quantify the result of a replication attempt. *Journal of Experimental Psychology: General*, 143(4), 1457–1475. <https://doi.org/10.1037/a0036731>
- Verkuil, B., Atasayi, S., & Molendijk, M. L. (2015). Workplace bullying and mental health: A meta-analysis on cross-sectional and longitudinal data. *PLOS ONE*, 10(8), e0135225. <https://doi.org/10.1371/journal.pone.0135225>
- Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology*, 100(3), 426–432. <https://doi.org/10.1037/a0022790>
- Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, 7(6), 632–638. <https://doi.org/10.1177/1745691612463078>
- Westfall, J., & Yarkoni, T. (2016). Statistically controlling for confounding constructs is harder than you think. *PLOS ONE*, 11(3), e0152719. <https://doi.org/10.1371/journal.pone.0152719>
- Woods, C. M. (2006). Careless responding to reverse-worded items: implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment*, 28(3), 186–191. <https://doi.org/10.1007/s10862-005-9004-7>
- Zondervan-Zwijenburg, M., van de Schoot, R., & Hoijtink, H. (2019). *Testing ANOVA replications by means of the prior predictive p-value* [Preprint]. <https://doi.org/10.31234/osf.io/6myqh>

## Appendix A

### Summary of Peer-Review Comments and Subsequent Changes to the Manuscript

The manuscript was submitted to the journal *Collabra* in June 2019 and I received the first round of peer reviews in late October. The outcome of the submission was that revisions were required, meaning that at least another round of peer review would be required before In Principle Acceptance was granted (i.e., before data collection could commence). Due to the short timeline for completing a one-year Master's thesis, the decision was made to not resubmit for a second round of reviews so that data collection could begin. However, reviewers' comments were incorporated in the manuscript within this thesis, which will be submitted as a complete manuscript to *Collabra* in the normal publication format. This is anticipated to take place in early 2020. Below is a summary of the major points that the reviewers suggested addressing and the action taken in response (in bold, below each point). Note that the original submission was not included in this thesis for brevity; the manuscript within this thesis incorporates these changes.

Some of the major points raised by the peer reviewers were (other minor points are excluded for brevity):

- That there should be an opening paragraph clearly explaining the goals and the purpose of the manuscript.
  - **Thus, I included a small paragraph titled “Purpose of the Study”.**
- That I needed to make it clear to the reader what theory of validity I was using, given that there are multiple theories available (i.e., what was meant by validity).
  - **Thus, I explicitly stated that we were relying on a combination of Cronbach and Meehl's, and Borsboom's theories of validity in the introduction of the manuscript.**

- That there needed to be a more detailed description of why the nomological variables should theoretically relate to workplace dignity.
  - **Thus, I included background information from prior studies theoretically situating workplace dignity within its nomological net.**
- That the Maximum Likelihood estimation method was more appropriate for this CFA (I originally intended to use the Diagonally Weighted Least Squares estimation method).
  - **Thus, in considering the reviewer's comment and investigating recommendations in the literature, the estimation method was changed to ML (because the items have more than 5 response options). This is detailed in the Method section.**
- That we needed to explicate why nomological relationships with Indignity were not hypothesised about.
  - **Thus, in the Hypotheses section of the manuscript I detailed our rationale for only having hypotheses about Dignity's relationship with nomological variables (i.e., lack of prior theory about how Indignity should relate to nomological variables, and wanting to keep in line with Thomas and Lucas' hypotheses).**

## Appendix B

### Low-Risk Ethics Application

#### 1. Application title

Low-risk Ethics Application for: A Validation of The Workplace Dignity Scale (WDS)

#### 2. Project summary

The purpose of this research is to conduct a second validation of a newly developed scale aimed to assess people's perceptions of Workplace Dignity (Workplace Dignity Scale [WDS]; Thomas & Lucas, 2019). The main aim of this research is to address concerns around the practice of validating psychometric scales. Specifically, researchers often fail to assess a scale's validity beyond the assessment that occurs during its initial development. Thorough assessment is necessary, especially in accordance with open science principles, because there are many steps during scale development (and subsequent validation in later studies) in which researchers' decisions can influence validity assessments. The current study will examine the factor structure, nomological validity and reliability of the Workplace Dignity Scale. In doing so, we will be partially replicating the work of Thomas and Lucas (2019). This observational study will be cross-sectional, where participants (N = 850) will complete a survey using the online survey platform Qualtrics. Recruited via Prolific Academic, participants will be sought from the USA that are over 21, and currently work at least 31 hours per week (as per the sample of Thomas & Lucas). Participant responses to the WDS, Workplace Status, Need for Competence, Interpersonal Justice, Workplace Incivility, Work Alienation, and Work-Related Flow Scales will be measured and participants will be reimbursed for their time (approximately 10 minutes). Only when we have enough confidence in the validity of this measure of workplace dignity can further empirical insights be developed. As such, the current study provides a necessary platform from which applied research can be developed that helps us understand and, consequently, improve workplace wellbeing.

#### 3. Describe the peer review process that has been used to discuss and analyse the ethical issues present in this project.

This project was peer reviewed by Dr Stephen Hill who was satisfied with the project. He suggested to clarify the nature of making the data available when gaining informed consent from participants. Accordingly, the following information is included in the information sheet at the informed consent process of data collection:

“The data we collect will initially only be accessible by the project team. Once the data has been analysed, we will ensure that any information in the dataset that might indicate who you are (e.g., your Prolific ID number) has been removed, and then post the data in an online repository (on the Open Science Framework). Other researches and members of the public will be able to obtain the data from this repository. Any identifiable data you provide (e.g., your Prolific ID number) will be deleted as soon as the project is complete [February 2020]. This de-identified data will be stored indefinitely.”

#### 4. List the ethical issues considered and explain how they have been addressed.

Some of the scales contain questions that may remind participants about previously upsetting experiences (e.g., “At work, I am valued as a human being”, “ I am treated as less valuable than objects or pieces of equipment”, “My dignity suffers at work”, “Have your co-workers/supervisor ever made jokes at your expense?” “I feel estranged/disconnected from myself”). However, the questionnaire itself is not likely to cause discomfort. Furthermore, the likelihood of psychological harm from this survey is low, given that participants are not recruited from a vulnerable group and because participants are only required to provide answers via Likert-scales (as opposed to if they were encouraged to reflect more deeply, through written answers). Nonetheless, if participants need support, contact information will be provided at the end of the survey for psychological support resources (e.g., <https://www.helpguide.org/> and <https://www.helpguide.org/articles/stress/stress-in-the-workplace.htm>).

Respect for the privacy and confidentiality of participants will be upheld. We will not request participant’s names, contact details, the names of the organisations they work for, or any other identifiers. Prior to beginning the survey, participants will be informed and required to consent to the condition that the raw anonymous data from the study will be stored in an open online depository (as per open science principles).

Data will be collected from Prolific Academic whereby participants will be reimbursed for their time spent undertaking the survey. The compensation provided to participants is deliberately not set at a level such that it would plausibly comprise an inducement (i.e., not a payment “extensive enough to persuade prospective participants to consent to participate in research against their better judgement” – see the [MOH Operational Standard for Ethics Committees, 2002](#)).

#### 5. With whom did you peer review your research?

Dr Matt Williams (Supervisor)

Peer Reviewed by: Dr Stephen Hill – School of Psychology

Date: 20 May 2019

Dear Casey Scott Campbell

Re: Ethics Notification - **4000021064 Low risk Ethics Application for: A Validation of The Workplace Dignity Scale (WDS)**

Thank you for your notification which you have assessed as Low Risk.

Your project has been recorded in our system which is reported in the Annual Report of the Massey University Human Ethics Committee.

The low risk notification for this project is valid for a maximum of three years.

If situations subsequently occur which cause you to reconsider your ethical analysis, please contact a Research Ethics Administrator.

Please note that travel undertaken by students must be approved by the supervisor and the relevant Pro Vice Chancellor and be in accordance with the Policy and Procedures for Course Related Student Travel Overseas. In addition, the supervisor must advise the University's Insurance Officer.

**A reminder to include the following statement on all public documents:**

*"This project has been evaluated by peer review and judged to be low risk. Consequently, it has not been reviewed by one of the University's Human Ethics Committees. The researcher(s) named in this document are responsible for the ethical conduct of this research."*

*If you have any concerns about the conduct of this research that you want to raise with someone other than the researcher(s), please contact Professor Craig Johnson, Director - Ethics, telephone 06 3569099 ext 85271, email [humanethics@massey.ac.nz](mailto:humanethics@massey.ac.nz)."*

Please note, if a sponsoring organisation, funding authority or a journal in which you wish to publish requires evidence of committee approval (with an approval number), you will have to complete the application form again, answering "yes" to the publication question to provide more information for one of the University's Human Ethics Committees. You should also note that such an approval can only be provided prior to the commencement of the research.

Yours sincerely



Professor Craig Johnson  
Chair, Human Ethics Chairs' Committee and Director (Research Ethics)

## Appendix C

### Participant Information Sheet

#### **Research Description**

The purpose of this research is to assess people's perceptions of dignity in their workplace and how this relates to other workplace experiences. If you decide and are eligible to participate you will be required to complete a short online survey that will take approximately 10 minutes to complete.

#### **Eligibility Criteria**

To be eligible to participate in this research you must be at least 21 years old, and currently be working at least 31 hours per week on average in the United States.

#### **Reimbursement for Participation**

You will be reimbursed USD \$1.30 for your time spent completing the survey.

#### **Data Management**

The data collected as part of this research will be used solely for research purposes. Survey answers will be anonymous (you will not be asked to provide your name or any contact information). The data we collect will initially only be accessible by the project team. Once the data has been analysed, we will ensure that any information in the dataset that might indicate who you are (e.g., your Prolific ID number) has been removed, and then post the data in an online repository (on the Open Science Framework). Other researchers and members of the public will be able to obtain the data from this repository. Any identifiable data you provide (e.g., your Prolific ID number) will be deleted as soon as the project is complete [February 2020]. This de-identified data will be stored indefinitely.

#### **Participant's Rights**

You are under no obligation to accept the invitation to participate in this research. If you decide to participate, you have the right to decline to answer any particular question and to stop answering questions at any time.

If you are interested in hearing about the results of this research you will be provided with an opportunity to indicate this at the end of the survey.

#### **Project Contacts**

Please feel welcome to contact me with any questions about this research.

Casey Scott-Campbell

#### **Committee Approval Statement**

This project has been evaluated by peer review and judged to be low risk. Consequently it has not been reviewed by one of the University's Human Ethics Committees. The researcher(s) named in this document are responsible for the ethical conduct of this research. If you have any concerns about the conduct of this research that you want to raise with someone other than the researcher(s), please contact Professor Craig Johnson, Director (Research Ethics), email [humanethics@massey.ac.nz](mailto:humanethics@massey.ac.nz)