ABSTRACT

STEGEMOLLER, SARA DEVIN. Just a MIsunderstanding?: Investigating Measurement Invariance across Remote and In-Person Workers on Job Satisfaction, Autonomy, and Social Support (Under the direction of Dr. S. Bartholomew Craig).

The practice of remote work has seen an unprecedented increase since the COVID-19 pandemic. As researchers and practitioners strive to understand the differences between remote and in-person workers, the ability to make meaningful comparisons across these groups is especially pertinent. Using the item response theory framework to assess differential item functioning, this study investigated measurement invariance on the job satisfaction survey (JSS), the Work Design Questionnaire (WDQ) autonomy scale, and the WDQ social support scale. Results suggested that several items on the JSS and WDQ social support scale exhibit measurement non-invariance across remote and in-person workers, with one WDQ autonomy item exhibiting measurement non-invariance across these groups. Further, the nature of a well-established relationship changed when non-invariant items were removed on the JSS and the WDQ autonomy scale. Future research directions and limitations are discussed.

Just a MIsunderstanding?: Investigating Measurement Invariance across Remote and In-Person Workers on Job Satisfaction, Autonomy, and Social Support

by
Sara Devin Stegemoller

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Psychology

Raleigh, North Carolina
2023

APPROVED BY:

Dr. S. Bartholomew Craig Committee Chair	Dr. Adam Meade
Dr. Rupert Nacoste	Dr. Paul Mulvey

BIOGRAPHY

Sara Stegemoller received her Bachelor's in Psychology from North Carolina State University in 2015.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The prevalence of remote work— a work arrangement in which employees spend at least some portion of their time away from the conventional office (Golden & Veiga 2005; Golden & Gajendran, 2019; Nilles, 1994)— has been steadily increasing for decades. This increase is largely due to the rapid advancement of information and communication technologies (ICTs) which have placed less constraints on where and how work can be accomplished (Allen et al., 2015; Gajendran & Harrison, 2007). Even so, prior to 2020, remote work was not practiced extensively (Wang et al., 2021). The Bureau of Labor Statistics reported that from 2017-2018, while 24.8% of workers indicated doing work at home, only 14.7% reported having days where they worked the entirety of their workday from home. Even less common was exclusively remote work, where employees worked at home five or more days per week, which was only practiced by 2.1% of U.S. workers (U.S. Bureau of Labor Statistics, 2019). The Coronavirus (COVID-19) pandemic, however, saw a significant climb in remote work (Kniffin et al., 2021), with 20% of U.S. employees being exclusively remote in April 2020 (Ozimek, 2020).

Of course, not all employees have the option to work remotely. Given the nature of their jobs, employees working in segments such as healthcare, retail, food production, and construction are less likely to be given the option to work remotely (Dey et al., 2020). However, according to an analysis based on data from Occupational Information Network (O*NET), American Time Use Surveys (ATUS), and the National Longitudinal Survey of Youth 1979 (NLSY79), roughly half (45%) of U.S. occupations are compatible with remote work (U.S. Bureau of Labor Statistics, 2020). These consist of occupations such as management, business, financial operations, and professionals. Amongst this group are knowledge workers— workers whose performance is heavily reliant on human intellect as opposed to physical labor (Powell &

Snellman, 2004). Therefore, the increase seen in remote work was especially notable among knowledge workers, with over 25% of all knowledge workers in the U.S. shifting to remote work in the weeks following the pandemic (Slack, 2020).

Among those with remote-compatible work, this forced exposure appears to have had a positive impact on feasibility and perceptions of remote work among both employees and employers (Barrero et al., 2021; Meluso et al., 2022; Ozimek, 2020), suggesting that this immediate upsurge may have some degree of permanence. Indeed, 38% of U.S. employees performed all or a portion of their work at home in 2021 (U.S. Bureau of Labor Statistics, 2022). Thus, while there are still workers for whom remote work is not a feasible option, it has become much more common for workers who can take advantage of remote work to do so. This has made the frequencies of remote and in-person employees much more similar than ever before.

Employee Experience across Remote and In-Person Workers

Though remote work has seen an unprecedented increase since the COVID-19 pandemic (Barrero et al., 2021; Kniffin et al., 2021), research on remote work is nowhere near its infancy. Studies dedicated to investigating the practice of working at home began in the 1980s (Ramsower, 1983). Nevertheless, findings are conflicting at best and there is still much to learn about how the experiences of remote and in-person workers differ (Boell et al., 2016; Brunelle & Fortin, 2021). These differences are particularly pertinent given that the percentage of remote workers has increased from around two percent (U.S. Bureau of Labor Statistics, 2019) to roughly 40% (U.S. Bureau of Labor Statistics, 2022) in the past four years. This makes these investigations evermore relevant.

Perhaps the most common tool for gathering information about workers and their work experiences are survey measures (e.g., Podsakoff & Organ, 1986; Rogelberg, et al., 2008).

Surveys allow researchers and practitioners to link abstract constructs or variables which cannot be directly observed (e.g., job satisfaction) to empirical observations (e.g., high vs. low job satisfaction; Lang & Tay, 2021). Typically, organizational researchers and practitioners focus on measuring behaviors, traits, attitudes, or perceptions of individual employees, teams, or organizations (Lang & Tay, 2021). A central goal in this area of research is to make comparisons between remote and in-person workers on relevant constructs or variables to draw conclusions, make recommendations, and discuss the implications of these different working environments (e.g., Allen et al. 2015; Charalampous et al., 2019; Gajendran & Harrison, 2007; Zöllner & Sulíková, 2021). For a given construct, this would involve identifying a measure to represent that construct, collecting and aggregating responses for remote and in-person workers, and analyzing whether the observed scores differ (Lang & Tay, 2021). Given their importance in the work context, variables of interest have included job satisfaction, autonomy, and social support (Allen et al., 2015; Charalampous et al., 2019; Gajendran & Harrison, 2007). Relatedly, there is evidence to suggest that these relationships play out differently depending on whether an employee works remotely or in-person (e.g., Brunelle & Fortin, 2021; Gajendran & Harrison, 2007; Schall, 2019). The ability to make meaningful comparisons across these groups, however, requires that the survey measures being used exhibit measurement invariance (MI; Vandenberg & Lance, 2000).

MI investigates whether the relationships between a construct of interest (e.g., job satisfaction, autonomy, social support) and the observed score on its measure are equivalent across groups of interest (Vandenberg & Lance, 2000). Without establishing MI, one cannot be sure if mean level differences represent true latent differences (e.g., true differences on job satisfaction) or are the result of some measurement artifact (e.g., conceptual differences on the

given job satisfaction measure across remote and in-person workers; Lang & Tay, 2021). If MI does not exist, mean scores across worker groups (i.e., remote and in-person workers) cannot be meaningfully compared and observed score differences may be misleading (Vandenberg & Lance, 2000). For instance, even if studies found remote workers to be more satisfied with their jobs than in-person workers, without the establishment of MI, it's possible that these differences are due to factors other than true job satisfaction levels. That is to say, conclusions based on observed score differences alone (e.g., remote workers have higher job satisfaction than in-person workers) may not accurately reflect true job satisfaction differences, which would call into question much of what we think we know.

Job Satisfaction

Locke (1976) defined job satisfaction as "a pleasurable or positive emotional state resulting from the appraisal of one's job or job experiences" (p. 1304). In addition to global job satisfaction, there are different facets of job satisfaction, such as satisfaction with pay, supervisor, coworkers, and others (Spector, 1985). Job satisfaction is one of the most widely studied attitudes in the organizational literature (Nübold et al., 2020). Simply typing "job satisfaction" into Google Scholar yields over 1.9 million results. This is not surprising, considering that satisfied employees are less likely to leave an organization (i.e., turnover; Lee & Mitchell, 1994; March & Simon, 1958; Mobley, 1977), which is a critical work outcome. Turnover can cost an organization over two times the salary of a given vacant job (Allen et al., 2010) and can lead to significant cuts in productivity (Hausknecht et al., 2009), damaging overall financial organizational performance (Park & Shaw, 2013). Employees with higher job satisfaction are also thought to perform better (Judge et al., 2001) and engage in counterproductive work behaviors with lower frequency (Dalal, 2005), both of which have additional implications for

overall firm performance and costs. Moreover, job satisfaction is studied in applied settings (Dey, 2022) to better diagnose problem areas and gauge intervention effectiveness (Spector, 1997). Thus, it is clear why job satisfaction is a popular topic in both academic and applied settings and that the ability to accurately measure it is of value.

Many studies have sought to investigate job satisfaction levels of remote and in-person workers to better understand its role in these different work environments (e.g., Allen et al 2015; Brunelle & Fortin, 2021; DuBrin, 1991; Gajendran & Harrison, 2007; Golden & Veiga, 2005; Vega et al., 2015). Unfortunately, findings have been mixed, making it difficult to fully understand how job satisfaction operates in these different conditions (Golden & Veiga, 2005). While there is less evidence for a null relationship (i.e., remote and in-person workers do not differ in their job satisfaction levels; Morganson et al., 2010) or a negative linear relationship between remote working and job satisfaction (i.e., remote workers are less satisfied with their jobs than in-person workers; Chapman et al., 1995; Cooper & Kurland, 2002; Pinsonneault & Boisvert, 2001), it is less clear whether the relationship is positive and linear (remote workers have higher job satisfaction than in-person workers), or curvilinear (the relationship between extent of remote working and job satisfaction is positive up until some threshold of remote work is exceeded).

The theoretical foundation for the idea that remote workers have higher job satisfaction than in-person workers is that remote workers are afforded higher levels of autonomy and flexibility (Dubrin, 1991; Wang et al., 2021). Given their psychological and spatial distance from direction supervision (Dubrin, 1991), remote workers can achieve a better balance between their personal and professional endeavors (Gajendran & Harrison, 2007; Golden & Veiga, 2005; Schall, 2019; Wang et al., 2021) and have more control over the timing and execution of their

tasks (Raghuram et al., 2003). Indeed, the positive linear relationship between remote working and job satisfaction has considerable support in the literature. Dating back to over two decades ago, Dubrin (1991) and Guimaraes and Dallow (1999) in their respective studies, found that remote workers had higher satisfaction than in-person workers. Meta analytic evidence echoes this notion, with evidence from 46 studies and approximately 13,000 employees finding higher job satisfaction levels among those who engage in remote work at a higher frequency (Gajendran & Harrison, 2007). Findings from a British household survey from 1993 to 2009 (Wheatley, 2012) found that, when comparing male and female home-based (i.e., remote) and employer's premises (i.e., in-person) workers, the mean job satisfaction was 13.5% higher for remote male employees and 7.5% higher for remote female employees. A within-person examination of remote working revealed that employees reported higher job satisfaction on days where they work from home (Vega et al., 2015). More recently, Brunelle and Fortin (2021) and Taboroši and colleagues (2022), in their respective studies, found that job satisfaction was higher amongst employees who were considered remote workers by their organization. Taken together, these results suggest that job satisfaction levels tend to be higher for remote workers.

The speculation of a curvilinear relationship between remote working and job satisfaction came about when Golden and Veiga (2005) set out to resolve contradictory findings of negative and positive relationships between remote working and job satisfaction. To accomplish this goal, they measured the extent of teleworking by asking participants to report the number of hours per week they spent working away from their central office. Results confirmed an inverted U-shaped relationship between extent of telework and job satisfaction, suggesting that job satisfaction increases with extent of remote work to a certain point, but decreases thereafter. Specifically, job satisfaction began to decrease at 15.1 hours of remote work per week. Golden (2006) replicated

these findings and found almost identical results, with the inflection point being 14.56 hours of remote work per week. Similarly, Virick and colleagues (2010) found support for a curvilinear relationship between extent of telecommuting and job satisfaction, with extent of telecommuting being measured in number of days spent working from home. The above suggests that the impact of remote work on job satisfaction is transient, and thus, remote work should not be implemented full time.

Researchers posit that working remotely only a portion of the time will serve to strike a balance between employees' need for autonomy, social support, and interaction (Golden & Veiga, 2005; Golden, 2006), as well as attenuate any negative consequences of social or professional isolation (Allen et al., 2015). Even still, support for a curvilinear relationship is mixed. For instance, a study of 89 remote workers and 103 in-person workers found that those performing remote work at least three days per week had higher levels of job satisfaction than in-person workers (Fonner & Roloff, 2010). Additionally, Schall (2019) found that, regardless of how often remote work was performed, remote workers had overall higher job satisfaction than in-person workers.

There is also evidence that remote workers and in-person workers are satisfied in different ways (Brunelle & Fortin, 2021). In their study, Brunelle and Fortin (2021) hypothesized that the relationships between (1) the psychological need for autonomy, (2) the psychological need for competence, and (3) the psychological need for relatedness and employee's job satisfaction would be moderated by whether an employee was considered remote or in-person. Their hypothesis was supported; the need for autonomy was more significantly related to job satisfaction for in-person workers, the need for competence was only significantly related to job satisfaction for in-person workers, while the need for relatedness was only significantly related

to job satisfaction for remote workers. These findings imply that the experience of a remote worker is "truly different" from an in-person worker (Brunelle & Fortin, 2021, p. 9) and that efforts to increase job satisfaction may look differently for these two groups.

Taken together, it seems there are still some unanswered questions about the relationship between type of worker (remote versus in-person) and job satisfaction. Even so, it appears that remote working can enhance job satisfaction, at least to a point. Given the claim that the additional autonomy remote workers enjoy enhances their working experience, many researchers have included autonomy in their studies to dig deeper into the nuances of remote and in-person work (Allen et al. 2015; Brunelle & Fortin, 2021; Fonner & Roloff, 2010; Gajendran & Harrison, 2007; Schall, 2019).

Autonomy

Autonomy has been a focal theme in organizational research for decades (Gajendran & Harrison, 2007; Hackman & Oldham, 1976; Morgeson & Humphrey, 2006). Autonomy is defined as "the extent to which a job allows freedom, independence, and discretion to schedule work, make decisions, and choose methods used to perform tasks" (Morgeson & Humphrey, 2006, p. 1323). While the above suggests the benefits of autonomy are unique to remote workers, levels of autonomy can be found in any job and has significant benefits. Specifically, autonomy is associated with outcomes such as lower turnover intentions (Thompson & Prottas, 2006) and higher job performance (Parker, et al., 2019; Spector, 1986). Perhaps most studied, however, are its links to motivation and satisfaction (e.g., Hackman & Oldham, 1976; DeCarlo & Agarwal, 1999; Liu et al., 2005; Parker et al., 2019; Spector 1986). Dating back roughly five decades ago, the job characteristics model (JCM; Hackman & Oldham, 1976) was developed to illustrate how job characteristics (e.g., autonomy) lead to job satisfaction and motivation by increasing the

meaningfulness of, and responsibility over, work. In addition to the JCM (Hackman & Oldham, 1976), these links can be described through two central theories of motivation: self-determination theory (SDT; Deci & Ryan, 1985) and the job-demands resources model (JD-R; Demerouti et al., 2001).

SDT describes the process through which motivation leads to behavior. According to this theory, a given behavior can be described in terms of the type of motivation that drives it. Expanding on the dichotomy of intrinsic to extrinsic motivation, SDT asserts a continuum of motivation for behavior ranging from externally controlled behavior to autonomous behavior: amotivation, external regulation, introjected regulation, identified regulation, integrated regulation, and intrinsic motivation (Gagné & Deci, 2005; Brunelle & Fortin, 2021). The closer a given behavior is to intrinsic motivation on this continuum, the more aligned it is with internal beliefs, preferences, and values. Relatedly, the more intrinsically motivating it is, the more autonomous it is. Given that autonomy is a basic psychological need (Ryan & Deci, 2000), employees with the latitude to structure their days and tasks in accordance with their own preferences and values (Chirkov et al., 2003) are more likely to have their need for autonomy met, and thus, are more likely to be satisfied.

The JD-R model (Demerouti et al., 2001) is a model illustrating the dual processes of stress and motivation, classifying work into two main domains: job demands— physical, psychological, social, and/or organizational factors that require extended effort and may interfere with an employees' goal attainment, and job resources— physical, psychological, social, and/or organizational factors that assuage this extended effort and allow employees to better reach their goals. High job demands (e.g., time pressures, role conflict) may lead to stress, exhaustion and burnout whereas job resources (e.g., autonomy, social support) can counterbalance these

demands by increasing employee motivation and engagement. Autonomy is considered a valuable resource for both remote workers and in-person employees (Bakker & Demerouti, 2007) due to its relation to intrinsic motivation (i.e., the psychological need for autonomy as discussed above; Deci & Ryan, 1985) and extrinsic motivation (i.e., allowing employees to better meet their work goals; Bakker & Demerouti, 2007; Sardeshmukh et al., 2012).

The link between autonomy and job satisfaction is apparent regardless of the type of worker (remote workers or in-person workers). Empirically, Brunelle and Fortin (2021) found that need for autonomy was positively related to job satisfaction for both remote workers and in-person workers. However, given that remote workers have more control and less constraints regarding how and where their days are spent, they have a different experience of autonomy when compared to in-person workers. Indeed, remote workers report higher levels of perceived autonomy than those who do not perform remote work (Gajendran et al., 2015). Relatedly, when compared to in-person workers, remote workers report higher levels of satisfaction of the need for autonomy (Brunelle & Fortin, 2021). While many in-person workers have the autonomy to decide how to execute tasks, they are still required to collocate with their coworkers, and thus do not have the physical separation that remote workers enjoy.

Autonomy has been identified as a critical component of remote work (Allen et al. 2015; Brunelle & Fortin, 2021) and has been found to mediate the relationship between remote work and job satisfaction (Gajendran & Harrison, 2007; Hornung & Glaser, 2009; Schall, 2019). That is to say, remote work leads to higher perceptions of autonomy which then results in higher levels of job satisfaction. This can materialize in various ways, such as remote workers being able to start their days earlier or later, work at coffee shops (Allen et al., 2015), or take their kids to and from school (Hornung & Glaser, 2009). This lends itself well to meeting one’s need for

autonomy, as remote workers may be better suited to balance the demands of both their professional and personal lives (Golden & Veiga, 2005). Further, they have autonomy over aspects such as clothing, lighting, layout, and other physical ambient features, making further distinctions between their experience of autonomy when compared to in-person workers (Elsbach, 2003; Standen, 2000).

Studied to a lesser extent is the idea that remote work has a negative impact on autonomy. While remote workers do have certain advantages regarding autonomy, some have speculated that electronic monitoring (Ambrose & Adler, 2000) and micromanaging (Harrison et al., 2000) may increase in cases where supervisors do not trust or understand the efforts of their employees. Thus, though autonomy may increase for remote workers, using ICTs at home may place a unique set of constraints on their behaviors in a setting (e.g., home) that was previously beyond the visibility of managerial control (Sewell & Taskin, 2015). Additionally, Warr's (1987) vitamin model suggests that too much of a good thing (e.g., autonomy) can have a negative impact. For instance, too much autonomy may in fact increase stress as workers have to exercise more self-control to achieve work goals and meet deadlines.

All in all, the construct of autonomy has been investigated heavily as it relates to work, and it is commonly perceived as a benefit. Another central theme in the literature on remote work is the interplay between autonomy and social support, which researchers have coined "the teleworking paradox" (Gajendran & Harrison, 2007). This paradox underlines how the higher levels of autonomy that are inevitable for remote workers may have the "mutually incompatible consequence" (Gajendran & Harrison, 2007; p. 1526) of hindering their interpersonal bonds with their coworkers, as well as reducing the frequency and quality of communication, when compared to in-person workers.

Social Support

Social support has become one of the most popular constructs in organizational research (French et al., 2018), and it has been identified specifically as an ideal work characteristic for remote workers (Wang et al., 2021). Unfortunately, the literature on social support is somewhat fragmented, leading to different conceptualizations of this construct across studies (Jolly et al., 2021). In this section, while variables discussed may not be labeled "social support" in their respective studies, they will only be discussed if they have definitional or measurement overlap with the definition of social support being utilized in this study. Thus, as definitions of social support vary heavily (Jolly et al., 2021), it is important to note that the definition of social support being applied in this paper is "the degree to which a job provides opportunities for advice and assistance from others" (Morgeson & Humphrey, 2006, p. 1324). This includes aspects such as an employee having the chance to meet with and get to know coworkers and supervisors, feeling as though their coworkers and supervisors take personal interest in them, and having the opportunity to establish friendships at work (Sims et al., 1976; Morgeson & Humphrey, 2006). To this end, jobs high in social support will be associated with opportunities to gain information and resources from coworkers as well as have friendly interactions with them. This, in turn, may result in more learning opportunities, more ease in clarifying expectations, and lower stress and uncertainty (Brunelle & Fortin, 2021; Demerouti et al., 2001; Golden & Gajendran, 2019; Humphrey et al., 2007).

The benefits of social support at work can be further illustrated through the theories discussed above, namely SDT (Deci & Ryan, 1985) and JD-R (Demerouti et al., 2001). Regarding the SDT, social support can satisfy individuals' need for relatedness, which is a basic psychological need (Ryan & Deci, 2000). In a work context, the need for relatedness is fulfilled

when an employee feels connected and supported (Brunelle & Fortin, 2021), can establish a positive connection with interaction partners (Baumeister & Leary, 1995), and can develop friendships at work (Morgeson & Humphrey, 2006). Regarding the JD-R, social support is classified as a valuable job resource that can help buffer against experiences of stressors and strains that come with high work demands (Demerouti et al., 2001; Viswesvaran et al., 1999). Empirically, meta-analytic evidence (Kossek et al., 2011; Mathieu et al., 2019) reveals that social support has a positive relationship job satisfaction and a negative relationship with stressors (e.g., role conflict) and strains (e.g., disengagement). Altogether, social support is a resource for employees "which enriches the psychological context of work and erodes the effects of stress" (Hughes et al., 2022; p. 388).

Researchers ascribed the term "teleworking paradox" to illustrate the contradiction that the increased autonomy enjoyed by remote workers will be the very thing that hinders their opportunity for social connections at work (Gajendran & Harrison, 2007). This is not to say that remote workers cannot experience social support. The methods through which they must do so, however, will differ from in-person workers, which can differentiate certain aspects of the experience. As working remotely comes with a reduced social presence, remote workers must rely heavily on ICTs to establish relationships and communicate with their coworkers and managers, given they no longer have sustained face-to-face interactions with them (Brunelle & Fortin, 2021). This "alters the dynamics of work-related interpersonal processes" (Allen et al., 2015, p. 52) and differentially impacts how relationships are developed and maintained (Brunelle & Fortin, 2021).

According to media richness theory (Daft & Lengel, 1986), the forms of communication necessary for remote working (e.g., instant messaging, emails, and video conferencing) are "less

rich" than the face-to-face interactions in-person workers have, as they diminish the experience of nonverbal cues and have a "reduced personal focus" (Nesher Shoshan & Wehrt, 2022, p. 830). This has the potential to make these interactions less personal and efficient for remote workers, which in turn may reduce the bonds and connections made with coworkers and supervisors (Brunelle & Fortin, 2021; Golden, 2006). Empirically, face-to-face interactions have been identified as the most important form of communication above telephone calls, instant messaging, and e-mails regarding maintaining friendships at work (Allen et al., 2015), a key aspect of social support. In fact, remote workers have reported missing the face-to-face aspect of being in the office, such as general office interactions and idle hallway conversations with colleagues (Allen et al., 2015; Charalampous et al., 2019; Gajendran & Harrison, 2007; Grant, et al., 2013; Kurland & Cooper, 2002).

By definition, remote workers are spatially distanced from their coworkers for at least some portion of the time. This has led many researchers to deem social isolation, which is a significant and negative correlate of social support (Bentley et al., 2016), one of the greatest challenges of remote work (Allen et al., 2015; Golden et al., 2008). Without intentionally creating opportunities for formal and informal interactions, remote workers may feel excluded from their coworkers or supervisors (Pyöriä, 2011). Indeed, a study by Vander Elst and colleagues (2017) found that the extent of remote working was negatively related to social support, which increased emotional exhaustion— a job demand, a key component of burnout, and a significant correlate of turnover (Demerouti et al., 2001). This suggests that higher perceptions of social support can work to diminish feelings of emotional exhaustion, burnout, and turnover intentions. More recently, in a sample of remote workers, social support was found to have a negative relationship with emotional exhaustion through lower levels of loneliness and

a positive relationship with life satisfaction through lower levels of loneliness in a sample of remote workers (Wang et al., 2021). This suggests that social support may counteract the potential loneliness or social isolation that has been discussed as a challenge for remote workers.

Additional findings suggest more nuanced relationships regarding social support. In their meta-analysis, Gajendran and Harrison (2007) looked at relationship quality, which included many aspects of social support such as opportunities to have interactions with coworkers and supervisors as well as create and maintain friendships. They found that both the extent to which an employee works remotely as well as the target of the relationship (i.e., coworker vs. supervisor) are important to understanding the role of remote work in this context. Specifically, among those who worked remotely a majority of the time, relationship quality with coworkers was more negatively impacted than relationship quality with supervisors (Gajendran & Harrison, 2007). In fact, the extent of remote working was positively related to relationship quality with supervisors. The authors speculate that this may be due to increased focus on building quality relationships in this context given the spatial distance. These findings bring into question the reality of the "teleworking paradox."

Recent work obfuscates the role of social support among remote workers even further. When exploring the relationship between remote work and job performance, Golden and Gajendran (2019) found a positive relationship only when social support was low. When social support was high, the relationship between remote work and job performance was nonsignificant. In other words, it appears that low levels of social support matter more for remote workers' job performance than high levels, which do not have much impact. Authors speculate that employees who endure unfavorable relationships at work benefit from the spatial distance remote working permits.

When looking at need for relatedness (which overlaps with social support in terms of its feeling of connection to others at work), researchers found that this need was better satisfied among remote workers than in-person workers (Brunelle & Fortin, 2021). Upon closer examination, researchers found that there were organization-wide practices (e.g., opportunities to meet coworkers) implemented to encourage social connection among remote workers. Even so, this suggests that the collocation in-person workers are afforded did not foster the same level of relatedness remote workers reported. All in all, the above indicates that there are differences between these groups in terms of establishing and maintaining social support and connection at work, though the intricacies may not be entirely clear (Jolly et al., 2021).

As can be gleaned from the above sections, there are many aspects of the work experience of remote and in-person employees that are distinct from one another. For decades, researchers have investigated the observed mean differences between remote and in-person workers on constructs like job satisfaction, autonomy, and social support. With the recent increase in remote work, it is likely that these distinctions will be highlighted and illuminated even further (Wang et al., 2021). However, the knowledge gained from research studies is only valuable to the extent that the assumptions behind these investigations are being met and tested as a study is carried out. A uniting feature of the studies discussed thus far is that none of them mentioned performing tests of MI on their measures across remote and in-person workers prior to making conclusions about observed mean differences. This alone could alter the veracity of these findings.

Measurement Invariance Applied to Organizational Constructs

Given the recent upsurge in remote work, the comparability between remote workers and in-person workers is an especially important topic. Moreover, as the above review highlights, the

experiences of remote and in-person workers differ on more than simply physical surroundings, but also in terms of the psychological experience. Direct comparisons between these groups are a requisite for any researcher or organization hoping to better understand these different working environments.

MI is an assessment of equivalence in psychometric properties of a given measure across groups to understand whether the construct being measured has the same meaning to those groups (Putnick & Bornstein, 2016). The underlying question of MI, then, is "whether or not, under different conditions of observing and studying phenomena, measurements yield measures of the same attributes. If there is no evidence indicating presence or absence of measurement invariance—the usual case—or there is evidence that such invariance does not obtain, then the basis for drawing scientific inference is severely lacking: findings of differences between individuals and groups cannot be unambiguously interpreted" (Horn & McArdle, 1992, p. 117). Put differently, an invariant measure is one in which individuals belonging to different groups, but with equal standing on a latent variable, have the same probability of receiving the same score for that latent variable using that measure (Meriac et al., 2010). Without establishing MI, it cannot be determined if mean level differences reflect true latent differences or are the result of some measurement artifact (Lang & Tay, 2021).

In the context of this study, the observed differences between remote and in-person workers reviewed above might be a signal that the constructs being measured (e.g., job satisfaction, autonomy, social support) are being conceptualized differently in these groups (Vandenberg & Lance, 2000). MI tests are an important and necessary step in identifying the appropriateness of measurements used to advance psychological research (Drasgow, 1987; Vandenberg & Lance, 2000; Tay et al., 2015), as it would be inappropriate to make comparisons

based on mean differences using measures lacking MI (Meredith, 1993; Vandenberg & Lance, 2000). Taken together, establishing a measure as invariant or non-invariant prior to utilizing it in research is critical to the generalizability, validity, and overall usefulness of the measure itself and the findings it reveals.

In organizational research, one method commonly used to test for MI is the Confirmatory Factor Analysis (CFA) approach called multiple-group confirmatory factor analysis (MGCFA; Vandenberg & Lance, 2000). At its core, a CFA is an a priori modeling structure which tests the hypothesis that a given latent construct(s), which cannot be measured directly, accounts for the variance and covariance among a set of indicator variables (Fabrigar et al., 1999). Testing this a priori structure involves observing how well this structure is reproduced in the observed covariance matrix using these indicator variables (Brown, 2015). The goal of MGCFA, then, is simply to demonstrate that this a priori structure is comparable across groups so that meaningful interpretations about observed group differences can be made (Tay et al., 2015). For a one-factor measure of job satisfaction with four items, an MGCFA would be testing the hypothesis that understanding someone's underlying job satisfaction level can be accomplished by using these four items to represent this attitude across different groups. This involves testing a series of models for three different types of invariance, each with greater constraints than the previous, in order to determine whether the relationships between indicator variables (i.e., items on a survey) and latent constructs emerge similarly across groups (Vandenberg & Lance, 2000).

The first, configural invariance, involves specifying the same factor structure in the groups of interest separately while allowing all model parameters to be estimated freely between the groups. This essentially ensures that the specified factor structure fits in both groups. While this is not a direct test of measurement invariance, or an actual test, adequate fit of this MGCFA

is a prerequisite to testing for measurement invariance. Assuming configural invariance is established, more direct tests of measurement invariance can be performed. The second, metric invariance, compares the baseline to a model in which like-item factor loading estimates are constrained to be equal across groups of interest. These two models are then compared using a chi-square difference test (Thissen et al., 1988). Though fit is likely to be worse, given that the factor loading estimates are no longer optimized to fit these groups independently, this test serves to gauge whether fit is significantly worse. A non-significant chi-square indicates that the like-item factor loading estimates in the two groups are not significantly different from one another. In other words, latent variables are being defined similarly by both groups. The final test, scalar invariance, compares the previous model to a model where like-item intercepts are constrained to be equal across groups. A chi-square difference test comparing these two models allows the researcher to determine whether mean levels of the latent variables can be meaningfully compared across groups, with a non-significant chi-square indicating that mean scores for the same level of the latent variable are not significantly different from one another in the two groups. Studies have employed MGCFA to better understand diverse topics such as performance of organizational citizenship behaviors scores (OCB) across gender (Jang et al., 2022), personality test scores across internet and paper-and-pencil formats (Meade et al., 2007), and performance appraisal rating scores across rater groups (Facteau & Craig, 2001).

Another common method for assessing MI is based on the item response theory (IRT) framework termed differential item functioning (DIF; Drasgow, 1984). IRT DIF analysis provides more information than the MGCFA approach as it is an item-level analysis, and so, it can differentiate between items that do or do not exhibit MI (Lang & Tay, 2021; Tay et al., 2015). This specificity allows researchers and practitioners to exclude items that will attenuate the

validity of findings. IRT refers to a series of mathematical models used to estimate theta (i.e., one's standing on a latent variable) using the probabilistic relationship between that latent variable and parameters of an item on a given measure. Assumed under the IRT framework is monotonicity, where, as one's standing on an underlying variable increases, the probability of getting an item correct also increases (i.e., the conditional probability of response; Embretson & Reise, 2000). When thinking about polytomous data, such as when we are using ordered response options, the probability of choosing higher-ordered response option increases monotonically with theta. To illustrate, someone with higher theta levels of job satisfaction would have a higher probability of choosing "agree" than "disagree" on an item assessing job satisfaction.

The purpose of DIF, then, is to understand whether, on a unidimensional scale, something other than a given latent variable affects the conditional probability of response on an item (Embretson & Reise, 2000; Tay et al., 2015). To accomplish this, models in which like-item parameters are freely estimated across groups are compared to constrained models in which like-item parameters are constrained to be equal across groups (Meade & Wright, 2012). Like the above MGCFA approach, chi-square difference tests are used to assess whether DIF is present for a particular item across groups. When DIF is present, some trait-irrelevant variance shared within-groups is being captured by the item, and as such, it should not be used when making direct comparisons.

The item parameters of interest to this study item discrimination (*a* parameters) and item difficulty or item location (*b* parameters). Item discrimination can be conceptualized as the slope of the item and signifies how rapidly the conditional probability changes with the theta. For dichotomous items, difficulty corresponds to the theta level at which an individual has a 50%

chance of correct response. For instance, if an item has a *b* parameter of 1, 50% of individuals with a theta level of 1 on a given latent trait would be expected to endorse that item. For polytomous data, the number of *b* parameters is one less than the number of response options for a given item and represents the theta level needed to respond above a certain threshold with .50 probability (Embretson & Reise, 2000). A given item can display *a* parameter DIF, *b* parameter DIF, or both. Additionally, there can be uniform DIF (usually when there is *b* parameter DIF) or non-uniform DIF (usually when there is *a* parameter DIF). Uniform DIF is the term used to describe DIF that always favors one group over another while non-uniform DIF is the term used to describe DIF that favors different groups at different theta levels (Embretson & Reise, 2000).

The IRT framework has been applied to understand item functioning in areas such as job satisfaction scores across computerized and paper-and-pencil administrations (Donovan et al., 2000), burnout and emotional exhaustion scores across gender (Mitropoulou & Zampetakis, 2022), and workaholism scores across gender (Beiler-May et al., 2017). Findings from such studies can guide researchers and practitioners on whether groups can be meaningfully compared, the appropriate scales and items to use, and the conditions under which to employ a particular scale or format (e.g., Hammack-Brown, 2018; Facteau & Craig, 2001; Jang et al., 2022; Meade et al., 2007)— all of which elevate the utility of findings.

As noted above, none of the studies mentioned thus far reported testing for MI before analyzing mean differences on job satisfaction, autonomy, or social support measures across remote and in-person workers. That is not to say that measures of these constructs have not been tested for MI in any context. For example, researchers have examined MI on a job satisfaction measure across different cultural groups (Liu et al., 2004) and work contexts (Watson et al., 2007), a measure of autonomy across cultural groups (Brien et al., 2012), and gender (Izquierdo

& Pérez, 2022), the relationship between autonomy and job satisfaction across gender (Hornung & Glaser, 2009), as well as a social support measure across gender (Izquierdo & Pérez, 2022). MI was supported in all but one of these studies. Watson and colleagues (2007) found DIF in just under half (14 of 32) of the items on the Job Satisfaction Survey (JSS; Spector, 1985)— a 36 item, nine-facet scale of job satisfaction— between patrol officers and administrative officers. They concluded that interpretability of job satisfaction scores using the JSS to compare employees across these contexts may suffer. They caution against the use of the JSS across work contexts until evidence of MI is found. To restate, investigations of this nature provide guidance to researchers and practitioners regarding the appropriateness of scales and the functionality of findings. With this, future researchers and practitioners can adjust if necessary and proceed with higher confidence in their results.

Additionally, though none of the above studies mentioned testing for MI between remote and in-person workers before analyzing mean differences in constructs of interest, research on MI between these groups has been conducted on other constructs. In four such studies, MI was tested across these groups on a perceptions of coaching behaviors scale (Hammack-Brown, 2018), a technostress scale (Molino et al., 2020), a perception of safety at work during the COVID-19 pandemic scale (Converso et al., 2021), and the relationship between autonomy and cross-role interrupting behaviors (i.e., interrupting a work role to attend to non-work demands and vice versa; Santarpia et al., 2021). Of these, measurement non-invariance was found by Santarpia and colleagues (2021). Their results revealed that the path from job autonomy using the autonomy scale of the Work Design Questionnaire (WDQ; Morgeson & Humphrey, 2006) and work to non-work interrupting behaviors using a scale adapted by Kossek and colleagues (2012) was nonsignificant in the remote worker group but significant in the in-person worker

group. Unfortunately, the authors did not specify which of these two scales exhibited measurement non-invariance, so it is not possible to conclude which measure was the source of non-invariance. Thus, it can only be concluded that the WDQ scale of autonomy has potentially been found to exhibit non-invariance between remote and in-person workers in the past. If this is the case, mean group comparisons using this scale would be inappropriate.

Remote work has become much more common in recent years (Gifford, 2022; Global Workplace Analytics, 2020). As such, understanding the "truly different" (Brunelle & Fortin, 2021, p. 9) experiences of remote and in-person workers is as pertinent as ever. Organizational researchers and practitioners rely on survey measures to better understand variables such as employee behaviors, traits, attitudes, and perceptions (Lang & Tay, 2021). Relatedly, they rely on the results of these surveys to make decisions. For instance, the finding that remote workers have overall higher job satisfaction than in-person workers would seem to indicate that some intervention may be needed to better understand why this is the case and how job satisfaction of in-person workers can be increased to match the levels of remote workers. However, these findings are only valid if survey measures used to draw conclusions between these groups exhibit MI. For instance, say the satisfaction facet of satisfaction with one's supervisor is being used to make comparisons between remote and in-person workers. Let's assume this construct has a stronger relationship (i.e., factor loading) with an item measuring perceived competence of the supervisor for remote workers, whereas it has a stronger relationship with an item measuring how fair the supervisor treats others for in-person workers. It would be inappropriate to make direct comparisons between these groups on satisfaction with one's supervisor because how this construct is being defined differs between Thus, mean level comparisons are not as informative.

Present Study: Measurement Invariance across Remote and In-Person Workers

As the above review highlights, the experiences of remote and in-person workers on job satisfaction, autonomy, and social support are clearly of interest. Unfortunately, much of previous research has focused on mean differences between these groups without explicit examination of whether the scales being used are assessing the same constructs. Thus, this study aims to contribute to the research literature by assessing MI in job satisfaction, perceptions of autonomy, and social support across remote and in-person workers. In doing so, this study will investigate whether measures commonly used to assess these variables are comparable across these increasingly similarly sized worker groups (i.e., remote and in-person workers).

While this study will only investigate this phenomenon using one measure for each of these constructs, it will serve as an impetus for future research while also providing insight into the newer, albeit increasingly common and sought after, working arrangement of exclusively remote workers (Dart, 2022; Ozimek, 2020). Additionally, it will allow researchers to better understand whether the relationships found in their respective studies are real or spurious, as well as provide practitioners with more direction when analyzing survey results. Given that surveys are the central method for understanding employee experiences (Lang & Tay, 2021), this study is an important next step.

Worker Groups

The focus of the above, the comparison between remote and in-person workers, will specifically focus on comparing those who work exclusively remotely compared to workers who travel to the office all or a portion of the time (i.e., in-person or hybrid workers, respectively). Hybrid work is certainly on the rise, having increased 13.3% from 2018 to 2022 (U.S. Bureau of Labor Statistics, 2022), and is even sometimes deemed “the future of work” (Hilberath et al.,

2020). However, the experience of exclusively remote workers is likely starker of a contrast than that of hybrid and in-person workers.

Exclusively remote workers must rely solely on ICTs to communicate, collaborate, and build relationships with their coworkers and supervisors (Bentley et al., 2016). The experience of "engaging online is totally different than being present and interacting with [others]" (Knight et al., 2022, p. 6), contrasting the experience of someone who works exclusively remote to those who travel to the office at least a portion of the time. Additionally, the experience of organizational climate and culture is likely much more similar for hybrid and in-person workers than either of these groups and exclusively remote workers.

Organizational climate is the meaning employees derive from their experiences, whereas culture is a function of the values and beliefs that go into creating what an employee experiences (Schneider et al., 2011). Described in more detail, "culture-embedding mechanisms are visible artifacts [relics such as objects, works of art, and uniforms] of the emerging culture and they directly create what would typically be called the 'climate' of the organization (Schneider, 1990; Ashkanasy, et al., 2000)" (p. 246). This suggests that the "meaning employees derive from their experience" (Schneider et al., 2011, p. 390) is likely heavily influenced by the physical environment of being at a company office with other colleagues. Exclusively remote work eliminates the visible and tangible aspects of the organization (e.g., offices, collocated employees) and instead forces employees to rely primarily on their psychological experience to represent their organization (Wiesenfeld et al., 2001). Therefore, there are likely more pronounced differences in the experience of those who work exclusively remote and those who work in the office at least a portion of the time (Wang et al., 2021).

The practice and attraction of exclusively remote work, while a much newer concept, appears to be on the rise. A 2022 Gallup poll revealed that one-third (34%) of remote-capable employees prefer exclusively remote work (Dart, 2022). Additionally, more than half (60%) of exclusively remote employees claimed that they would be "extremely likely" to leave their organization if long-term remote options were not offered (Dart, 2022). This trend is only expected to increase, with experts anticipating full-time remote work to grow by 65% in the next five years (Ozimek, 2020). As such, this study will focus investigating MI across fully remote employees and those that are in-person for at least a portion of their work week.

Job Satisfaction

Belonging to a particular worker group (i.e., remote or in-person) has been found to influence job satisfaction (e.g., Brunelle & Fortin, 2021; Fonner & Roloff, 2010; Dubrin, 1991; Gajendran & Harrison, 2007; Schall, 2019; Wheatley, 2012). Remote work researchers have described remote workers as "... [having] autonomy over work location and time…" (Sparrow, 2000, p. 2014), which may impact their perceptions of job satisfaction (Morganson et al., 2010). Others have theorized that simply working away from a physical office can result in changed employee attitudes and perceptions (Feldman & Gainey, 1997). Though differences in job satisfaction levels across remote and in-person workers do not necessarily indicate measurement non-invariance, it is plausible that these different work experiences alter the conceptualization of job satisfaction between them.

If this is confirmed and measurement non-invariance is established, the ability to meaningfully compare job satisfaction scores across remote and in-person workers may be impeded. Potential evidence of measurement non-invariance between these groups was revealed in a study by Brunelle and Fortin (2021). Specifically, they found different patterns of

relationships between the predictors need for autonomy, need for competence, and need for relatedness and the outcome variable of job satisfaction across in-person workers and remote workers. While the focus of the present study is not on inter-variable relationships, and further, MI was not tested on these scales prior to the authors concluding that "[remote] and [in-person] workers are satisfied in different ways" (Brunelle & Fortin, 2021; p. 9), it indicates that the composition of job satisfaction may differ in these two worker groups. Indeed, previous research has confirmed measurement non-invariance in job satisfaction using the JSS (Spector, 1985) across patrol officers and administrative officers, signaling that differences in work context influence the relationship between job satisfaction and observed scores on certain facets of this scale (Watson et al., 2007).

In this study, the JSS (Spector, 1985) will be used to investigate measurement invariance across remote and in-person workers on job satisfaction. Since 2019, this scale has been cited roughly 1,300 times, indicating it is still being used with frequency. The JSS comprises nine facets of job satisfaction, each assessed using four items. The nine facets are: pay, promotion, supervision, fringe benefits, contingent rewards, operating conditions, coworkers, nature of work, communication.

The upsurge in remote work trends violates the previously conceived assumption that employment relationships are bound by physical confines (Minbaeva, 2020). There are various disparities made inevitable by the experiences of remote versus in-person workers. These include aspects such as physical location and presence, frequency of face-to-face interactions, and communication mediums (Harrington & Santiago, 2006). As such, job satisfaction may be susceptible to measurement non-invariance.

Research Question 1: Will measurement invariance hold for the JSS across remote and in-person workers?

Autonomy

Deemed "the most widely studied work characteristic" (Morgeson & Humphrey, 2006, p. 1323), autonomy was established as a critical component for work motivation and satisfaction in the 1970s (i.e., Hackman & Oldham, 1976). More recently, it has been examined in the remote work literature quite extensively, being identified as a distinctive benefit of remote work given the physical and psychological separation it creates (Brunelle & Fortin, 2021; Gajendran & Harrison, 2007; Hornung & Glaser, 2009; Schall, 2019; Wang et al., 2021). Though the benefits of autonomy are not specific to remote workers, they have a unique experience with autonomy in that they have control over factors that are not plausible for in-person workers. These include things such as where to work (Wang et al., 2021), what to wear (Elsbach, 2003), when to begin and end their day (Golden & Veiga, 2005), as well as being able to adhere to personal demands more easily such as doing laundry or taking their children to school (Wang et al., 2021). Again, while these differences do not equate to measurement non-invariance, they may produce different conceptualizations of autonomy in these two groups.

Recent evidence appears to underscore the disparity between remote and in-person workers in relation to autonomy. Using the Work-Related Basic Need Satisfaction Scale (W-BNS; Van den Broeck et al., 2010), Brunelle and Fortin (2021) found that the relationship between the need for autonomy and job satisfaction differed depending on which worker group (remote or in-person) one was in. Specifically, the link between the satisfaction of the need for autonomy and job satisfaction was significantly greater for in-person workers, suggesting that satisfaction of the need for autonomy does not relate to job satisfaction in the same way for these

worker groups. Unfortunately, given that MI was not investigated, it is difficult to interpret these results directly.

Using the Work Design Questionnaire (WDQ; Morgeson & Humphrey, 2006) Santarpia and colleagues (2021) found non-invariant relationships between autonomy, work to non-work interrupting behaviors, and work-family conflict across remote and in-person workers. They concluded that autonomy "differentially relates to employees' work-family conflict when engaged in [in-person] and remote working arrangements" (Santarpia et al., 2021; p. 17). Though not all the above variables are of interest to this study, and the focus here is not in inter-variable relationships, these findings may indicate differences in how autonomy is perceived across remote and in-person workers. Unfortunately, however, the authors did not specify which of these scales exhibited measurement non-invariance, so it is not possible to conclude which measure was the source of non-invariance. Thus, it can only be concluded that the WDQ scale of autonomy has potentially been found to exhibit non-invariance between remote and in-person workers in the past.

If this is the case, these results as they relate to autonomy are less interpretable. This would potentially make the comparability of autonomy scores across remote and in-person workers faultier and the results less defensible. Given the recent increase in remote work (Dart, 2022) and the importance of autonomy in the study of work (e.g., Hackman & Oldham, 1976; Gajendran & Harrison, 2007; Wang et al., 2021), this is especially pertinent to understand. In this study, the Work Design Questionnaire (WDQ; Morgeson & Humphrey, 2006) will be used to test for MI in autonomy between remote and in-person workers. Since 2019, the scale development paper for the WDQ has been cited 1,600 times, indicating it is still being used with frequency and is appropriate to use.

Research Question 2: Will measurement invariance hold for the autonomy scale of the WDQ across remote and in-person workers?

Social Support

Social support is one of the most widely studied constructs in the organizational literature (French et al., 2018). Social support involves the opportunity to gain advice and assistance at work, develop friendships, and meet others at work (Morgeson & Humphrey, 2006). Given the physical separation of remote workers, ICTs become the means through which these events must occur for these workers, while in-person workers can continue to engage in face-to-face interactions (Allen et al., 2015). In other words, obtaining social support materializes differently for remote and in-person workers. Given the above discussion on social support and the conceptual issues surrounding this construct in the literature (Jolly et al., 2021), the following will be investigated using the social support measure from the WDQ (Morgeson & Humphrey, 2006):

Research Question 3: Will measurement invariance hold for the social support scale of the WDQ across remote and in-person workers?

METHOD

Sample

Data for this study were collected through the Prolific web-based research platform. The total dataset included 560 participants. To ensure the intended population was sampled, the following Prolific prescreens were used: (1) at least 18 years old, (2) fluency in the English language, (3) residing in the United States, and (4) employment status of Full-Time. Participants were paid $1.70 for their participation. Based on recommendations from Meade and Craig (2012), two careless responding screens, in the form of instructed-response items (e.g., “If you

are reading this, please select Agree"), were included. These careless responding checks were failed by four participants (0.7%), who were removed from the sample, resulting in a final sample of 556 respondents. Across this sample, the mean age was 38.31 (SD = 10.29) and the gender breakdown was 60% men, 39% women, and 1% non-binary. Additionally, the racial breakdown was as follows: 80% White/Caucasian, 7% Asian/Asian-American/Pacific Island, 7% Hispanic or Latino, 5% Black/African American, 1% Other, and <1% American Indian or Alaskan Native.

Procedure

Interested Prolific participants self-selected into the survey based on the title, a brief description of the survey, expected time to complete, and compensation amount. Those who chose to participate were directed to Qualtrics where the surveys were housed. After providing informed consent, participants responded to an initial demographics questionnaire to identify their age, race, gender, and the number of days they spend in the same physical location as others in their organization during a typical work week. To get a deeper understanding of their working situation, participants were asked to indicate whether they have been in their current working situation (i.e., remote, hybrid, or in-person) for more or less than one year, if this working situation matches their preference, if they live alone or with others, as well as whether they perceive their employer to use some form of electronic monitoring. Finally, participants responded to three surveys to measure their job satisfaction, perceptions of autonomy, and perceptions of social support at work. After completing these surveys, participants were debriefed and rerouted to Prolific to complete their submission.

Measures

Worker Groups

Worker group (remote or in-person) was identified in two ways. Prolific has a prescreen question titled “Remote/office work” with options such as: “I always work from a central place of work”, “I sometimes work from a central place of work and sometimes remotely” and “I always work remotely.” Per Prolific protocol, this study was duplicated such that one version was only available to those who selected “I always work from a central place of work” or “I sometimes work from a central place of work and sometimes remotely” and the other version was only available to those who selected “I always work remotely.” As it was unclear how up-to-date these responses were, this question was cross-referenced in Qualtrics by asking: “During a typical work week, how many days do you spend in the same physical location as others in your organization?” with response options of “0 days”, “1 day”, “2 days”, “3 days”, “4 days”, “5 days”, or “More than 5 days.” While responses should be consistent across platforms, the Qualtrics response was used in inconsistent cases given it is a more recent account of which worker group one belongs to. For this study, remote workers were defined as employees who selected the response “0 days.” All others were classified as in-person workers. This resulted in a breakdown of 247 remote workers (44%) and 309 in-person workers (56%).

Job Satisfaction

Job satisfaction was measured using the 36 item Job Satisfaction Survey (JSS), which has been found to demonstrate expected patterns of convergent and discriminant validity (Spector, 1985). The JSS intends to measure nine facets of job satisfaction: pay, promotion, supervision, fringe benefits, contingent rewards, operating conditions, coworkers, nature of work, and communication. Sample items include: “I feel I am being paid a fair amount for the work I do,” “I am not satisfied with the benefits I receive,” “I like the people I work with,” and “I like doing the things I do at work.” Given that IRT assumes unidimensionality, an EFA was conducted to

assess whether this scale could be sufficiently represented by one factor. Results of the EFA (described in further detail in the Results section) supported a one factor solution; thus, this scale was treated as a global job satisfaction measure (alpha = .96).

In the original version of this scale, participants are instructed to "Please circle the one number for each question that comes closest to reflecting [their] opinion about it" with choices ranging from 1 = "Disagree very much" to 6= "Agree very much." This was slightly adapted to reflect that this is an online questionnaire (i.e., circling is not an option), to reflect the more common Likert agreement scale options, and to include a neutral option. Additionally, one item is somewhat outdated, reading: "I have too much paperwork" and was deleted. In this study, this measure instructed participants to indicate their agreement with the 35 statements regarding their current job and was scored from 1= "Strongly disagree" to 5= "Strongly agree."

Autonomy

Perceived autonomy was measured using the validated nine-item autonomy scale within the WDQ (Morgeson & Humphrey, 2006). This scale was designed to measure three facets of autonomy: work scheduling autonomy, decision making autonomy, and work methods autonomy. Sample items include: "I am allowed to make my own decisions about how to schedule my work," "The job provides me with significant autonomy in making decisions," and "I am allowed to make decisions about what methods I use to complete my work." Again, given that IRT assumes unidimensionality, an EFA was conducted to assess whether this scale could be sufficiently represented by one factor. Results of the EFA (described in further detail in the Results section) supported a one factor solution; thus, this scale was treated as one autonomy measure (alpha = .96). This measure instructed participants to indicate their agreement with the nine statements regarding their perceptions of autonomy at their current job and was scored from 1= "Strongly disagree" to 5= "Strongly agree."

Social Support

Perceived social support was measured using the validated six-item social support scale within the WDQ (Morgeson & Humphrey, 2006; alpha = .89). Sample items include: “I have the opportunity to develop close friends in my job” and “People I work with are friendly.” This measure instructed participants to indicate their agreement with the six statements regarding their perceptions of social support at their current job and was scored on a five-point scale from 1= “Strongly disagree” to 5= “Strongly agree.”

Analyses for Examining Psychometric Properties

Unidimensionality Analyses

All data analyses were conducted using R software (R Core Team, 2022). IRT analyses rest on the assumption that the construct undergoing DIF testing is unidimensional. Thus, the dimensionality of each scale was first assessed using a parallel analysis (PA; Horn, 1965) to determine the upper bound for the optimal number of factors to retain for each scale. PA accounts for the issues sampling error by constructing random data with the same dimensions as the observed data (i.e., same number of variables and sample size) and comparing the eigenvalues from both datasets. The number of eigenvalues in the observed data that are higher than those derived from the random data serves as the number of suggested factors to be retained (Hayton et al., 2004). Given this, the PA served as an upper bound, such that the number of factors retained did not exceed the number specified by the PA results. A scree plot was also observed for a visual depiction of the number of factors present. Exploratory factor analysis (EFA) with a maximum likelihood extraction method and oblique rotation was performed using the “psych” R package (Revelle, 2021) for each scale in each worker group, using a combination of the PA results and the scree plot to determine the number of factors to

extract. Unidimensionality was assumed where a dominant first factor, accounting for at least 20% of the variance (Reckase, 1979), was present.

Item Parameter Estimation and Model Fit

To model the polytomous data and estimate item parameters, the graded response model (GRM; Samejima, 1969) was employed using the "mirt" package (Chalmers, 2012) in R software (R Core Team, 2022). The GRM assumes that, for a positively worded item, higher standing on theta will increase the likelihood of an ordinally higher response to an item. For example, someone with a higher theta on job satisfaction will have a higher probability of choosing "Agree" than "Neither agree nor disagree" on an item assessing job satisfaction.

In the GRM, each item is represented by one *a* parameter, whereas the number of *b* parameters is equal to one minus the number of response options for the given item (Embretson & Reise, 2000). So, for an item with five response options, four *b* parameters are estimated. The *a* parameter corresponds to slope and is termed "item discrimination", signifying how rapidly the probability of response changes with theta level. An item with a high *a* parameter indicates that the item is more sensitive to differences in theta levels, as it has a large impact on the probability of response. The *b* parameter corresponds to the location of a between category threshold and represents the theta level necessary to respond above a certain threshold (i.e., response option) with .50 probability.

Estimating these threshold parameters is a two-step process. First, an operating characteristic curve (OCC) is calculated for each threshold parameter using the 2-parameter logistic model (2PL; see Embretson and Reise, 2000 for formal equations). These OCCs represent the probability of a given individual's response "falling in or above a given category threshold conditional on [theta]" (Embretson & Reise, 2000, p. 98). Next, category response curves (CRCs) are calculated by subtracting the probability of endorsing a given response by

the probability of endorsing the next highest response option, with the number of CRCs being equal to the number of response options for a given item. Fit was assessed using the "mirt" package (Chalmers, 2012) to understand whether the GRM was an appropriate model for the data. Fit was examined using the item-level fit index S-X^2, an appropriate index for psychological research (Kang & Chen, 2007; Orlando & Thissen, 2003; Tay et al., 2014).

Research Question Analyses

Once item parameters were estimated across groups, direct tests of MI were conducted to examine Research Questions 1, 2, and 3. An important step in investigating DIF is to identify items to serve as anchors, or non-differential functioning (DF) items, for which to compare nested models. Initially, a free baseline model was specified for each unidimensional scale with all item parameters estimated freely (i.e., taking on their optimal parameter estimate values), which is simply used to examine item parameters in both worker groups separately. The All Others As-Anchors Method (AOAA; Meade & Wright, 2012; Lopez Rivas et al., 2009) was then employed and a chi-square difference test was performed between two nested models.

Specifically, a constrained baseline model, where all like-item parameters are forced to be equal across the two worker groups, was compared to an augmented model, where the software performs iterations such that item parameters are freed up for one item at a time. Items that have a significant chi-square statistic, signaling that allowing parameters to differ across groups will significantly improve model fit, were flagged as potential DF items. However, a drawback to AOAA approach is that for a noninvariant scale, DF items will be included among the set of anchor items. When conducting invariance analyses, choosing the correct anchor items is essential, as failure to identify invariant anchor items can lead to inflated Type I and Type II errors (Johnson et al., 2009; Lopez Rivas et al., 2009). Thus, the AOAA method served

to identify anchor items (i.e., those with high *a* parameters and a nonsignificant chi square) for use in additional DIF testing (Lopez Rivas et al., 2009; Meade & Wright, 2012), again using a chi-square difference test. In this additional method, the constrained baseline model is specified such that all chosen anchor items' parameters are constrained to be equal while the rest are estimated freely. An augmented model then constrains parameters for one non-anchor item at a time and compares it back to the baseline model. Items with a significant chi-square value are then identified as DF items.

For Research Question 1, examining whether MI would hold for the JSS (Spector, 1985) across remote and in-person workers, the MaxA5 method was utilized (i.e., 5 invariant items with the highest *a* parameters; Meade & Wright, 2012) for identification of non-DF anchor items. For Research Question 2 and Research Question 3, examining whether MI would hold for the smaller scales measuring autonomy and social support across remote and in-person workers, procedures by Meade and Wright (2012) were followed, who suggested 20-25% of the total number of items on a given scale as suitable when determining the number of anchors to select. Thus, two items were selected as anchors for the WDQ autonomy scale (Morgeson & Humphrey, 2006) using the same criteria as above (i.e., two invariant items with the highest *a* parameters). For anchor selection using the WDQ social support scale (Morgeson & Humphrey, 2006), this would result in only one anchor item being chosen, which can make for an underpowered test. On the other hand, inclusion of a potential non-invariant anchor would increase the likelihood of a Type I error. Thus, the above procedure was implemented twice for social support, once with one anchor item being selected and again with two anchor items being selected, with anchor items being chosen using the same criteria as the other scales. Additional iterations of anchor selection for autonomy were performed in cases where misfitting items

were excluded from analyses; the specific details of anchor selection in these cases are provided in the results section.

Finally, effect sizes of the expected scores for the referent (in-person workers) and focal (remote workers) groups were used to assess the extent of DF present in items flagged as containing significant DF using the above procedures (Meade, 2010). Expected scores for members in the focal group (i.e., remote workers) were computed using the focal group's theta and item parameters for a given item in that group. For the referent group (i.e., in-person workers), expected scores were computed using the focal group's theta and the referent group's item parameters. Thus, for a given member of the referent group, the expected score represents what their expected score would have been had they been a member of the focal group. Effect sizes of these expected scores, then, provide different ways to understand the impact of the DF.

The reported effect sizes at the item and scale level will include signed difference of the sample (SIDS), unsigned difference of the sample (UIDS), expected score standardized difference (ESSD), and expected test score standardized difference (ETSSD; Meade & Wright, 2012). SIDS and UIDS are easy to interpret given that they are on the metric of the scale being used, and so they can be directly interpreted as the amount of DF that would be expected on a given item or scale (ranging from 1-5 in this study, given all measures are scored on a 5-point scale). SIDS represents the average of the difference in observed scores between groups for a given item or scale while UIDS represents the absolute value of this average difference. A positively signed SIDS indicates that, given equal standing on the latent trait, remote workers are expected to have a higher observed score than in-person workers. While this is a convenient way to think about DF in the context of the scale being using, it also means that these metrics are not inherently meaningful across scales with differing numbers of response options. ESSD and ETSSD, which can be interpreted like Cohen's *d* statistic (Cohen, 1988), provide a

standardized metric for evaluating how big of an effect the DF is expected to have across groups of interest for a given item or scale. Thus, the ESSD and ETSSD will be the focus of discussion on effect sizes. According to guidelines set out by Cohen (1988), an effect size below .20 is classified as a trivial effect, an effect size of 0.20 is small, an effect size of .50 is moderate, and .80 is a large effect.

RESULTS

Descriptive statistics for the whole sample are provided in Table 1. Additionally, bivariate correlations for the whole sample, remote workers, and in-person workers are shown in Table 1.

Unidimensionality Analyses

Following recommendations that IRT's assumption of unidimensionality can be considered met when a first factor accounts for at least 20% of the variance in a construct (Reckase, 1979), each scale was considered unidimensional. Using a combination of the PA results and the scree plot, an EFA extracting three factors was run for remote and in-person workers on the JSS (Spector, 1985). In both groups, a dominant first factor, accounting for 25% of the variance in both cases, was present. It should be noted, however, that two items had factor loadings lower than .35 and were removed from further analysis (Hair et al., 1998). This included JS15 ("My efforts to do a good job are seldom blocked by red tape") and JS24 ("I have too much to do at work"). With these removed, factor loadings ranged from .36 to .92 in the remote group and .35 to .90 in the in-person group. An additional EFA was run extracting one factor in both groups. Results revealed that 41% and 42% of the was variance explained by one factor in the remote and in-person worker groups, respectively. Factor loadings ranged from .40 to .80 in the remote group and .54 to .80 in the in-person group, again without the items identified above.

For the WDQ autonomy scale (Morgeson & Humphrey, 2006), results of the PA and visual inspection of the scree plot suggested one factor be extracted, with 74% and 70% of the variance explained by this factor in the remote and in-person worker groups, respectively. Factor loadings ranged from .71 to .92 in the remote group and .70 to .91 in the in-person group. For the WDQ social support scale (Morgeson & Humphrey, 2006), results of the PA and visual inspection of the scree plot suggested one factor be extracted, with 60% and 59% of the variance explained by this factor in the remote and in-person worker groups, respectively. Factor loadings ranged from .47 to .93 in the remote group and .56 to .87 in the in-person group.

Model Fit

To ensure the GRM provided appropriate fit to the data, item-level fit was assessed using the S-X^2 statistic (Kang & Chen, 2007; Orlando & Thissen, 2003; Tay et al., 2014), as seen in Table 2, alongside item parameter estimates for each item. As can be observed, one item from the JSS (JS36) had a significant S-X^2 statistic, signaling misfit. However, conclusions made from analyses with and without this item included were identical. Thus, results will be discussed for models using all items, apart from those removed due to low loadings (JS15 and JS24). Two and three items from the WDQ autonomy scale and WDQ social support scale (Morgeson & Humphrey, 2006), respectively, showed a significant S-X^2 statistic. Given analyses for these scales with and without misfitting items according to the S-X^2 were somewhat discrepant, results of both approaches will be presented.

Research Question Analyses

The following sections are organized by construct per the numerical order of the research questions. The item content for each item identified as containing significant DF is displayed in Table 3. Expected scores and effect sizes for all items identified as DF items in the

non-AOAA method, as well as test level effect sizes, are presented in Table 4. Finally, Figures 1-3 show expected score curves for the JSS items, the WDQ autonomy scale items, and the WDQ social support items, respectively.

Job Satisfaction

Research Question 1 asked whether MI would hold for the JSS (Spector, 1985) across remote and in-person workers. Using the AOAA method, 10 items were initially flagged as containing significant DF (i.e., had a significant chi-square). Using the MaxA5 method, items JS5, JS14, JS19, JS23, and JS32 were identified as non-DF anchor items for further DIF testing. These 10 items were still identified as DF items using the MaxA5 method. Further, an additional item (JS35) was identified as a DF item. This resulted in a total of 11 of 33 JSS items (33%) being non-invariant across remote and in-person workers, indicating employees from these worker groups are interpreting these items differently.

All four items regarding satisfaction with coworkers (JS7, JS16, JS25, JS34) were found to be DF items. Specifically, JS7 (“I like the people I work with”) and JS25 (“I enjoy my coworkers”) exhibit *b* parameter DF, whereas JS16 (“I find I have to work harder at my job because of the incompetence of people I work with”) and JS34 (“There is too much bickering and fighting at work”) exhibit both *a* and *b* parameter DF.

For JS7, the *b* parameter DF was such that, given the same level of theta, remote workers were less likely to endorse the most extreme disagreement (i.e., “Strongly disagree”) at lower levels of theta, whereas at moderate to high levels of theta, they were less likely to endorse the most extreme agreement (“Strongly agree”). In other words, for both low and high theta levels, remote workers appeared less likely to endorse the more extreme response option than in person workers (i.e., remote workers had a higher probability of selecting “Disagree” vs. “Strongly disagree” and “Agree” vs. “Strongly agree”). This suggests that, for workers with lower levels of

theta on liking their coworkers, remote workers would be less likely to indicate extreme dissatisfaction, as well as being less likely to indicate extreme satisfaction at higher levels of theta on liking their coworkers. For JS25, the pattern of the *b* parameter DF is somewhat similar; remote workers were less likely to endorse the more extreme response option at moderate to high levels of theta.

Items JS16 and JS34 exhibited both *a* and *b* parameter DIF. Like the above, the *b* parameter DIF was such that, for lower and higher theta levels, remote workers were less likely to endorse the more extreme response options. It should be noted that these items were reverse coded, and so, higher expected scores indicate disagreement with the item. This pattern was more pronounced for JS34, and most specifically at lower to moderate levels of theta, where remote employees were more likely to endorse the less extreme option (i.e., "Agree" over "Strongly agree"). For both items, the *a* parameter DF was such that this item was more strongly related to overall satisfaction levels for in-person workers than remote workers.

Three of the four items regarding satisfaction with supervisor (JS3 JS12, JS21) were found to be DF items. Specifically, JS3 ("My supervisor is quite competent in doing their job") exhibited *a* and *b* parameter DF, JS12 ("My supervisor is unfair to me") exhibited *a* parameter DF, and JS21 ("My supervisor shows too little interest in the feelings of subordinates") exhibited *b* parameter DF. For JS3, a similar pattern as the above items emerged, where, given equal theta levels, remote workers were less likely than their in-person counterparts to endorse the more extreme response option. For this same item, the *a* parameter DF was such that this item was more strongly related to overall satisfaction levels for in-person workers than remote workers. Similarly, JS12 exhibited *a* parameter DF such that responses to this item were more strongly related to satisfaction levels for in-person workers than remote workers. Lastly, item JS21

(reverse coded) exhibited *b* parameter DF such that at low to moderate theta levels, remote workers were less likely to endorse the most extreme response option (i.e., remote workers had a higher probability of selecting "Agree" over "Strongly agree" or "Neither agree nor disagree" over "Agree").

Three of the four items regarding satisfaction with the nature of the work (JS8, JS17, and JS35) were identified as DF items. Specifically, all three items exhibited *a* and *b* parameter DF with similar patterns as those items described above. The nature of *b* parameter DF for all three items— JS8 ("I sometimes feel my job is meaningless"), JS17 ("I like doing the things I do at work"), and JS35 ("My job is enjoyable")— was such that remote workers were less likely to endorse the more extreme option at lower and higher levels of theta. Additionally, the *a* parameter DF was such that these items were more strongly related to overall satisfaction levels for in-person workers than for remote workers. Additionally, the *a* parameter DF again indicates that these items are more strongly related to overall satisfaction levels for in-person workers than for remote workers.

Finally, one item regarding satisfaction with communication was identified as a DF item (JS9, "Communication seems good within this organization"). This item exhibited both *a* and *b* parameter DF. As with the above items, the same patterns emerge such that the *b* parameter DF indicated remote workers being less likely to endorse the more extreme options given equal levels of theta. Relatedly, the *a* parameter DF was such that this item was more strongly associated with overall satisfaction levels for in-person workers than remote workers.

Using the above criteria for effect sizes of expected scores (Cohen, 1988; see Table 4), none of the JSS items had large effect sizes, with ESSD ranging from -.07 to .76. Five of the 11 DF items had a trivial effect size (less than .20; JS7, JS9, JS17, JS25, and JS35), five had a small

to medium effect size (greater than .20 but less than .50; JS3, JS8, JS12, JS16, and JS21), and one had a medium effect size (.76; JS34). As for test level effect size, the ETSSD (.00) suggested the impact of DF at the scale level was trivial. Thus, the impact of DF may differ depending on the specific items used and the use of the scale.

Autonomy

Research Question 2 asked whether MI would hold for the WDQ scale of autonomy (Morgeson & Humphrey, 2006) across remote and in-person workers. According to the S-X^2 statistic for assessing item-level fit on the sample, items Auto1 and Auto2 were identified as misfitting items. Thus, the AOAA method and the two-anchor item method were performed twice, once with all items and once without the misfitting items. With the former approach, items Auto6 and Auto8 were selected as non-DF anchors. One item, Auto1, was found to have significant DIF using both AOAA and the two-anchor item approach**.** In other words, one of the nine WDQ autonomy items was identified as being non-invariant across remote and in-person workers in the run with all items included, indicating that this item does not function the same across worker groups.

In the latter approach without items Auto1 and Auto2, the two-anchor item method was employed again, resulting in two of the seven remaining items (29%) being selected as anchors. Since this is higher than suggestions from Meade and Wright (2012), the same analysis was performed with only one anchor, in effort to balance out the issues of selecting a potential DF anchor item on inflating Type I error rates. Specifically, Auto6 and Auto8 were chosen as anchors for the first iteration, followed by two additional runs such that both Auto6 and Auto8 served as non-DF anchors. All results were identical, with none of the items identified as non-invariant across remote and in-person workers in the run without misfitting items. Thus, only one

item (Auto1) on the nine item WDQ autonomy scale (11%) was identified as being non-invariant across remote and in-person workers, indicating this item does not function the same across worker groups.

More specifically, Auto1 ("The job allows me to make my own decisions about how to schedule my work") exhibited *b* parameter DF such that, for a given theta, remote workers were less likely to endorse the more extreme response option at lower and moderate levels of theta, becoming slightly more likely at higher levels of theta. Regarding effect sizes, the amount of DF in item Auto1 was considered a small to medium effect at the item level (ESSD of .28). At the test level, the effect size (ETSSD of .06) was trivial. Again, this suggests that the impact of DF may be dependent upon the actual use of the scale.

Social Support

Research Question 3 asked whether MI would hold for the social support scale of the WDQ (Morgeson & Humphrey, 2006) across remote and in-person workers. According to the S-X^2 statistic for assessing item-level fit on the sample, items Social1, Social4, and Social6 were identified as misfitting. Thus, the same approach as with autonomy was employed; the AOAA method and the method with one or two (depending on the run) invariant items with the highest *a* parameters selected as non-DF anchor items were performed twice. In the run with all items included, the AOAA approach flagged items Social3, Social4, and Social6 as DF items. In the additional method with the two non-DF anchors (identified as Social1 and Social2), an additional item, Social5, was identified as a DF item. Thus, four of the six items (67%) were flagged as containing significant DF in the run with all items included, indicating that these items do not function the same across worker groups.

However, given the above two-anchor approach resulted in 33% of items being selected as anchors, which again, is higher than recommendations by Meade and Wright (2012) and increases the chances of a non-invariant item being chosen as an anchor, an additional run with only one anchor item (Social2) was performed. In this run, Social3, Social4, and Social6 were again identified as significant DF items. Social5 was marginally significant in both the AOAA method ($p = .077$) and the method with the one non-DF anchor item ($p = .062$). Thus, it appears that the one-anchor item method lacked sufficient power to detect the DF in Social5 that was caught in the previous method.

In the latter approach, items Social1, Social4, and Social6 were removed due to misfit, leaving only three items (Social2, Social3, and Social5) being tested for DIF. The results of the AOAA method indicated that item Social3 contained significant DF, which left only two items (Social2 and Social5) as candidates for the anchor item. Since, again, a one-anchor item test may be underpowered (i.e., subject to Type II errors), but using a potential DF item as an anchor can induce false positives (i.e., Type I error rates), additional chi-square difference tests were performed to allow both items to serve as the anchor to better understand the results. With this, Social3 was the only item identified as containing DF. This resulted in one of the three well-fitting items as being non-invariant across remote and in-person workers.

Given the evidence to suggest that detection of DF in Social5 was subject to Type II error in the one-anchor item run with all items included, the subsequent sections will be described as if Social5 is a significant DF item. In sum, then, items Social3, Social4, Social5, and Social6 were found to contain significant DF. This indicates that four items on the six-item WDQ social support scale (67%) were non-invariant across remote and in-person workers, indicating that these items do not function the same across worker groups.

Items Social3 (“I have the opportunity to meet with others in my work”), Social4 (“My supervisor is concerned about the welfare of the people that work for him/her”), and Social5 (“People I work with take a personal interest in me”) exhibited *a* and *b* parameter DF. For Social3, the *b* parameter DF was such that, for a lower to moderate levels of theta, remote workers were more likely to endorse the more extreme response option. Regarding *a* parameter DF, this item was more strongly related to overall social support levels for remote workers than in-person workers.

For Social4, the *b* parameter DF was similar in nature to the JSS items described above, where, for a given theta, remote workers were less likely to endorse the more extreme response option (i.e., remote workers had a higher probability of selecting “Agree” over “Strongly agree” or “Neither agree nor disagree” over “Agree”). For Social5, the *b* parameter DF was such that, for lower to moderate levels of theta, remote workers were less likely to endorse the more extreme option, but more likely to endorse the more extreme option at higher levels of theta. The *a* parameter DF for these two items was such that were more strongly related to overall social support levels for in-person workers than remote workers.

Social6 (“People I work with are friendly”) exhibited *b* parameter DF similar in nature to Social5. Specifically, for lower to moderate levels of theta, remote workers were less likely to endorse the more extreme option, but more likely to endorse the more extreme option at higher levels of theta. Regarding effect sizes, two of the four DF items had a small to medium effect size (-.23 and .25; Social3 and Social5) and two had a medium effect size (.71 and .59; Social4 and Social6) at the item level. The effect size was trivial at the test level (ETSSD of .15). This once again suggests that the impact of DF will be dependent on item or scale usage.

Supplemental Analyses

Examination of Mean Differences

To provide additional insight into the research questions, supplemental analyses were conducted. First, patterns of mean differences across remote and in-person workers were assessed. Non-invariance precludes meaningful comparisons of mean differences, suggesting that the groups of interest are not interpreting or responding to DF items in such a way as to make meaningful comparisons. Thus, mean differences are provided for all three scales in their entirety and after removal of DF items and were assessed using one-way ANOVAs (see Table 5).

Observed scores on the JSS (Spector, 1985) with the DF items included indicated that there were no significant mean differences between remote and in-person workers on job satisfaction using this scale. Upon removal of DF items, these mean differences were significant, with in-person workers reporting higher job satisfaction on the JSS. As there is evidence for a curvilinear relationship between extent of remote work and job satisfaction (Golden, 2006; Golden & Veiga, 2005; Virick et al., 2010), this was also assessed, first with the DF items included and again without DF items. In both cases, the pattern of results was consistent with the literature, showing an inverted U-shaped relationship between extent of remote work and job satisfaction, though the relationships were nonsignificant.

For autonomy, there were no significant mean differences between remote and in-person workers using the WDQ (Morgeson & Humphrey, 2006) with or without the DF items. Consistent with much of the literature (e.g., Brunelle & Fortin, 2021; Gajendran et al., 2015), mean observed scores were higher for remote workers on perceptions of autonomy. For social support, significant mean differences were found using the WDQ (Morgeson & Humphrey, 2006) with and without the four DF items, with the mean observed scores for in-person workers being significantly higher than the remote worker group in both cases. Additionally, when only

looking at mean differences for the item that was found to be non-invariant in both runs, that is, the model with all social items and the model without DF items that had misfit according to the S-X^2 statistic (i.e., Social3), significant differences were still evident. Table 5 shows means only for the run with all social items included. Further, the statistics shown for non-DF items includes any items flagged as containing significant DF during the non-AOAA method (i.e., removing items Social3, Social4, Social5, and Social6).

Additionally, questions regarding more specific aspects of each employee's situation were asked; these included duration of their current worker group status (i.e., remote, hybrid, or in-person), whether their preference matched this worker group status, whether they lived alone or with others, as well as if they perceived the use of electronic monitoring by their employer. Of these, a match between preference and worker group status was the only aspect found to be significantly different ($p < .001$) across remote and in-person workers. Results were such that 97% remote workers (defined as working "0 days" in the same physical location as others in their organization in a typical week) indicated their preference matched their current situation, whereas 69% of in-person workers (all others that did not choose "0 days") indicated the same.

Mediation Analysis

To further investigate the impact of making conclusions based on a scale(s) with non-invariant items, a mediation analysis was conducted. Specifically, the well-established indirect effect of autonomy on the relationship between the extent of remote work and job satisfaction (e.g., Gajendran & Harrison, 2007; Hornung & Glaser, 2009; Schall, 2019) with and without DF items was assessed. Thus, a continuous variable with seven levels was created for extent of remote work with 1 = "5 or more days" and 7 = "0 days." The Hayes PROCESS macro, model 4 (Hayes, 2018) was used to perform the mediation tests. The test with all items (including DF

items) showed the indirect effect of autonomy was significant (*b* = -.014, SE = .006, LLCI = -.026, ULCI = -.002), as the confidence interval did not contain zero. This is in line with much of the previous research described above (e.g., Gajendran & Harrison, 2007; Hornung & Glaser, 2009; Schall, 2019). In the test without DF items included, the indirect effect of autonomy was no longer significant (*b* = -.011, SE = .006, LLCI = -.024, ULCI = .001), with the confidence interval containing zero. This serves as an additional example of how conclusions can be misconstrued when using scales that contain DF items.

DISCUSSION

Remote work has seen an unprecedented increase since the COVID-19 pandemic (Dart, 2022; Ozimek, 2020; U.S. Bureau of Labor Statistics, 2022). A great deal of research has explored the differences between remote and in-person workers on job satisfaction, autonomy, and social support. Unfortunately, much of this research has assumed that the measures being employed function the same across these groups without explicitly testing that assumption. This study sought to empirically investigate MI across these two groups using IRT analyses to determine whether they can be meaningfully compared on job satisfaction, autonomy and social support using three measures: the JSS, the autonomy scale of the WDQ, and the social support scale of the WDQ. Results indicate that DF by worker group is present for each of these scales. Item level effect sizes range from trivial to medium and test level effect sizes are trivial. Overall, these results aid in our understanding of job satisfaction, autonomy, and social support across remote and in-person workers.

A pattern that seemed to emerge with many of the DF items (11 of 16) is that they tend to concern attitudes towards or perceptions of other organizational members (i.e., satisfaction with coworkers, satisfaction with supervisors, and perceptions of social support). The upsurge in remote work trends violates the previously conceived assumption that employment

relationships are bound by physical confines (Minbaeva, 2020). Certain differences are made inevitable by the nature of remote versus in-person work, such as physical location and presence, frequency of face-to-face interactions, and communication mediums (Harrington & Santiago, 2006), which may be at play in the DF of these items.

Specifically, items pertaining to supervisors may be conceptualized differently across remote and in-person workers given the idea that supervisors must develop different strategies for remote and in-person workers to attune to the different needs of these groups (Lautsch & Kossek, 2011). Items pertaining to coworkers and colleagues may exhibit measurement non-invariance for similar reasons. For instance, remote workers are "dependent upon technology…for coordinating activities and collaboration" with their colleagues (Bentley et al., 2016, p. 208), which may result in different drivers for how they perceive their interactions and relationships with coworkers across these groups. Indeed, research has described that reliance on ICTs versus face-to-face interactions reshapes interpersonal processes (Allen et al., 2015; Daft & Lengel, 1986) and impacts relationship development and maintenance (Brunelle & Fortin, 2021). The results of this study suggest that this may be a microcosm for DIF across remote and inperson workers. Additionally, remote and in-person workers differ in the range of behaviors and situations for which they are exposed to and able to observe in their respective supervisors and coworkers, which has the potential of making certain aspects more or less salient depending on which worker group one is in.

Results of the JSS (Spector, 1985) MI investigation suggest that comparisons between remote and in-person workers are potentially misleading, especially for items pertaining to satisfaction with coworkers, supervision, and nature of work. Roughly one-third of the items tested on this scale were found to contain significant DF. In most instances, the DF was such that remote workers appear less likely to select the more extreme options at both high and low

levels of theta (see Figure 1). This may be due to different thresholds or sensitization to factors that influence satisfaction. For instance, in-person workers may be more sensitized to the causes of their satisfaction given the physical cues and more tangible inputs to their satisfaction levels, whereas remote workers do not have the same level of physical interaction with these aspects. Indeed, research has described the power of organizational culture, climate, and symbols as shaping meaning and outcomes such as satisfaction (Schein, 1990). Physical manifestations or symbols help shape the organizational culture and climate and have significant consequences for how employees derive meaning from and evaluate their experiences (Rafaeli & Worline, 2000). Relatedly, these symbols become associated with internalized feelings and states such that the physical presence of these symbols can evoke certain feelings. As such, they may serve as vehicles through which latent trait levels are formed and reinforced. This implies that the meaning in-person employees extract from their experiences is entangled with their physical surroundings, such as being in the office and with other colleagues (Schneider et al., 2011). As remote workers are physically separated from their organizations and do not have the same level of exposure to symbols, their experiences and internalizations are less entangled. Thus, remote workers and in-person workers may systematically differ in their threshold or sensitization to factors that influence their satisfaction levels.

Further, in all but one case, items with significant *a* parameter DF were more strongly related to the respective latent traits for in-person workers than remote workers. This suggests that these items have higher salience to in-person workers. This is not entirely surprising, given many of the scales investigated were developed when remote work was much less prominent. This may signal the need for development of new items or measures that are similar in their salience across remote and in-person workers on this construct.

Results of the WDQ autonomy scale MI analysis suggest that meaningful comparisons between these two groups may be inappropriate where the interest is in work scheduling autonomy. The nature of DF for the work scheduling autonomy item reveals that, for equal theta levels, in-person workers are more likely to select the more extreme option at lower levels of theta and less likely to select the more extreme option at higher levels of theta. This finding may signal that in-person workers are hyper aware of their autonomy levels such that they are especially sensitive to instances where they are and are not able to make their own work scheduling decisions. For instance, SDT (Deci & Ryan, 1985) posits that autonomy is a basic psychological need, increasing motivation through satisfaction of an intrinsic need for self-determination. As the physical separation afforded by remote workers is obvious, in-person workers may be more sensitized to their need for autonomy being met, given the comparative autonomy remote workers enjoy in terms of flexibility of physical location, which is so visible. Thus, in-person workers may have a higher threshold for what constitutes a satisfactory level of autonomy, and specifically, the ability to make decisions about how to schedule their work.

Finally, results of the WDQ social scale MI analysis suggest that comparisons between these two groups using this scale may be unwarranted, as two-thirds of the items tested were identified as containing significant DF. The nature of DF for items on this scale suggest that remote workers may have different expectations of their organization to help orchestrate coworking meetups, whether this occurs at the onboarding stage or for general socialization purposes. Empirical research seems to reinforce these findings, with remote workers who had opportunities to "physically meet other people (both [remote] workers and [in-person] workers" (Brunelle & Fortin, 2021, p. 7) reported higher levels of their need for relatedness being satisfied (an aspect of which includes social support). While MI investigations were not carried

out in their study, this nonetheless suggests that opportunities to meet colleagues may be especially salient to remote workers' standing on social support.

Additionally, the physical separation of remote workers means their main source for gathering information on whether their supervisor is concerned with the welfare of others, whether people at work take personal interest in them, and whether people they work with are friendly, is through ICTs. In other words, the situations in which remote workers formulate their standing on social support are likely planned much of the time (e.g., set scheduled meetings or 1:1's with others). For in-person workers, however, the flow of communication and opportunities to sample these behaviors in others is more informal, given in-person employees are bound to the same physical space, at least for a portion of their workweek. As such, employees who see each other regularly have access to "relatively unrestricted channels of communication" (Fonner & Roloff, 2010), which may differentially impact how they perceive these items. Taken together, these findings have major implications for research and practice, especially given the rise in the practice of remote work.

Research Implications

This study has important research implications. First, this study provided an empirical investigation of MI across remote and in-person workers on three widely used, validated scales to measure popular organizational constructs, which much of previous literature investigating these variables has failed to consider (e.g., Brunelle & Fortin, 2021; Fonner & Roloff, 2010; Golden, 2006; Hornung & Glaser, 2009). Given the vast growth of remote work since the COVID-19 pandemic (Daft, 2022; U.S. Bureau of Labor Statistics, 2022), it is critical to understand whether comparisons being made using popular scales are warranted.

Results of this study suggest that making conclusions and studying relationships across remote and in-person workers using the JSS (Spector, 1985) and the WDQ autonomy scale

(Morgeson & Humphrey, 2006) may be problematic, with the meditation analysis results being incongruent when conducted with and without DF items. Specifically, with DF items included, autonomy was confirmed as a significant indirect effect in the relationship between extent of remote work and job satisfaction, which is consistent with past literature (Gajendran & Harrison, 2007; Hornung & Glaser, 2009; Schall, 2019). With the DF items removed for these scales, this relationship became inconsistent with past research, with autonomy no longer being a significant mediator. This suggests that our theories on how autonomy relates to extent of remote work and job satisfaction may need to be revisited, as the inequivalence of these scales alters how these variables relate to each other.

Interestingly, the findings regarding job satisfaction using the JSS (Spector, 1985) with no DF items run counter to much of previous research, which has found remote workers to have higher job satisfaction than in-person workers (e.g., Brunelle & Fortin, 2021; Golden & Gajendran, 2007; Schall, 2019; Taboroši et al., 2022; Virick et al., 2010), or found a curvilinear relationship between extent of remote work and job satisfaction (Golden, 2006; Golden & Veiga, 2005; Virick et al., 2010). However, it is unclear how valid these results are, as none of the authors mentioned performing MI testing prior to making comparisons. As this study classified remote workers as only those who perform exclusively remote work (e.g., spend 0 days during a typical work week in the same physical location as others in their organization), future research should continue to investigate satisfaction levels in this new group of workers to understand how their satisfaction levels compare to other worker groups. For instance, when looking at whether employees' preference matched their current worker group status, remote workers reported significantly higher levels of matching. This may be an especially salient feature for remote workers in respect to job attitude formation. Additionally, results from a recent study (Brunelle & Fortin; 2021) led authors to conclude that remote and in-person

workers are “satisfied in different ways” (p. 9), which, even though they do not mention performing MI tests, may serve as further evidence that different factors constitute satisfaction in these groups.

Practical Implications

There are also practical implications that should be noted. Results of this study suggest that practitioners should be more intentional in selecting scales or items for comparing remote and in-person. Particularly, extra caution should be taken where items are asking about other organizational members such as supervisors, colleagues, and coworkers. These items appear especially prone to differential functioning across remote and in-person workers. Alternatively, scales for these constructs that have already been found to be invariant across these groups should be employed.

Item level effect sizes (see Table 4) were small to medium for most items, insinuating that impact of DF may be minor. Even so, single-item measures have been gaining popularity in the organizational science literature, given long surveys suffer from issues such as high cognitive load, participant fatigue, careless responding, redundancy, and low retention rates (Matthews et al., 2022). Additionally, single-item measures, or at least reduced measures, may be more attractive to organizational practitioners who seek to gain insight on several constructs at once (e.g., job satisfaction, autonomy, social support, commitment, stress, turnover intentions, etc.). Thus, if one is unknowingly using only DF items to assess differences between these two worker groups, and make decisions or design interventions, the impact is likely to be more pronounced.

At the test level, effect sizes suggest the impact of DF is negligible, as these were trivial for all three scales (see Table 4). Regardless, there may still be reason for caution. In this study, the JSS and the WDQ scale of autonomy were investigated as global satisfaction and autonomy

measures, respectively. These measures, however, were intended to be facet-level measures. For practitioners using them to derive conclusions at the facet level rather than at the global level, the impact of DF will likely be much more problematic for facets with all or many DF items (i.e., satisfaction with supervision, satisfaction with coworkers, satisfaction with nature of work, and work scheduling autonomy), given that these scores would be based on a larger number of inequivalent items across groups. In other words, the implications of using these two scales are dependent on how the measures are being utilized. If they are being used as global measures, results would suggest it's reasonable to make mean comparisons across groups. If they are being used to derive separate scale scores, these two measures would not be the most desirable options for observing mean differences between remote and in-person workers.

The WDQ social support measure (Morgeson & Humphrey, 2007) is intended as one scale. Thus, test level effect size estimates would suggest that the amount of DF is trivial and would not be associated with considerable issues when making comparisons across remote and in-person workers. In other words, test level effect sizes seem to signal that using the WDQ for social support to make comparisons between remote and in-person workers is acceptable. However, as has been shown, that does not necessarily ensure that use this scale will be rid of erroneous conclusions.

Perhaps more importantly, examination of mean differences with and without DF items highlights how different conclusions can be reached when using nonequivalent versus equivalent measures. For the JSS, inclusion of DF items masked the significant differences in means across these two groups, which could result in vast misunderstandings when making decisions or interpreting results using this scale. For the WDQ scale for social support, patterns of mean differences with DF items included do not appear much different than the means without the DF items. However, the non-invariance on four of six of these items indicates that

these means are not representing the same thing across these groups. As such, while the actual mean values may not appear to be affected, it is the assumption that these measures are comparable and the resulting conclusions that would be problematic.

Limitations and Future Research

This study has several limitations which bear noting. First, although DF analyses for social support found four items that contained significant DF across remote and in-person workers, this was based on a model that included three misfitting items. Additionally, two of these misfitting items had significant DF. When looking more closely, it was found that Social4 fit in the in-person worker group but did not fit in the remote worker group. For Social6, this item fits in both groups independently, but does not fit in the overall sample. A similar situation arose for autonomy, where the only item found to have significant DF (Auto1) was a misfitting item. When evaluating fit in the worker groups separately, this item fit in the remote worker group, but not in the in-person group. This may simply be an indication that these items contain DF (Tennant et al., 2004), which was the case for 50% of the misfitting items, but the tests were too underpowered to detect it.

Relatedly, standard errors (SE) of parameter estimates for four of the DF items were above the desired value of .5 in one of the worker groups. Specifically, the SE for *b1* parameters on items JS12, JS25, JS34, and Social6 were .55, .55, 1.176, and .55, respectively. For JS12 and JS34, this was only the case in the in-person worker group. For JS25 and Social6, this was only the case in the remote worker group. Collapsing categories was performed in effort to mitigate this issue, though it did not help lower SEs. Further, it had no impact on the results of the DIF analyses. Future research should replicate these results in a larger sample to better understand the function of these items and to investigate whether the issues of misfit and SE can be alleviated with a larger sample.

Additionally, the original JSS is scored on 1 = “Disagree very much” to 6 = “Agree very much”, with no neutral option. In this study, a 5-point Likert scale was used, scored from 1 = “Strongly disagree” to 5 = “Strongly agree” with a neutral option as the midpoint to assess job satisfaction. This choice was made to mimic the more traditional Agree-Disagree Likert scale format and to be consistent across all study scales. However, it does offer the possibility that these results would not be the same if the response options were used as intended by Spector (1985). Relatedly, two items were dropped due to low factor loadings, so measurement invariance was not assessed for those two items. Thus, future research should replicate these findings using the original scale options to see if the results differ.

Further, this study sought to compare exclusively remote workers to in-person workers. In doing so, hybrid workers were combined with in-person workers such that all employees reporting traveling to the office at least once a week were considered in-person workers. Given the increasing popularity of hybrid work (Hilberath et al., 2020; U.S. Bureau of Labor Statistics, 2022), it is important to understand how the results of this study may change if hybrid workers, were treated as their own group. While the argument was made that hybrid workers would be more similar to in-person workers, this assumption should be assessed directly. As the overarching intention of this study was to analyze exclusively remote workers versus all others, hybrid workers did not constitute a large enough sample for IRT analysis. Future research should explore MI across remote, hybrid, and in-person workers in widely used, validated scales.

Finally, future research should investigate why some of these items are non-invariant across these two groups. Examining DF items may provide some insights on why this may be the case (see Table 3), and while some possibilities are provided above in an ad hoc sense, future research should take a more systematic approach to better identify what may be causing

these items to function differently in these groups. Based on findings of this study, items or scales focused on measuring constructs that involve other organizational members may be especially relevant. Relatedly, as this study only investigated MI in three organizational constructs, future researchers should investigate additional constructs relevant to the organizational literature to understand how they are functioning across different worker groups. This is especially necessary given the recent upsurge in the proportion of the workforce performing remote work (U.S. Bureau of Labor Statistics, 2022). With these investigations, organizational practitioners and researchers alike can have more confidence in their results, provide accurate recommendations, and make more educated decisions.

Conclusion

To conclude, researchers and practitioners alike have long been interested in the differences between remote and in-person workers on constructs such as job satisfaction, autonomy, and social support. This study illustrates how conclusions of prior research using the JSS, WDQ autonomy scale, and WDQ social support scale may not be as appropriate or informative for comparing remote and in-person workers as we may have assumed. While scale level effect sizes suggest the impact of differential functioning is trivial, this study illustrates how the findings of past research to compare these two groups can be misleading. In particular, disparate conclusions are drawn when using invariant versus non-invariant versions of these scales to understand relationships. Moreover, when the interest is in the facet or item level, particularly for those involving satisfaction with coworkers, satisfaction with supervisors, satisfaction with nature of work, social support, or work scheduling autonomy, the ability to make meaningful mean comparisons across remote and in-person workers using these scales is impeded.

REFERENCES

Allen, D. G., Bryant, P. C., & Vardaman, J. M. (2010). Retaining talent: Replacing misconceptions with evidence-based strategies. *Academy of management Perspectives*, *24*(2), 48-64.

Allen, T. D., Golden, T. D., & Shockley, K. M. (2015). How effective is telecommuting? Assessing the status of our scientific findings. *Psychological science in the public interest, 16*(2), 40-68.

Ambrose, M. L., & Adler, G. S. (2000). Designing, implementing, and utilizing computerized performance monitoring: Enhancing organizational justice. *Research in Personnel and Human Resource Management, 18,* 187–219.

Ashkanasy, N. M., Wilderom, C. P., & Peterson, M. F. (Eds.). (2000). *Handbook of organizational culture and climate*. Sage.

Bailey, D. E., & Kurland, N. B. (2002). A review of telework research: Findings, new directions, and lessons for the study of modern work. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *23*(4), 383-400.

Bakker, A. B., & Demerouti, E. (2007). The job demands-resources model: State of the art. *Journal of managerial psychology*, *22*(3), 309-328.

Barrero, J. M., Bloom, N., & Davis, S. J. (2021). *Why working from home will stick* (No. w28731). National Bureau of Economic Research.

Barría-González, J., Postigo, Á., Pérez-Luco, R., Cuesta, M., & García-Cueto, E. (2021). Assessing Organizational Climate: psychometric properties of the ECALS Scale. *Anales de Psicología/Annals of Psychology*, *37*(1), 168-177.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117*(3), 497–529. https://doi.org/10.1037/0033-2909.117.3.497

Beiler-May, A., Williamson, R. L., Clark, M. A., & Carter, N. T. (2017). Gender bias in the measurement of workaholism. *Journal of personality assessment*, *99*(1), 104-110.

Bentley, T. A., Teo, S. T., McLeod, L., Tan, F., Bosua, R., & Gloet, M. (2016). The role of organisational support in teleworker wellbeing: A socio-technical systems approach. *Applied ergonomics*, *52*, 207-215.

Boell, S. K., Cecez-Kecmanovic, D., & Campbell, J. (2016). Telework paradoxes and practices: The importance of the nature of work. *New Technology, Work and Employment*, *31*(2), 114-131.

Brien, M., Forest, J., Mageau, G. A., Boudrias, J. S., Desrumaux, P., Brunet, L., & Morin, E. M. (2012). The basic psychological needs at work scale: Measurement invariance between Canada and France. *Applied Psychology: Health and Well-Being*, *4*(2), 167-187.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford publications.

Brunelle, E., & Fortin, J. A. (2021). Distance makes the heart grow fonder: An examination of teleworkers' and office workers' job satisfaction through the lens of self-determination theory. *Sage open*, *11*(1), 1-11.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin*, *56*(2), 81.

Chapman, A. J., Sheehy, N. P., Heywood, S., Dooley, B., & Collins, S. C. (1995). The organizational implications of teleworking. In C. L. Cooper & I. T. Robertson (Eds.), International review of industrial and organizational psychology (pp. 229-248). New York, NY: Wiley.

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. Journal of Statistical Software, 48(6), 1–29.

Charalampous, M., Grant, C. A., Tramontano, C., & Michailidis, E. (2019). Systematically reviewing remote e-workers' well-being at work: A multidimensional approach. *European Journal of Work and Organizational Psychology*, *28*(1), 51-73.

Chirkov, V., Ryan, R. M., Kim, Y., & Kaplan, U. (2003). Differentiating autonomy from individualism and indepen- dence: A self-determination theory perspective on internal-ization of cultural orientations and well-being. *Journal of Personality and Social Psychology*, *84*, 97–110.

Cohen, J. (1988). The effect size. *Statistical power analysis for the behavioral sciences*, 77-83.

Converso, D., Bruno, A., Capone, V., Colombo, L., Falco, A., Galanti, T., ... & Loera, B. (2021). Working during a pandemic between the risk of being infected and/or the risks related to social distancing: First validation of the SAPH@ W questionnaire. *International Journal of Environmental Research and Public Health*, *18*(11), 5986.

Cooper, C. D., & Kurland, N. B. (2002). Telecommuting, professional isolation, and employee development in public and private organizations. Journal of Organizational Behavior, 23, 511-532.

Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management science*, *32*(5), 554-571.

Dalal, R. S. (2005). A meta-analysis of the relationship between organizational citizenship behavior and counterproductive work behavior. Journal of Applied Psychology, 90, 1241–1255. doi:10.1037/0021-9010.90.6.1241

Dart, S. (2022, September 29). 94 percent of knowledge workers prefer remote or hybrid: Gallup poll. www.a.team. Retrieved January 16, 2023, from https://www.a.team/mission/workers-prefer-remote-or-hybrid-gallup-poll

DeCarlo, T. E., & Agarwal, S. (1999). Influence of managerial behaviors and job autonomy on job satisfaction of industrial salespersons: A cross-cultural study. *Industrial Marketing Management*, *28*(1), 51-62.

Deci, E. L., & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. *Journal of research in personality*, *19*(2), 109-134.

Demerouti, E., Bakker, A. B., Nachreiner, F., & Schaufeli, W. B. (2001). The job demandsresources model of burnout. *Journal of Applied psychology*, *86*(3), 499-512.

Dey, M., Frazis, H., Loewenstein, M. A., Sun, H. (2020, June). *Ability to work from home: evidence from two surveys and implications for the labor market in the COVID-19 pandemic.* U.S. Bureau of Labor Statistics. https://www.bls.gov/opub/mlr/2020/article/ability-to-work-from-home.htm.

Dey, M. (2022, September 23). *Employee satisfaction in 2022: Measuring Job Satisfaction*. SelectHub raquo. Retrieved November 11, 2022, from https://www.selecthub.com/hris/measure-employee-satisfaction/

Donovan, M. A., Drasgow, F., & Probst, T. M. (2000). Does computerizing paper-and-pencil job attitude scales make a difference? New IRT analyses offer insight. *Journal of applied psychology*, *85*(2), 305-313.

Drasgow, F. (1987). Study of measurement bias of two standardized psychological tests. *Journal of Applied Psychology, 72*, 19-29.

Drasgow, F., Nye, C. D., Stark, S., & Chernyshenko, O. S. (2018). Differential item and test functioning. *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development*, 885-899.

Dubrin, A. J. (1991). Comparison of the job satisfaction and productivity of telecommuters versus in-house employees: a research note on work in progress. *Psychological Reports*, *68*(3_suppl), 1223-1234.

Elsbach, K. D. (2003). Relating physical environment to self- categorizations: Identity threat and affirmation in a non-territorial office space. *Administrative Science Quarterly, 48*, 622– 654.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Lawrence Erlbaum Associates Publishers.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological methods*, *4*(3), 272299.

Facteau, J. D., & Craig, S. B. (2001). Are performance appraisal ratings from different rating sources comparable?. *Journal of Applied Psychology*, *86*(2), 215-227.

Feldman, D. C., & Gainey, T. W. (1997). Patterns of telecommuting and their consequences: Framing the research agenda. *Human Resource Management Review*, *7*(4), 369-388.

Fonner, K. L., & Roloff, M. E. (2010). Why teleworkers are more satisfied with their jobs than are office-based workers: When less contact is beneficial. *Journal of Applied Communication Research*, *38*(4), 336-361.

French, K. A., Dumani, S., Allen, T. D., & Shockley, K. M. (2018). A meta-analysis of work–family conflict and social support. *Psychological bulletin*, *144*(3), 284-314.

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: current use, calculations, and interpretation. *Journal of experimental psychology: General, 141*(1), 2.

Gagné, M., & Deci, E. L. (2005). Self-determination theory and work motivation. *Journal of Organizational behavior*, *26*(4), 331-362.

Gajendran, R. S., & Harrison, D. A. (2007). The good, the bad, and the unknown about telecommuting: meta-analysis of psychological mediators and individual consequences. *Journal of applied psychology*, *92*(6), 1524-1541.

Gajendran, R. S., Harrison, D. A., & Delaney-Klinger, K. (2015). Are telecommuters remotely good citizens? Unpacking telecommuting's effects on performance via i-deals and job resources. Personnel psychology, 68(2), 353-393.

Gifford, J. (2022). Remote working: unprecedented increase and a developing research agenda. *Human Resource Development International*, *25*(2), 105-113.

Global Workplace Analytics. (2020). *Work-at-Home After Covid-19 – Our Forecast.*

Golden, T. D. (2006). The role of relationships in understanding telecommuter satisfaction. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *27*(3), 319-340.

Golden, T. D., & Gajendran, R. S. (2019). Unpacking the role of a telecommuter's job in their performance: Examining job complexity, problem solving, interdependence, and social support. *Journal of Business and Psychology*, *34*(1), 55-69.

Golden, T. D., & Veiga, J. F. (2005). The impact of extent of telecommuting on job satisfaction: Resolving inconsistent findings. *Journal of management*, *31*(2), 301-318.

Golden T. D., Veiga J. F., Dino R. N. (2008). The impact of professional isolation on teleworker job performance and turnover intentions: Does time spent teleworking, interacting face-to-face, or having access to communication-enhancing technology matter? *Journal of*

Applied Psychology, 93, 1412–1421.

Grant, C., Wallace, L. M., & Spurgeon, P. C. (2013) An exploration of the psychological factors affecting remote e-worker's job effectiveness, well-being and work-life balance. *Employee Relations, 35* (5): 527-546. http://dx.doi.org/10.1108/ER-08-2012-0059

Guimaraes, T., & Dallow, P. (1999). Empirically testing the benefits, problems, and success factors for telecommuting programmes. *European Journal of Information Systems, 8*(1), 40-54.

Hackman, J. R., & Oldham, G. R. (1976). Motivation through the design of work: Test of a theory. Organizational Behavior & Human Performance, 16, 250–279. http://dx.doi.org/10.1016/0030-5073(76)90016-7

Hair, J. F., Tatham, R. L., Anderson, R. E., and Black, W. (1998). *Multivariate data analysis. (Fifth Ed.)* Prentice-Hall: London.

Hammack-Brown, B. (2018). *Testing the measurement invariance and latent mean differences of data from the employee perceptions of supervisor/line manager coaching behavior measure between virtual and traditional employees from generation X and Y.* [Unpublished doctoral dissertation]. University of Texas at Taylor.

Harrington, S. J., & Santiago, J. (2006). Organizational culture and telecommuters' quality of work life and professional isolation. *Communications of the IIMA*, *6*(3), 1-10.

Harrison, D. A., Crockett, D., & Juban, R. (2000, August). *Time, place and control: Organizational antecedents and individual outcomes of flexibility in the when and where of work.* Paper presented at the meeting of the Academy of Management, Toronto, Ontario, Canada.

Hausknecht, J. P., Rodda, J., & Howard, M. J. (2009). Targeted employee retention: Performance-based and job-related differences in reported reasons for staying. *Human*

Resource Management: Published in Cooperation with the School of Business Administration, The University of Michigan and in alliance with the Society of Human Resources Management, *48*(2), 269-288.

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis (second edition): A regression-based approach*. Guilford Press.

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational research methods, 7*(2), 191-205.

Hilberath, C., Kilmann, J., Lovich, D., Tzanetti, T., Bailey, A., Beck, S., Kaufman, E., Khandelwal, B., Schuler, F., & Woolsey, K. (2020). Hybrid work is the new remote work. *Boston Consulting Group.[en línea], Recuperado de: https://www.bcg.com/publications/2020/managing-remote-work-and-optimizing-hybridworking-models*.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 32*, 179-185.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental aging research*, *18*(3), 117-144.

Horn, J. L., McArdle, J., & Mason, R. (1983). When is invariance not invariant: A practical scientist's look at the ethereal concept of factor invariance. *Southern Psychologist, 1*, 179-188

Hornung, S., & Glaser, J. (2009). Home-based telecommuting and quality of life: further evidence on an employee-oriented human resource practice. *Psychological Reports*, *104*(2), 395-402.

Hu, L. T., & 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.

Hughes, I. M., Freier, L. M., & Barratt, C. L. (2022). "Your help isn't helping me!" Unhelpful workplace social support, strain, and the role of individual differences. *Occupational Health Science*, *6*(3), 387-423.

Humphrey, S. E., Nahrgang, J. D., & Morgeson, F. P. (2007). Integrating motivational, social, and contextual work design features: a meta-analytic summary and theoretical extension of the work design literature. *Journal of applied psychology*, *92*(5), 1332-1356.

Izquierdo, A. L. G., & Pérez, A. M. C. (2022). Work characteristics and occupational health: validation and proposal of a shortened version of the Work Design Questionnaire. *Anales de Psicología/Annals of Psychology*, *38*(1), 149-162.

Jang, S., Allen, T. D., Kim, E., O'Brien, K. E., Cho, I., & Ceylan, S. (2022). Measurement invariance of organizational citizenship behavior across gender. *Applied Psychology*, *71*(4), 1535-1564.

Johnson, E. C., Meade, A. W., & DuVernet, A. M. (2009). The role of referent indicators in tests of measurement invariance. Structural Equa- tion Modeling, 16, 642–657. doi:10.1080/10705510903206014

Jolly, P. M., Kong, D. T., & Kim, K. Y. (2021). Social support at work: An integrative review. *Journal of Organizational Behavior*, *42*(2), 229-251.

Judge, T. A., Thoresen, C. J., Bono, J. E., & Patton, G. K. (2001). The job satisfaction-job performance relationship: A qualitative and quantitative review. Psychological Bulletin, 127, 376 – 407.

Kang, T., & Chen, T. T. (2007). An Investigation of the Performance of the Generalized SX 2 Item-Fit Index for Polytomous IRT Models. ACT Research Report Series, 2007-1. ACT, Inc.

Kniffin, K. M., Narayanan, J., Anseel, F., Antonakis, J., Ashford, S. P., Bakker, A. B., ... & Vugt, M. V. (2021). COVID-19 and the workplace: Implications, issues, and insights for future research and action. *American Psychologist, 76*(1), 63-77.

Knight, C., Olaru, D., Lee, J., & Parker, S. (2022). The loneliness of the hybrid worker. *MIT Sloan Management Review*.

Kossek, E. E., Pichler, S., Bodner, T., & Hammer, L. B. (2011). Workplace Social Support and Work–Family Conflict: A Meta-Analysis Clarifying the Influence of General and Work–Family-Specific Supervisor and Organizational Support. *Personnel Psychology, 64*(2), 289–313. https://doi.org/10.1111/j.1744-6570.2011.01211.x

Lang, J. W., & Tay, L. (2021). The science and practice of item response theory in organizations. *Annual Review of Organizational Psychology and Organizational Behavior*, *8*, 311-338.

Lautsch, B. A., & Kossek, E. E. (2011). Managing a blended workforce: telecommuters and non-telecommuters. *Organizational Dynamics*, *40*(1), 10-17.

Lee, T. W., & Mitchell, T. R. (1994). An alternative approach: The unfolding model of voluntary employee turnover. *Academy of management review*, *19*(1), 51-89.

Liu, C., Borg, I., & Spector, P. E. (2004). Measurement invariance of the German job satisfaction survey used in a multinational organization: Implications of Schwartz's culture model. *Journal of Applied Psychology, 89*(6), 1070-1082.

Liu, C., E Spector, P., & M Jex, S. (2005). The relation of job control with job strains: A comparison of multiple data sources. *Journal of Occupational and Organizational Psychology*, *78*(3), 325-336.

Locke, E. A. (1976). The nature and causes of job satisfaction. *Handbook of industrial and organizational psychology*.

Lopez Rivas, G. E., Stark, S., & Chernyshenko, O. S. (2008). *The Effects of Referent Item Parameters on Differential Item Functioning Detection Using the Free Baseline Likelihood Ratio Test. Applied Psychological Measurement, 33(4), 251–265.* doi:10.1177/0146621608321760

March, J. G. & Simon, H. A. (1958). Organizations. New York: Wiley

Mathieu, M., Eschleman, K. J., & Cheng, D. (2019). Meta-analytic and multiwave comparison of emotional support and instrumental support in the workplace. *Journal of Occupational Health Psychology, 24*(3), 387–409. https://doi.org/10.1037/ocp0000135

Maydeu-Olivares A., Joe H. (2005). Limited and full information estimation and testing in 2^n contingency tables: A unified framework. *Journal of the American Statistical Association*, 100, 1009–1020.

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological methods*, *17*(3), 437.

Meade, A. W., Michels, L. C., & Lautenschlager, G. J. (2007). Are Internet and paper-and-pencil personality tests truly comparable? An experimental design measurement invariance study. *Organizational Research Methods*, *10*(2), 322-345.

Meade, A. W., & Wright, N. A. (2012). Solving the measurement invariance anchor item problem in item response theory. *Journal of Applied Psychology*, *97*(5), 1016.

Meluso, J., Johnson, S., & Bagrow, J. (2022). Flexible Environments for Hybrid Collaboration: Redesigning Virtual Work Through the Four Orders of Design. *Design Issues, 38(*1), 55-69.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58,* 525-543. doi: 10.1007/BF02294825

Meriac, J. P., Woehr, D. J., & Banister, C. (2010). Generational differences in work ethic: An examination of measurement equivalence across three cohorts. Journal of business and psychology, 25, 315-324.

Minbaeva, D. (2020). Disrupted HR? Human Resource Management Review, 100820. https:// doi.org/ 10.1016/ j.hrmr.2020.100 820

Mitropoulou, E. M., & Zampetakis, L. (2022). Introducing the scr-OLBI: Examination of a short measure for assessing burnout using item response theory. *Psychological Test Adaptation and Development,* 3, 97-105.

Mobley, W. H. (1977) "Intermediate Linkages in the Relationship Between Job Satisfaction and Employee Turnover Process." *Journal of Applied Psychology,* 62, 237-240.

Molino, M., Ingusci, E., Signore, F., Manuti, A., Giancaspro, M. L., Russo, V., ... & Cortese, C. G. (2020). Wellbeing costs of technology use during Covid-19 remote working: An investigation using the Italian translation of the technostress creators scale. *Sustainability*, *12*(15), 5911.

Morganson, V. J., Major, D. A., Oborn, K. L., Verive, J. M., & Heelan, M. P. (2010). Comparing telework locations and traditional work arrangements: Differences in work-life balance support, job satisfaction, and inclusion. *Journal of Managerial Psychology, 25*(6), 578-595.

Morgeson, F. P., & Humphrey, S. E. (2006). The Work Design Questionnaire (WDQ): developing and validating a comprehensive measure for assessing job design and the nature of work. *Journal of applied psychology*, *91*(6), 1321.

Nilles, J. M. (1994). *Making telecommuting happen: A guide for telemanagers and telecommuters*.

Nesher Shoshan, H., & Wehrt, W. (2022). Understanding "Zoom fatigue": A mixed-method approach. *Applied Psychology*, *71*(3), 827-852.

Nübold, A., Van Quaquebeke, N., & Hülsheger, U. R. (2020). Be (com) ing real: A multi-source and an intervention study on mindfulness and authentic leadership. *Journal of Business and Psychology*, *35*(4), 469-488.

Orlando, M., & Thissen, D. (2003). Further investigation of the performance of S-X2: An item fit index for use with dichotomous item response theory models. *Applied Psychological Measurement, 27*(4), 289-298.

Ozimek, A. (2020). The future of remote work. Available at SSRN 3638597.

Park, T. Y., & Shaw, J. D. (2013). Turnover rates and organizational performance: A meta-analysis. *Journal of Applied Psychology, 98,* 268 –309. http://dx.doi.org/10.1037/a0030723

Parker, S. K., Andrei, D. M., & Van den Broeck, A. (2019). Poor work design begets poor work design: Capacity and willingness antecedents of individual work design behavior. *Journal of Applied Psychology*, *104*(7), 907.

Pinsonneault, A., & Boisvert, M. (2001). The impacts of telecommuting on organizations and individuals: A review of the literature. In N. J. Johnson (Ed.), *Telecommuting and virtual offices: Issues and* opportunities (pp. 163- 85). Hershey, PA: Idea Group Publishing.

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of management*, *12*(4), 531-544.

Powell, W. W., & Snellman, K. (2004). The knowledge economy. *Annual review of sociology*, *30*, 199–220.

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental review*, *41*, 71-90.

Pyöriä, P. (2011). Managing telework: risks, fears and rules. *Management Research Review*, *34*(4), 386-399.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rafaeli, A., & Worline, M. (2000). Symbols in organizational culture. *Handbook of organizational culture and climate, 7, 1*-84.

Raghuram, S., Wiesenfeld, B., & Garud, R. (2003). Technology enabled work: The role of self-efficacy in determining telecommuter adjustment and structuring behavior. *Journal of Vocational Behavior*, *63*(2), 180-198.

Ramsower, R. M. (1983). *Telecommuting: An investigation of some organizational and behavioral effects of working at home*. Unpublished doctoral dissertation, University of Minnesota, Minneapolis.

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics, 4,* 207–230.

Revelle, W. (2021). Package 'psych'. Retrieved from https://cran.r-project.org/web/packages/psych/psych.pdf

Rogelberg, S. G., Church, A. H., Waclawski, J., & Stanton, J. M. (2008). Organizational Survey Research. In *Handbook of Research Methods in Industrial and Organizational Psychology* (pp. 140–160). https://doi.org/10.1002/9780470756669

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*, 68–78.

Santarpia, F. P., Borgogni, L., Consiglio, C., & Menatta, P. (2021). The Bright and Dark Sides of Resources for Cross-Role Interrupting Behaviors and Work–Family Conflict: Preliminary Multigroup Findings on Remote and Traditional Working. *International Journal of Environmental Research and Public Health*, *18*(22), 12207.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. Psychometrika Monograph Supplement, Vol. 34(4), 100.

Sardeshmukh, S. R., Sharma, D., & Golden, T. D. (2012). Impact of telework on exhaustion and job engagement: A job demands and job resources model. *New Technology, Work and Employment, 27,* 193–207.

Schall, M. A. (2019). *The relationship between remote work and job satisfaction: The mediating roles of perceived autonomy, work-family conflict, and telecommuting intensity* (Doctoral dissertation, San Jose State University).

Schneider, B. (1990). The climate for service: An application of the climate construct. *Organizational climate and culture*, *1*, 383-412.

Schneider, B., Ehrhart, M. G., & Macey, W. H. (2011). Organizational climate research. *The handbook of organizational culture and climate*, *29*, 12169-012.

Sewell, G., & Taskin, L. (2015). Out of sight, out of mind in a new world of work? Autonomy, control, and spatiotemporal scaling in telework. *Organization studies*, *36*(11), 1507-1529.

Sims, H. P., Szilagyi, A. D., & Keller, R. T. (1976). The measurement of job characteristics. *Academy of Management Journal, 19,* 195–212

Sinar, E. F., & Zickar, M. J. (2002). Evaluating the robustness of graded response model and classical test theory parameter estimates to deviant items. Applied Psychological Measurement, 26(2), 181-191.

Slack. (2020, April 21). *Report: Remote work in the age of covid-19*. Slack. Retrieved January 16, 2023, from https://slack.com/intl/de-de/blog/collaboration/report-remote-work-during-coronavirus

Smith, P. C., Kendall, L. M., Huling, C. L., Balzer, W. K., Kihm, J. A., Irwin, J. L., & Parra, L. (1969). Job Descriptive Index (1997 Revision) and The Job in General Scales.

Sparrow, P. R. (2000). New employee behaviours, work designs and forms of work organization: *What is in store for the future of work? Journal of Managerial Psychology, 15,* 202–218.

Spector, P. E. (1985). Measurement of human service staff satisfaction: Development of the Job Satisfaction Survey. *American Journal of Community Psychology*, *13*, 693-713.

Spector, P. E. (1986). Perceived control by employees: A meta-analysis of studies concerning autonomy and participation at work. *Human relations*, *39*(11), 1005-1016.

Spector, P. E. (1997). *Job satisfaction: Application, assessment, causes, and consequences* (Vol. 3). Sage.

Standen, P. (2000). The home/work interface. In K. Daniels, D. A. Lamond, P. Standen (Eds.), *Managing telework: Perspectives from human resource management and work psychology* (pp. 83–92). London: BusinessPress.

Taboroši, S., Popović, J., Poštin, J., Konjikušić, M., & Nikolić, M. (2022). Job satisfaction in the conventionally employed and teleworkers: the impact of gender, age and education. *The Annals of the Faculty of Economics in Subotica, 58*(48)*,* 65-82.

Tay, L., Meade, A. W., & Cao, M. (2015). An overview and practical guide to IRT measurement equivalence analysis. *Organizational Research Methods*, *18*(1), 3-46.

Tennant, A., Penta, M., Tesio, L., Grimby, G., Thonnard, J. L., Slade, A., ... & Phillips, S. (2004). Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. *Medical care,* I37-I48.

Thompson, C. A., & Prottas, D. J. (2006). Relationships among organizational family support, job autonomy, perceived control, and employee well-being. *Journal of occupational health psychology*, *11*(1), 100.

U.S. Bureau of Labor Statistics. (2019, September 24). *Table 3. Workers who worked at home and how often they worked exclusively at home by selected characteristics, averages for the period 2017-2018.* Retrieved August 10, 2022, from https://www.bls.gov/news.release/flex2.t03.htm.

U.S. Bureau of Labor Statistics. (2020, June 24). *Ability to work from home: evidence from two surveys and implications for the labor market in the COVID-19 pandemic*. Retrieved August 10, 2022, from https://www.bls.gov/opub/mlr/2020/article/ability-to-work-from-home.htm.

U.S. Bureau of Labor Statistics. (2022, June 23). *Table 7. Employed persons working on main job at home, workplace, and time spent working at each location by class of worker,*

occupation, and earnings, 2021 annual averages. Retrieved August 10, 2022, from https://www.bls.gov/news.release/atus.t07.htm.

Van den Broeck, A., Vansteenkiste, M., De 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.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational research methods*, *3*(1), 4-70.

Vander Elst, T., Verhoogen, R., Sercu, M., Van Den Broeck, A., Baillien, E., & Godderis, L. (2017). Not extent of telecommuting, but job characteristics as proximal predictors of work-related well-being. *Journal of Occupational and Environmental Medicine, 59*(10), 180–186.

Vega, R. P., Anderson, A. J., & Kaplan, S. A. (2015). A within-person examination of the effects of telework. *Journal of Business and Psychology*, *30*(2), 313-323.

Virick, M., DaSilva, N., & Arrington, K. (2010). Moderators of the curvilinear relation between extent of telecommuting and job and life satisfaction: the role of performance outcome orientation and worker type. *Human relations*, *63*(1), 137-154.

Viswesvaran, C., Sanchez, J. I., & Fisher, J. (1999). The role of social support in the process of work stress: A meta-analysis. *Journal of Vocational Behavior*, 54, 314-334.

Wang, B., Liu, Y., Qian, J., & Parker, S. K. (2021). Achieving effective remote working during the COVID-19 pandemic: A work design perspective. *Applied psychology*, *70*(1), 16-59.

Warr, P. (1987). *Work, Unemployment, and Mental Health*, Clarendon Press: Oxford.

Watson, A. M., Thompson, L. F., & Meade, A. W. (2007, April). Measurement invariance of the job satisfaction survey across work contexts. In *22nd Annual Meeting of the Society for Industrial and Organizational Psychology, New York*.

Wheatley, D. (2012). Good to be home? Time-use and satisfaction levels among home-based teleworkers. *New Technology, Work and Employment*, *27*(3), 224-241.

Wiesenfeld, B. M., Raghuram, S., & Garud, R. (2001). Organizational identification among virtual workers: The role of need for affiliation and perceived work-based social support. *Journal of management*, *27*(2), 213-229.

Zagorsek, H., Stough, S. J., & Jaklic, M. (2006). Analysis of the reliability of the leadership practices inventory in the item response theory framework. International Journal of Selection and Assessment, 14(2), 180-191

Zöllner, K., & Sulíková, R. (2021). Teleworking and its influence on job satisfaction. *Journal of Human Resources Management Research*, *1*(1).

APPENDICES

Table 1

Correlations for study variables for the whole sample and within worker groups.

	Whole Sample					Remote			In-person		
Variable	M	SD	1	2	3	1	2	3	1	2	3
1 JS	3.51	.71	(.96)			(.96)			(.96)		
2 Auto	3.88	.85	.40*	(.96)		.41*	(.96)		.40*	(.95)	
3 Social	3.74	.83	.66*	.32*	(.89)	.64*	.28*	(.90)	.71*	.42*	(.89)

Note. Numbers in parentheses along the diagonal represent reliability estimates. JS = Job satisfaction, Auto = Autonomy, Social = Social support.
*$p < .001$.

Table 2

Item fit statistics in the full sample and independent item parameter estimates.

	Item Fit Statistics			Remote					In-Person				
Item	S-χ^2	*df*	*p*	*a*	*b1*	*b2*	*b3*	*b4*	*a*	*b1*	*b2*	*b3*	*b4*
JS1	94.0	97	.57	2.1	-2.0	-0.8	-0.4	1.3	2.0	-2.2	-0.9	-0.6	1.2
JS2	131.3	115	.14	2.2	-1.5	-0.3	0.4	1.9	1.6	-1.7	-0.5	0.3	1.9
JS3	96.9	94	.40	1.5	-3.2	-2.3	-1.1	0.5	2.0	-2.4	-1.6	-1.1	0.5
JS4	147.4	141	.34	1.5	-1.7	-0.7	-0.1	1.4	1.7	-1.9	-1.0	-0.5	1.1
JS5	102.2	89	.16	2.6	-1.8	-0.9	-0.2	1.3	2.7	-1.9	-1.0	-0.3	1.3
JS6	129.4	111	.11	1.6	-2.7	-1.4	-0.6	1.3	1.6	-2.4	-1.1	-0.6	1.3
JS7	92.3	75	.09	1.7	-3.3	-2.6	-1.1	1.0	1.7	-3.5	-2.2	-1.4	0.6
JS8	133.7	120	.19	1.5	-2.0	-0.6	-0.2	1.4	2.0	-2.1	-1.0	-0.5	0.8
JS9	99.4	107	.69	1.9	-2.2	-1.3	-0.5	1.5	2.3	-2.0	-1.0	-0.4	1.1
JS10	123.1	127	.58	2.2	-0.7	0.0	0.8	1.9	1.8	-1.2	-0.1	0.6	1.9
JS11	101.5	100	.44	2.3	-1.9	-0.7	0.2	1.6	2.0	-1.9	-0.9	-0.1	1.7
JS12	94.3	75	.07	1.5	-3.8	-2.7	-1.6	0.1	2.2	-2.7	-1.7	-1.2	0.1
JS13	138.5	125	.19	1.6	-2.0	-0.7	-0.1	1.6	1.6	-2.3	-1.1	-0.6	1.4
JS14	68.3	87	.93	2.9	-1.8	-0.9	-0.3	1.0	3.3	-2.0	-0.9	-0.5	0.8
JS16	1302	129	.45	1.1	-3.1	-1.7	-0.8	1.2	1.8	-2.2	-1.0	-0.5	1.0
JS17	92.3	96	.59	1.4	-2.7	-1.4	-0.6	1.7	2.6	-2.3	-1.4	-0.8	0.9
JS18	95.3	91	.36	1.4	-3.8	-1.9	-1.2	1.0	1.8	-2.9	-1.7	-1.2	0.5
JS19	107.2	103	.37	2.8	-1.3	-0.5	-0.1	0.9	3.1	-1.5	-0.7	-0.3	0.8
JS20	81.6	104	.95	2.4	-1.6	-0.5	0.4	2.2	1.5	-2.1	-0.6	0.4	2.6
JS21	111.1	101	.23	1.9	-2.3	-1.7	-0.8	0.6	2.2	-1.9	-1.2	-0.7	0.6
JS22	144.7	130	.18	1.8	-1.6	-0.7	-0.1	1.5	1.5	-2.4	-1.2	-0.5	1.5
JS23	73.7	91	.91	3.0	-1.6	-0.5	0.1	1.2	2.7	-1.9	-0.7	-0.1	1.1
JS25	78.9	81	.55	1.7	-3.2	-2.3	-0.8	1.2	1.7	-3.8	-1.9	-1.4	0.8
JS26	134.1	117	.13	1.7	-2.3	-0.9	-0.4	1.4	1.7	-2.3	-1.1	-0.6	1.0
JS27	95.2	112	.87	1.6	-2.5	-1.2	-0.4	1.6	1.8	-2.7	-1.5	-0.7	1.1
JS28	91.2	112	.93	2.6	-1.0	-0.2	0.3	1.4	2.1	-1.7	-0.5	0.0	1.4
JS29	117.0	130	.79	1.7	-1.4	-0.2	0.5	1.8	1.5	-1.9	-0.4	0.3	2.1
JS30	99.5	83	.10	1.8	-2.9	-2.2	-1.0	0.8	2.0	-2.5	-1.7	-1.0	0.6
JS32	90.1	99	.73	2.6	-1.3	-0.4	0.1	1.3	2.8	-1.7	-0.6	-0.1	1.2
JS33	110.5	107	.39	2.7	-1.2	-0.3	0.3	1.6	2.1	-1.7	-0.6	0.1	1.7
JS34	115.1	110	.35	0.8	-5.8	-3.9	-2.4	-0.1	1.5	-2.7	-1.7	-1.2	0.5
JS35	109.9	91	.09	1.7	-2.7	-1.3	-0.4	1.7	2.6	-2.2	-1.3	-0.5	1.1
JS36	**156.3**	**118**	**.01**	**1.2**	**-3.1**	**-1.5**	**-0.8**	**1.4**	**1.6**	**-2.8**	**-1.5**	**-0.9**	**0.9**
Auto1	**83.5**	**45**	**.00**	**2.2**	**-2.0**	**-1.2**	**-0.9**	**0.6**	**2.3**	**-1.8**	**-1.0**	**-0.5**	**1.1**
Auto2	**50.5**	**35**	**.04**	**3.5**	**-2.1**	**-1.3**	**-0.8**	**0.5**	**2.8**	**-2.2**	**-1.3**	**-0.9**	**0.9**
Auto3	40.7	30	.09	4.0	-2.0	-1.5	-1.0	0.4	3.7	-2.3	-1.3	-0.9	0.7
Auto4	25.9	27	.53	3.7	-2.3	-1.5	-0.9	0.5	3.9	-2.2	-1.4	-0.9	0.5
Auto5	22.0	23	.52	4.6	-2.2	-1.3	-0.8	0.5	5.8	-2.4	-1.2	-0.7	0.6
Auto6	27.3	26	.40	4.3	-2.0	-1.3	-0.6	0.6	3.8	-2.6	-1.2	-0.6	0.9
Auto7	30.6	30	.43	4.0	-2.0	-1.1	-0.7	0.5	3.4	-2.5	-1.4	-0.8	0.7
Auto8	27.2	24	.30	6.8	-1.9	-1.1	-0.7	0.4	4.1	-2.2	-1.3	-0.7	0.6

Auto9	25.7	21	.22	5.6	-2.1	-1.2	-0.7	0.7	5.5	-2.3	-1.4	-0.8	0.8
Social1	**41.3**	**21**	**.01**	**4.3**	**-1.3**	**-0.4**	**0.1**	**1.3**	**4.2**	**-2.3**	**-1.4**	**-0.6**	**0.8**
Social2	22.9	17	.15	7.4	-1.5	-0.7	-0.3	1.1	4.3	-2.3	-1.6	-1.2	0.6
Social3	26.1	21	.20	4.0	-1.3	-0.5	-0.1	1.2	3.4	-2.4	-1.7	-1.2	0.7
Social4	**65.9**	**45**	**.02**	**1.2**	**-3.1**	**-2.0**	**-0.9**	**1.2**	**1.5**	**-2.7**	**-1.7**	**-0.8**	**0.8**
Social5	34.1	25	.11	2.5	-1.9	-0.9	-0.2	1.4	2.6	-2.4	-1.5	-0.6	1.1
Social6	**33.4**	**21**	**.04**	**2.0**	**-3.0**	**-2.5**	**-1.3**	**0.6**	**2.2**	**-3.7**	**-2.4**	**-1.4**	**0.5**

Note. Lines shown in bold indicate misfitting items (p <.05).

Table 3

Item labels and content for study items identified as containing significant DF.

Item	Content
JS3	My supervisor is quite competent in doing their job.
JS7	I like the people I work with.
JS8	I sometimes feel my job is meaningless.
JS9	Communications seem good within this organization.
JS12	My supervisor is unfair to me.
JS16	I find I have to work harder at my job because of the incompetence of people I work with.
JS17	I like doing the things I do at work.
JS21	My supervisor shows too little interest in the feelings of subordinates.
JS25	I enjoy my coworkers.
JS34	There is too much bickering and fighting at work.
JS35	My job is enjoyable.
Auto1	The job allows me to make my own decisions about how to schedule my work.
Social3	I have the opportunity to meet with others in my work.
Social4	My supervisor is concerned about the welfare of the people that work for him/her.
Social5	People I work with take a personal interest in me.
Social6	People I work with are friendly.

Table 4

DIF effect sizes and expected scores.

Item	SIDS	UIDS	ESSD	ES Remote	ES In-Person
JS3	.17	.18	.24	3.06	2.89
JS7	-0.07	0.10	-0.12	2.96	3.03
JS8	-0.23	0.26	-0.28	2.30	2.54
JS9	0.15	0.20	0.19	2.57	2.42
JS12	0.18	0.21	0.27	3.28	3.10
JS16	0.22	0.26	0.31	2.71	2.50
JS17	-0.13	0.22	-0.18	2.59	2.71
JS21	0.25	0.25	0.30	2.92	2.67
JS25	-0.10	0.11	-0.19	2.83	2.94
JS34	0.43	0.43	0.76	3.31	2.90
JS35	-0.05	0.16	-0.07	2.51	2.56
Test-Level	.06	4.3	.00	-	-
Auto1	0.25	0.25	0.28	2.87	2.62
Test-Level	.50	.72	.06	-	-
Social3	-0.23	0.23	-0.23	2.25	2.48
Social4	0.49	0.49	0.71	2.78	2.29
Social5	0.22	0.22	0.25	2.39	2.17
Social6	0.37	0.37	0.59	3.11	2.75
Test-Level	.78	1.38	.15		

Note. SIDS = Signed difference of the sample, UIDS = Unsigned difference of the sample, ESSD = Expected score standardized difference, ES = Expected score. A positive SIDS indicates the focal group (remote workers) has a higher expected score.

Table 5

Mean differences across worker group with and without significant DF items.

Scale	Remote M (SD)	In-person M (SD)	Cohen's *d*	Remote M no DF (SD)	In-person M no DF (SD)	*d*
JSS	3.5 (.73)	3.6 (.74)	.13	**3.3 (.86)**	**3.4 (.78)**	.17
WDQ Auto	3.9 (.89)	3.8 (.82)	.10	3.9 (.90)	3.9 (.82)	.06
WDQ Social	**3.5 (.90)**	**3.9 (.72)**	.54	**3.3 (1.1)**	**3.9 (.84)**	.68

Note. JSS = Job satisfaction survey, WDQ Auto = Work Design Questionnaire Autonomy scale, WDQ Social = Work Design Questionnaire Social Support scale, M no DF = Mean without DF items included. Means that differ significantly across groups are bolded.

Figure 1

Expected score curves for JSS items..

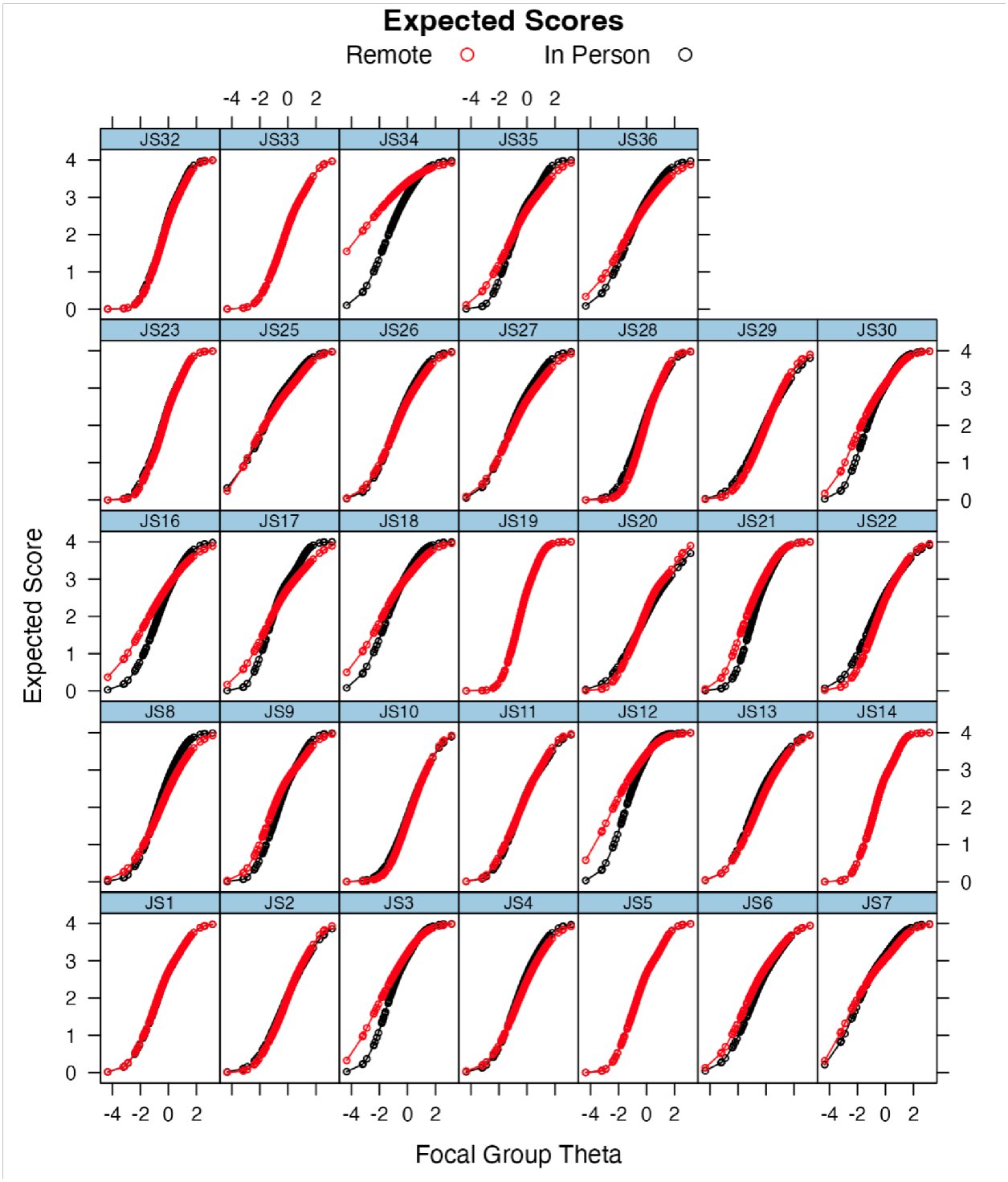

Figure 2

Expected score curves for WDQ autonomy scale items.

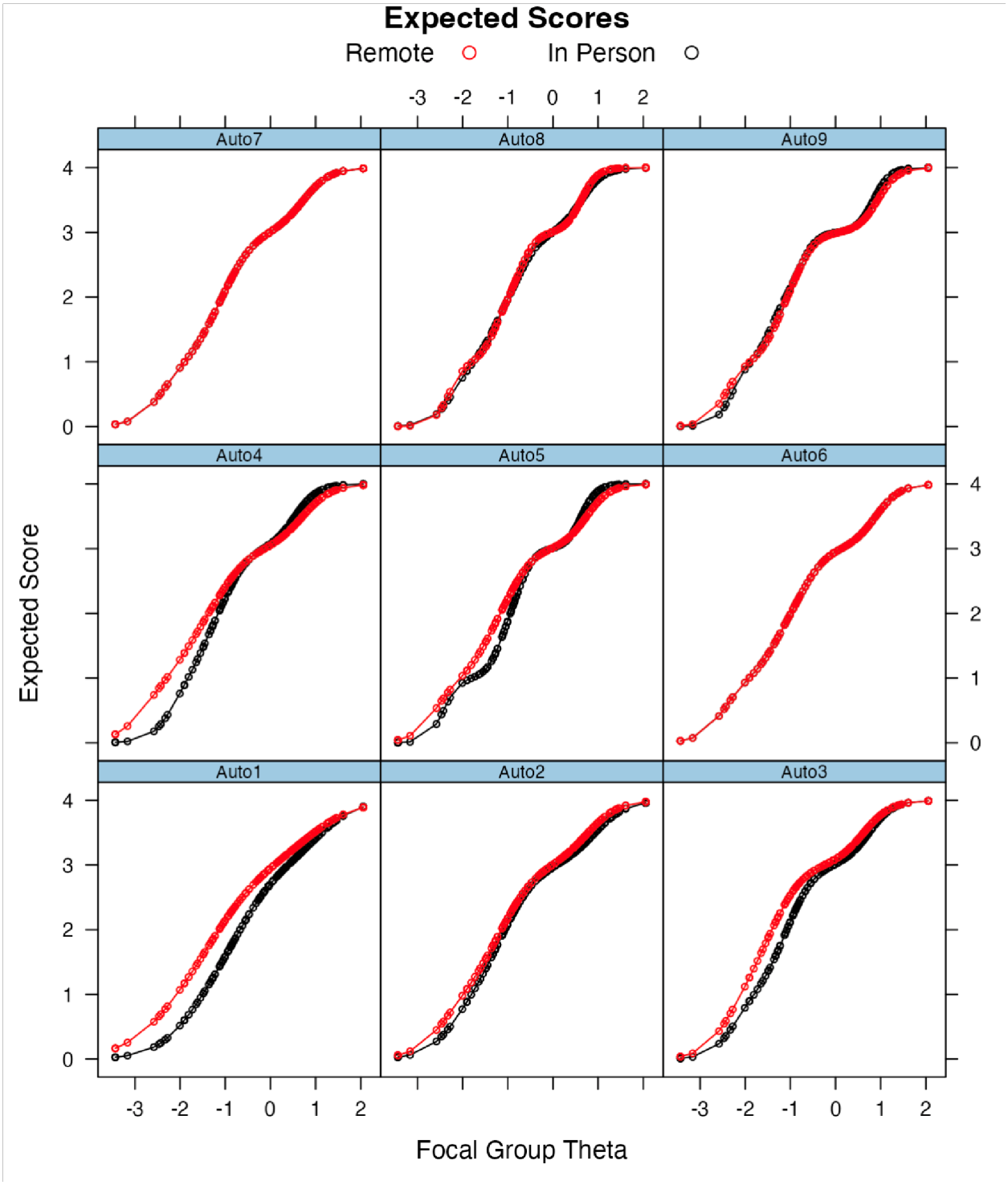

Figure 3

Expected score curves for WDQ social support scale items.