

THESIS

THE DEVELOPMENT OF A MEASURE OF WORK-RELATED UNDERLOAD

Thesis Submitted by

Megan N. Naude

Department of Psychology

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Fall 2015

Master's Committee:

Advisor: Gwenith Fisher

Zinta Byrne

Michael Steger

Chris Henle

Copyright by Megan Nicole Naude 2015

All Rights Reserved

ABSTRACT

THE DEVELOPMENT OF A MEASURE OF WORK-RELATED UNDERLOAD

This study outlines the development of the construct of work-related underload as well as a scale to measure underload. Underload has been given limited attention in the stress literature but could be a significant source of stress and other negative outcomes in the workplace. After reviewing the literature and the theoretical background related to underload, a clear definition of underload was established based on employees' perceptions, expectations, and desires related to workload. This definition informed the development of a three-factor scale to measure underload. Data from two development samples was collected in order to evaluate the reliability and validity of the scale. The results provide evidence of the internal consistency and construct validity of the three-dimensional scale. The work-related underload scale may be used in the future to explore the antecedents and consequences associated with the experience of underload at work.

TABLE OF CONTENTS

Abstract.....	ii
Introduction.....	6
Theoretical Background.....	2
Prior Empirical Research Related to Underload.....	7
Defining Underload	9
Measurement of Underload.....	9
Nomological Net.....	12
Potential Sources of Underload	14
Significance of the Study	16
Developing a Measure of Underload at Work	17
Study 1: Initial Item Development and Review.....	22
Study 1- Method	22
Participants.....	22
Procedure	22
Study 1- Results	26
Study 2: Scale Evaluation.....	27
Study 2- Method	27
Participants.....	27
Procedure	30
Measures	32
Study 2- Results	35
Statistical Analyses for Scale Evaluation	35
Sample 2 Results.....	39
Scale Reduction	41
Relations with Other Variables.....	43
Discussion.....	46
Theoretical Contribution.....	50
Practical Implications.....	50
Limitations	52
Future Research	53

Conclusion	56
References.....	85
Appendix A: Previous Measures of Underload	95
Appendix B: 25-Item Underload Scale for Study 2.....	96
Appendix C: Job Boredom Scale (Lee, 1986)	98
Appendix D: Scale of Perceived Overqualification (Maynard et al, 2006)	99
Appendix E: Role Ambiguity Scale (Breugh & Colihan, 1994).....	100
Appendix F: Job Satisfaction Scale (Cammann et al., 1979)	101
Appendix G: Stress in General Scale (Stanton et al., 2001)	102
Appendix H: Work and Meaning Inventory (Steger et al., 2012)	103
Appendix I: Role Overload Scale (Cammann et al., 1983)	104
Appendix J: Crowne-Marlowe Social Desirability Scale (Crowne & Marlowe, 1960)	105
Appendix K: Underload Scale (Caplan et al., 1975)	107
Appendix L: Final 18-Item Underload Scale.....	108

LIST OF TABLES

Table 1- Initial item pool and SME ratings for perceptions dimension from Study 1.....	57
Table 2- Initial item pool and SME ratings for desires dimension from Study 1.....	58
Table 3- Initial item pool and SME ratings for expectations dimension from Study 1.....	59
Table 4- Perceptions items included in MTurk survey- difficulties and discriminations for Sample 1.....	60
Table 5- Desires items included in MTurk survey- difficulties and discriminations for Sample 1.....	61
Table 6- Expectations items included in MTurk survey- difficulties and discriminations for Sample 1.....	62
Table 7- Sample 1 correlation table for all 25 items included in MTurk survey.....	63
Table 8- Sample 1 EFA rotated factor matrix.....	65
Table 9- Sample 1 eigenvalues from EFA for actual and random data.....	66
Table 10- Sample 1 correlation table and reliabilities for underload dimensions- 25-item scale.....	67
Table 11- Perceptions items- difficulties and discriminations for Sample 2.....	68
Table 12- Desires items- difficulties and discriminations for Sample 2.....	69
Table 13- Expectations items- difficulties and discriminations for Sample 2.....	70
Table 14- Sample 2 correlation table for all 25 items included in MTurk survey.....	71
Table 15- Sample 2 CFA factor loadings for 3-factor 25-item scale.....	73
Table 16- Sample 2 discrepancy matrix for all 25 items included in MTurk survey.....	74
Table 17- Sample 2 correlation table and reliabilities for underload dimensions- 25-item scale.....	76
Table 18- Sample 2 CFA factor loadings for 3-factor 18-item scale.....	77
Table 19- Sample 2: Comparing CFA fit statistics for alternative models.....	78
Table 20- Sample 2 correlation table and reliabilities for underload dimensions- 18-item scale.....	79
Table 21- Means, standard deviations, skewness, and kurtosis for all variables.....	80
Table 22- Correlation matrix for all variables.....	Error! Bookmark not defined.

LIST OF FIGURES

Figure 1. Sample sizes for each phase of the study.....	82
Figure 2. Scree plots for actual data and randomly generated data.....	83

Introduction

The psychological literature has paid a great deal of attention to the notion of work-role overload as a source of work stress. Overload refers to having too much work to do, particularly during a specific timeframe (Kahn, Wolfe, Quinn, Snoek, & Rosenthal, 1964). However, another issue that has received some but far less attention in the organizational and occupational health psychology literature is work-related *underload*. Sales (1970) was one of the first to define role underload as “a social condition in which the focal person is faced with obligations which, taken as a set, require him to do considerably less than he is able to do in the time available” (p. 593). Frankenhaeuser and Gardell (1976) further distinguished between two types of underload—quantitative underload and qualitative underload. Quantitative underload involves not having enough work to complete in a specified amount of time, while qualitative underload involves being engaged in work which does not allow an individual to fully utilize his or her skills or talent. An employee experiencing qualitative underload may be kept busy throughout the work day, but the tasks that he or she completes are repetitive, tedious, or lack significance. These initial definitions of underload are supported in many of the subsequent studies that address this construct (Froggat & Cotton, 1984; Keenan & Newton, 1984; Shaw & Weekley, 1985).

Work-related underload is important to understand because prior research has indicated that there are a number of negative consequences associated with underload. For example, survey and experimental research has linked the experience of underload with psychological outcomes such as depression as well as physical health outcomes such as higher blood pressure (Melamed, Ben-Avi, Luz, & Green, 1995b). The consequences of underload at an organizational level are also significant, as underload is associated with higher levels of absenteeism and

turnover (Melamed, Ben-Avi, Luz, & Green, 1995a). Additionally, the fact that there are multiple potential sources of underload, including individual differences and leadership style, suggests that further research is needed to determine best practices when it comes to addressing underload. Seemingly simple proposed solutions such as giving employees more work to do will not necessarily address the underlying issues that lead to underload. In fact, Parasuraman & Hancock (2001) suggested that “coping with underload may prove to be an even bigger challenge than overload in the coming era of automation” (p. 317).

The purpose of this study is to develop the construct of work-related underload by defining, conceptualizing, and developing a measure of the construct. I will begin by describing psychological theories that provide a background for the study. Then I will review the prior research that has contributed to the limited literature regarding underload and propose a definition. Finally, I will detail the process of developing and evaluating a measure of underload.

Theoretical Background

Eustress and distress. Although stress is typically conceptualized as being a negative human experience, Selye (1987) defined stress as “the non-specific response of the body to any demand placed upon it” (p. 17). Selye also distinguished between eustress (i.e., “good” stress) and distress (i.e., “bad” stress). Eustress represents the optimal amount of stress that an individual experiences. The Yerkes Dodson law has historically been used to demonstrate this idea of eustress and distress, particularly related to work performance (Le Fevre, Matheny, & Kolt, 2003). Depicted by a curvilinear or inverted U-shaped diagram, the law posits that increasing stress will consequently increase an individual’s motivation to perform until a certain point at which an opposite relationship between stress and performance motivation occurs. While the Yerkes Dodson law continues to inform contemporary management practices related to

distress and eustress, Le Fevre et al. (2003) noted that the original study actually explored the relationship between a stimulus and task acquisition in mice. This was eventually equated to demands and performance, but there is no evidence to support the idea that these principles can be similarly applied to human employees and organizations. Le Fevre et al. (2003) suggest that research related to eustress should instead be tied to the concept of person-environment fit.

Person-environment fit. Person-environment fit occurs when the characteristics of an individual match the characteristics of the work environment (Kristof-Brown, Zimmerman, & Johnson, 2005). Fit may exist based on a number of individual characteristics such as traits, needs, goals, or abilities as well as a variety of environmental characteristics including job demands, working conditions, and rewards. Overall, the theory of person-environment fit posits that behavior is a function of both the individual and the environment (i.e., workplace) (Kristof-Brown & Guay, 2011). One type of person-environment fit, person-job fit, focuses on the match between an employee's characteristics and the characteristics of the job or task that they perform at work. More specifically, needs-supplies fit occurs when employees' needs, desires, or preferences are met by the jobs that they perform (Edwards, 1991). Further, complementary fit occurs through a match between an individual's talents and the corresponding needs of the environment (Muchinsky & Monahan, 1987). Although each of these types of fit are distinct, the presence of all types of fit are related to outcomes such as adjustment, well-being, and satisfaction (Edwards, 1991). Ideally, an employee and an organization should fulfill each other's needs through the performance of job tasks that are satisfying to the employee and beneficial to the organization.

Person-environment fit theory provides a basis for understanding characteristics of employees and organizations that may lead to perceptions of underload. Employees who desire

or prefer higher workloads but are not given enough work to do will experience a lack of needs-supplies fit, as their needs are not being met by the organization. Additionally, employees whose abilities exceed the demands of the organization will experience a lack of complementary fit. Although this situation reflects the definition of qualitative underload (Frankenhaeuser & Gardell, 1976), overqualified employees may complete tasks more quickly than those whose abilities match the demands of the job (Erdogan, Bauer, Peiro, & Truxillo, 2011). Thus, a mismatch between the characteristics of the person and those of the job or organization regarding preferences for workload and job demands is likely to lead to perceived quantitative underload.

Cognitive appraisal of stress. Underload is relevant to the stress-related literature through theories of cognitive appraisal of stress. Lazarus (1966) posited that individuals experience strain when they perceive that they cannot adequately cope with the demands being made on them. The cognitive appraisal process begins with a potentially stressful event or situation. First, the individual conducts a primary appraisal during which he or she evaluates the encounter to determine whether there is a potential harm or benefit to himself or herself. Next, the individual conducts a secondary appraisal to determine whether there is anything that can be done to prevent harm or increase benefits. As a result of the secondary appraisal, the individual may choose to engage in coping strategies that can either regulate stressful emotions (i.e., emotion-focused coping) or actively alter the situation that is causing distress (i.e., problem-focused coping). Ultimately, the cognitive appraisal process leads the individual to determine whether the situation could impact his or her well-being and whether it can be considered threatening or challenging (Folkman, Lazarus, Dunkel-Schetter, DeLongis, & Gruen, 1986). The implication of the cognitive appraisal of stressors is that not every individual will perceive the same event to be a stressor that leads to strain.

Lazarus and Folkman (1984) further addressed the threatening or challenging nature of stressors by distinguishing between hindrance stressors and challenge stressors. Cavanaugh, Boswell, Roehling, and Boudreau (2000) define challenge stressors as “work-related demands or circumstances that, although potentially stressful, have associated potential gains for individuals” whereas hindrance stressors “tend to constrain or interfere with an individual’s work achievement and...do not tend to be associated with potential gains for the individual” (p. 68). Challenge stressors have been associated with positive work-related outcomes such as higher learning motivation, performance, and job satisfaction as well as lower turnover (Widmer, Semmer, Kalin, Jacobshagen, & Meier, 2012). The idea of positive outcomes of stressors relates to Selye’s conceptualization of positive stress, or eustress.

The cognitive appraisal theory of stress is an appropriate model for examining underload because it emphasizes the fact that individuals perceive stressors in different ways. According to Lazarus (1966), the process begins with a *potentially* stressful event, suggesting that some events may be stressful for some individuals but will not elicit a stress response for others. The focus of the framework is on the evaluation of events and the perception of stress which varies between individuals. Thus, when measuring underload it is important to target subjective perceptions of workload rather than objective indicators. Further, the distinction between hindrance and challenge stressors is important because underload is likely perceived as a lack of challenge stressors rather than an excess of hindrance stressors. Although both of these conditions have negative implications for both employees and organizations, the outcomes associated with them are different. Thus, the outcomes related to underload will not necessarily be the same as those associated with underload.

Perceptions, expectations, and desires related to underload. Based on the psychological theories discussed above, the process underlying the construct of work-related underload should involve the employee's perceptions of underload as well as his or her expectations and desires related to workload. According to the cognitive appraisal process (Lazarus & Folkman, 1984) as well as person-environment fit (Kristof-Brown et al., 2005), two employees may have the same amount of work to do, but one employee may perceive that this workload is not ideal (i.e., there is too much or too little) and experience strain as a result, whereas the other may not perceive the workload as just fine. An individual may rely on information regarding his or her expectations and desires when determining whether the workload is too much or not enough.

Research related to the violation of psychological contracts further supports the idea that expectations and desires contribute to employees' perceptions regarding their experiences at work. Psychological contracts involve mutual expectations about reciprocal obligations within an employee-organization relationship (Levinson, Price, Munden, Mandl, & Solley, 1962). Rousseau (1989) further clarified that psychological contracts are primarily based on the expectations of the individual regarding reciprocal obligations and that these expectations are not always shared by the organization. When an organization does not fulfill these obligations, the employee may perceive that the psychological contract has been violated. Morrison and Robinson (1997) posit that "violation involves disappointment, frustration, and distress stemming from the perceived failure to receive something that is both expected and desired" (p. 231).

The theoretical background also provides an understanding of the outcomes that may be associated with underload. Under certain circumstances, high levels of workload can be

perceived as challenge stressors (Cavanaugh et al., 2000). Thus, underload is associated with a lack of challenge stressors and can subsequently lead to low levels of eustress. This has important implications for employee motivation and job satisfaction. Additionally, according to the general theory of person-environment fit, an employee who wants to contribute to an organization through his or her specific job but is not given the opportunity to do so would experience a lack of needs-supplies fit. A lack of complementary fit can also occur through the experience of underload if an employee is not given enough work to do because his or her talents exceed the needs of the organization.

Prior Empirical Research Related to Underload

Although the existing body of empirical research on underload is somewhat limited, several studies have found relations between underload and negative psychological and physical outcomes. Survey data collected across a variety of occupations including contractors, service and sales workers, and blue-collar workers have provided evidence that quantitative underload is associated with job dissatisfaction, depression, and psychological distress (Ganster, Fusilier, & Mayes, 1986; Melamed, Ben-Avi, Luz, & Green, 1995a; Shultz, Wang, & Olson, 2010). Additionally, Melamed, Ben-Avi, Luz, & Green (1995b) used survey data from the Cardiovascular Occupational Risk Factors Determination in Israel (CORDIS) study to examine overload and underload. CORDIS was the first study to find negative health outcomes associated with underload, including higher systolic blood pressure and higher cholesterol.

Lab studies have also been conducted to examine the relationship between underload and negative outcomes. Sales (1970) studied undergraduate students in conditions of underload and overload in a lab setting. Participants were given sets of anagrams to complete in five minute intervals. Those in the overload condition were given 35% more anagrams than they were able to

decode within the five minute period, and those in the underload condition were given a smaller amount of anagrams so as to keep them waiting for 30% of the time spent in the lab. Measures included heart rate, task enjoyment, tension and anger, and reported attempts to increase productivity. Findings of the study indicated that the underloaded participants experienced less enjoyment of the task and reported less attempts to increase productivity than overloaded participants. This suggests that conditions of underload at work may not provide meaningfulness and fulfillment for employees and may result in lower productivity.

Shaw and Weekley (1985) conducted a similar lab study using anagrams to explore the effects of overload and underload as well as transient workloads. The underload and overload conditions were designed the same way as in the Sales (1970) study, and the transient workload condition alternated periods of underload and overload. The researchers measured performance, task enjoyment, resentment, and affective states including anxiety, depression, and hostility. Negative outcomes were associated with qualitative underload; however, there were no significant findings for quantitative underload. The lack of significant findings can be attributed to certain limitations of the study including the nature of the task used to create conditions of overload and underload. Although it has been convenient to use anagrams as a task because of the objectivity of completion, these tasks are not representative of what employees are responsible for in the workplace. Study participants may find it difficult to become invested in the anagram task because there is no connection to any purpose greater than that of the study itself. Additionally, a limitation across both survey and lab studies of underload may be the improper measurement of the construct of underload. After providing additional clarity regarding the definition of underload, I will review and critique how underload has been measured in the research literature.

Defining Underload

Before establishing a measure of underload, it is important to establish a clear definition that will be used in this study and can be set as the standard for future studies. This study specifically aims to provide a definition and measure for the quantitative component of underload. Although the qualitative component of underload is related to quantitative underload and has important implications (Frankenhaeuser & Gardell, 1976), they are fundamentally different experiences and should therefore be associated with different outcomes and implications. Given that qualitative underload has been extensively researched within the underemployment literature, this study focuses on quantitative underload.

Previous definitions of underload focus on the idea that an individual does not have enough work to complete. Although I agree that the employee's subjective evaluation of his or her workload is a critical aspect of the experience of underload, I propose that the desires and expectations of the individual play a role in this process as well based on the theoretical background discussed in the above sections. Thus, underload is defined as a condition in which an individual perceives that he or she does not have enough to do at work. The desires and expectations of an employee related to his or her workload contribute to perceptions regarding work-related underload. This definition guides the development of a new measure of work-related underload.

Measurement of Underload

Although the concept of underload has been present in the literature for decades, few measures have been established thus far, and little information about the reliability and validity of these measures has been offered. In Sales's (1970) original lab study, a single item measure was used to assess subjective workload (i.e., "How much work did you feel you had to do during

the work period?”). Shaw and Weekley (1985) used similar measures to assess underload in their lab study (i.e., “I could have easily solved many more anagrams in the allotted time”). While these measures may address the fact that participants did not have enough work to do in a single specified period of time, single-item measures may not be useful in measuring underload that occurs over time. Additionally, single-item measures are often insufficient methods of measuring constructs, and their internal consistencies cannot be estimated (Nunnally, 1978). A participant may indicate that they did not have much work to do during the work period or that they could have easily solved more anagrams in the allotted time, but the extent to which an employee is experiencing work-related underload over time may not be captured in one item.

Shultz et al. (2010) used a single item measure to assess underload as well (i.e., “How well do you think your skills matched the demands imposed on you by your job?”). Rather than assessing quantitative underload, this measure captures the qualitative underload that participants might experience. The item does not measure underload as defined in this paper because someone who indicates that their skills do not match the demands of the job may either be underqualified or underemployed. In either case, the item response does not tell us anything about the amount of work that the individual is given.

In contrast to single item measures, several lengthier scales have been developed in previous studies. For example, Caplan, Cobb, French, Harrison, & Pinneau (1975) developed an 11-item scale to measure quantitative underload. The scale assesses the amount of work that individuals are assigned, how often they are assigned this amount of work, and the pace at which they are required to work. The items assess the quantitative workload more extensively than the single-item measures, although the scale includes several items that do not seem to measure underload effectively. For example, one item asks “How many lulls between heavy workload

periods do you have?” with a response scale ranging from “hardly any” to “a great deal.” This item would not effectively distinguish between an individual who had several heavy workload periods with a large amount of down time in between and an individual who did not experience any heavy workload periods at all. This is an important distinction to make when considering the negative outcomes associated with underload, thus Caplan et al.’s (1975) scale should be improved in order to successfully measure underload.

Additionally, Froggat and Cotton (1984) used a three-item measure as a manipulation check for subjective workload in their anagram lab study. Two of the items are similar to those mentioned above in that they assessed the quantitative workload, and the third item asks about the extent to which participants felt that the rate at which work was assigned should have been increased or decreased. This item addresses one aspect of the participants’ expectations, as someone who desires or expects more work will indicate that they feel the rate of work should be increased. In this sense, Froggat and Cotton’s (1984) scale captures an additional dimension of the process that may lead underloaded employees to experience negative outcomes. However, the item taps content domain outside the scope of the definition of underload and should therefore not be included in a measure of the construct.

Based on a review of previously developed scales, the measurement of underload should assess the quantitative nature of the workload that an individual is assigned. I will focus on measuring quantitative underload rather than developing a scale that will measure qualitative underload because this concept is more closely aligned with the constructs of underemployment and overqualification for which several measures have already been established (i.e., Scale of Perceived Overqualification; Maynard, Joseph, & Maynard, 2006). What seems to be missing is a multi-item scale in which every item captures the extent to which an individual experiences

quantitative underload. Although some of the scales discussed in the above section seem promising, several are single-item measures and others include items that do not seem to truly assess underload as defined in this paper. Additionally, we cannot truly evaluate the effectiveness of any of these scales without evidence of reliability or validity. Finally, previous measures have not addressed the components of desires or expectations related to workload which play an important role in the perception of work-related underload. In this study, I draw on existing measures of underload (i.e., Caplan et al.'s (1975) 11-item measure) and create new items that are based on a theoretically-based definition of underload. I use my knowledge of previous research related to underload and exploring the nomological net of the construct to ensure that the new scale captures this definition of underload. See Appendix A for a full list of items from these existing measures of underload.

Nomological Net

In order to fully understand underload as a meaningful construct, it is helpful to explore the context of the nomological net by reviewing similar and distinct constructs.

Overload. Overload has been recognized as a stressor in the workplace for many decades (Kahn et al., 1964; Sales, 1970; Keenan & Newton, 1984; Shultz et al., 2010). Referring to an excess amount of assigned work, overload is often conceptualized as the opposite of quantitative underload. Given this inverse relationship, it may be tempting to assume that opposite outcomes would occur in conditions of overload and underload. In other words, it may be incorrectly inferred that because role overload often leads to negative outcomes such as strain and low job satisfaction, role underload would be related to positive outcomes including low strain and high job satisfaction. However, the literature reviewed in previous sections of this paper suggests that similar outcomes can occur as a result of both role overload and role underload (Sales, 1970;

Shultz et al., 2010). Therefore, overload should be negatively related to underload but should be associated with similar consequences.

Boredom. Boredom is a topic that has been explored in the literature and that has direct links to quantitative underload. Fisher (1991) posited that boredom occurs when an individual subjectively appraises an activity as deficient when compared to the type of activity that he or she desires. Thus, these two concepts are similar because they involve an expectation or desire that is not met by the stimulation provided. Fisher (1987) conducted a qualitative study that investigated incidents of boredom both on and off the job. Fifty-five percent of participants' reported incidents of boredom on the job were attributed to "having nothing to do," the key characteristic associated with underload. Many respondents in the study reported that they engaged in non-work activities as a way of coping with boredom at work. Dysfunctional responses to boredom at work include unsafe work practices, excessive horseplay, sabotage, or employee theft (Fisher, 1991). Consistent boredom at work may lead to more detrimental emotional outcomes including restlessness, agitation, and emotional distress (Robinson, 1975).

Fisher (1991) explored further evidence of the long-term consequences of frequent boredom at work as well. Job satisfaction, absenteeism, and turnover were all found to be outcomes of consistent boredom at work. Additionally, physical and mental health outcomes of boredom include cardiovascular disease, psychosomatic disease, learned helplessness, and reduced intellectual flexibility. These findings highlight the potential for quantitative underload to have significant negative effects on employees.

Although we can identify the similarities between underload, disengagement, and boredom, it is important to note that these constructs are in fact distinct. The essential difference that distinguishes these constructs is that underload is by definition a direct consequence of a

lack of assigned duties or tasks at work. While boredom and disengagement can be consequences of underload, they can also both occur during conditions of overload. For example, an employee who has too much work to do can become bored with their work and can therefore disengage from their role, but this employee would not be experiencing underload.

Underemployment. Underload is also different than the construct of underemployment. Previous literature that distinguished between quantitative and qualitative underload (e.g., Frankenhaeuser & Gardell, 1976) defined qualitative underload as a condition in which an employee's skills are underutilized. I propose, however, that this concept is not distinct from that of underemployment, which by definition occurs when the skills and abilities of an employee do not match the demands of the job (McKee-Ryan & Harvey, 2011). Thus, the idea of quantitative underload is a better way to define the construct and is clearly distinct from underemployment.

Overqualification. Similar to the concept of underemployment, overqualification occurs when an employee's education, skills, or work experience exceed the requirement of the job (Maynard et al., 2006). Perceived overqualification, as opposed to objective overqualification, is often measured in the literature because negative outcomes such as turnover and counterproductive work behaviors are more strongly associated with employees' perceptions of overqualification rather than an objective comparison of their skills and those required by the job (Luksyte, Spitzmueller, & Maynard, 2011). In this sense, overqualification and underload are similar because the process leading from these conditions and the negative outcomes that they are associated with depends on the perceptions of the employee. Despite this similarity, overqualification is more closely related to the idea of qualitative underload and is therefore distinct from the construct of quantitative underload.

Potential Sources of Underload

Underemployment. The recent global economic recession has led to an increase in unemployment and competition for jobs and in some cases has created unique situations for individuals seeking employment during this time. As financial reasons for seeking employment become more influential, job seekers may focus less on finding positions that meet their intellectual and motivational needs and settle for a job that will pay the bills. As a result, more employees may find that their skills and abilities exceed the demands of the job in which they are placed—a concept known as underemployment (McKee-Ryan & Harvey, 2011). McKee-Ryan and Harvey (2011) reported statistics from a Gallup report in May 2010 which used the percentage of the U.S. workforce that was employed part time but desired full time work to estimate the rate of underemployment at 9.7%. The same Gallup report in January 2014 estimated the underemployment rate in the U.S. at 10% (Gallup, 2014). These results provide evidence that while the unemployment rate slowly decreases, rates of underemployment continue to increase.

Although underemployment is a concept that is distinct from underload, quantitative underload may be experienced as a consequence of underemployment. When employees perform tasks that are below their levels of ability, they may complete these tasks quickly and find themselves without enough work to fill the time that they spend at work. Alternatively, underemployed individuals may have the mental capacity to juggle several challenging tasks at once, but the nature of the job may determine a slower working pace. In this case underemployed individuals would likely experience underload as they might desire a heavier workload to keep themselves busy throughout the day.

Low workload. There are several other potential causes of underload that may come from different sources in the workplace. First, it is possible that there is truly a lack of work to be

completed in the organization. For example, consulting firms that rely on contracts may experience slower periods during which there are a limited number of projects to complete. During this time, full-time employees are likely still required to work the same amount of hours, but the amount of work to complete is significantly reduced. Another possibility is that managers in an organization may not delegate enough work to employees. Employees who are new to an organization may be particularly susceptible to this source of underload because they have not yet had an opportunity to gain experience and demonstrate the ability to complete certain tasks. This may cause managers to be hesitant when assigning projects and may limit the amount of work that these employees have to do.

Role ambiguity. Role ambiguity is another potential source of underload. Role ambiguity refers to a lack of clarity about expected behavior within a job (Kahn et al., 1964). When employees do not have a clear understanding of their responsibilities at work, they may not complete certain tasks that take up time in the work day. For example, an employee in a retail store may not be aware that he is responsible for performing certain tasks to prepare the store for closing each night. This could be due to a variety of reasons including poor communication between managers and employees. This employee may spend the last hour of the work day experiencing underload because he is not aware that he should be counting the cash in the drawers, restocking empty shelves, and locking storage closets, and he has not received the proper training to do so. These are just a few examples of the many ways in which certain behaviors or practices may lead to quantitative underload.

Significance of the Study

Previous literature has suggested that underload can be a significant source of strain in the workplace. However, without an established definition and method of measuring the

construct, we cannot ensure that we are providing theoretically and statistically sound evidence to support this idea. A theoretically developed and empirically tested measure of underload is needed in order to further explore the issue of underload. Previously developed measures of underload have not been grounded in theory and have therefore been conceptualized in a variety of ways without consensus, making it difficult to build a body of empirical evidence to support the construct. Building a measure that is informed by theories of cognitive appraisal, person-environment fit, and psychological contract violation will provide a meaningful, multi-dimensional conceptualization of underload and will set the stage for future empirical research to test relationships between underload and outcomes such as psychological strain in the workplace.

Additionally, a validated measure of underload will have important practical implications for organizations. The ability to identify underload as a source of stress in the workplace will help organizations to better address worker health and well-being, and future research can examine how job design, leadership, communication, and training can all be used to prevent situations in which underload is likely to occur. Further, evidence of the roles that expectations and desires play in the experience of underload will have important implications for selecting the right employees for job positions and managing expectations regarding workload. Recent research suggests that automation of jobs will continue to increase in the future and will replace many tasks that employees previously were responsible for completing (Autor, 2015). This change suggests that the prevalence of underload is likely to increase as well, although it is difficult to assess the prevalence without a psychometrically supported measure.

Developing a Measure of Underload at Work

While previous studies have taken a limited perspective of underload, I propose that work-related underload is comprised of three distinct dimensions: subjective perceptions of

underload, desires related to workload, and expectations related to workload. These dimensions are based on the theoretical framework that contributes to the concept of underload as a stressor. In order to confirm that underload is made up of these three dimensions, I will examine the correlations between the factors. The factors should not correlate strongly, as this would indicate that they are not in fact distinct (Raykov & Marcoulides, 2011). Additionally, the three-factor model should fit the data better than a single factor model. This would provide further evidence of the existence of three dimensions that make up underload.

H1: Underload is comprised of three distinct dimensions: perceptions, desires, and expectations.

To establish underload as a unique construct, I will explore underload within the context of the nomological net discussed in previous sections. Assessment of convergent and discriminant validity is an important step in the construct validation process (Cook & Campbell, 1976; Guion, 1998). Convergent validity refers to finding significant positive relationships among similar constructs, and discriminant validity refers to finding negative or insignificant relationships among constructs that are not similar. Because not having enough to do at work is often an antecedent of boredom (Fisher, 1991), underload should be positively related to boredom. Empirical evidence shows that underload is a prevalent cause of boredom in the workplace (Fisher, 1991). Thus, these constructs should be related, but not redundant. Theories of person-environment fit suggest that a mismatch between the characteristics of an individual and the characteristics of the work environment can lead to negative outcomes in the workplace (Kristof-Brown et al., 2005). According to these theories, underload and underemployment should be related because they both involve a mismatch between an employee's abilities and the expectations of the organization or specific role. Further, existing literature related to underload

classified underload and underemployment as two related dimensions—the former referring to the quantitative nature of the workload and the latter referring to the qualitative nature of the workload (Frankenhaeuser & Gardell, 1976). However, underload and underemployment are distinct constructs and therefore should not be strongly related. Another important part of assessing convergent validity is evaluating the relationship between the new measure and previous measures that aim to assess the same construct (Murphy & Davidshofer, 1991). Although Caplan et al.'s (1975) scale of underload does not include a comprehensive assessment of the three-dimensional definition of underload, it should be positively related to the new measure of underload. Anastasi and Urbina (1997) posit that there are not specific cutoff scores regarding validity coefficients but that convergent validity is generally indicated by moderate to strong correlation coefficients (e.g., .3-.5).

***H2a:** All three dimensions of underload (perceptions, desires, and expectations) are moderately but not strongly positively related to boredom.*

***H2b:** All three dimensions of underload (perceptions, desires, and expectations) are moderately but not strongly positively related to underemployment.*

***H2c:** Previous measures of underload are strongly positively related to all three dimensions (perceptions, desires, and expectations) of the new measure of underload.*

Discriminant validity evidence can include a negative or non-significant relationship between constructs that are theoretically dissimilar (Murphy & Davidshofer, 1991). Overload refers to an excess amount of assigned work and can therefore be conceptualized as the opposite of quantitative underload. Thus, overload and underload should be negatively related.

Additionally, there are no theoretical similarities between underload and social desirability, and

they should not be significantly related. Discriminant validity is generally indicated by weak correlations (e.g., .1-.2; Anastasi & Urbina, 1997).

***H3a:** All three dimensions of underload (perceptions, desires, and expectations) are strongly negatively related to overload.*

***H3b:** All three dimensions of underload (perceptions, desires, and expectations) are not related to social desirability.*

It is also important to assess the criterion-related validity of a measure by evaluating the empirical associations with other criteria that should be related to the construct (DeVellis, 2012). I will specifically assess the concurrent criterion-related validity of underload because all constructs are measured at the same point in time. According to job characteristics theory, employees who experience low skill variety, task identity, or task significance find it more difficult to reach critical psychological states including experienced meaningfulness of work and are less likely to be satisfied in their jobs. Thus, I expect that underload will be negatively related to job satisfaction and meaningfulness of work. Additionally, because role ambiguity is a potential source of underload, I expect that they will be positively related. Finally, theories of cognitive appraisal of stress, person-environment fit, and psychological contracts suggest that an employee who perceives a low workload, expected more work, and wants to have more work would perceive underload as a stressor. Thus, there should be a positive relationship between underload and perceptions of stress.

***H4a:** All three dimensions of underload (perceptions, desires, and expectations) are negatively related to job satisfaction.*

***H4b:** All three dimensions of underload (perceptions, desires, and expectations) are negatively related to meaningfulness of work.*

***H4c:** All three dimensions of underload (perceptions, desires, and expectations) are positively related to work role ambiguity.*

***H4d:** All three dimensions of underload (perceptions, desires, and expectations) are positively related to perceived work stress.*

Using accepted best practices for scale development within the field of industrial/organizational psychology (e.g., Hinkin, 1995; DeVellis, 2012), I will complete the necessary steps to test the above hypotheses. These steps include generating, testing, and evaluating the items that will make up the underload scale. I will then complete the statistical analyses to test the structure of the measure and collect evidence of reliability and validity. The steps of this process are detailed in the following sections.

Study 1: Initial Item Development and Review

The purpose of Study 1 was to develop an initial pool of items for the underload scale and to conduct a review and evaluation of the new items based on feedback from subject-matter experts.

Study 1- Method

Participants. Participants included subject matter experts (SMEs) who reviewed the initial pool of items for the underload scale. As per recommendations from Lindell, Brandt, & Whitney (1999), a sample of eleven industrial/organizational psychology doctoral students who had successfully completed a course in measurement and test construction were chosen as SMEs because of their knowledge of measurement principles and psychological theories related to work and underload. The SMEs received a survey containing each of the original items for the three dimensions of the scale and were asked to rate the relevance of each item to the construct based on the definition provided to them. They were also given the opportunity to comment on the general clarity and content of the items.

Procedure. Using DeVellis (2012) and Hinkin (1995) as guides for scale development, I completed the following steps:

Determine the purpose of the scale. DeVellis (2012) stresses the importance of considering the purpose of the scale before devoting time to generating items and collecting data. As detailed in the introduction, I developed a measure of underload that is based on the theoretical background related to stress and other work processes. Thus far, there have not been any validated measures of underload that follow a clear definition. In this initial step, it is also important to consider how specific the scale will be. While previous studies have measured

underload in both a qualitative and quantitative sense, my measure considers the quantitative nature of the experience of underload. DeVellis (2012) also suggests that scale developers should consider whether the construct is distinct from other constructs. Exploring the nomological net and including measures of these related constructs in the study provide evidence that underload is a unique construct, and the theoretical background suggests that this is true as well. Overall, I believe that my extensive review of the literature related to underload as well as relevant theories have guided the determination of the purpose of the scale.

Generate items for the scale. Hinkin (1995) posits that the primary concern in this stage is the content validity of the scale. Two methods of item generation are proposed—the deductive method (i.e., using theory to guide item generation) and the inductive method (i.e., using individual responses rather than theory to guide construct identification). I used the deductive method to generate items for the underload scale. This is the same approach outlined in DeVellis (2012).

I began with a pool of 30 items that reflect the proposed three-dimensional construct of underload. I used the theories discussed above to guide the development of the initial items. In accordance with Lazarus's (1965) cognitive appraisal of stress model, I wrote items for the perceptions dimension with the purpose of assessing subjective perceptions of workload in relation to how participants evaluated the way in which they spent their time at work. I also used person-environment fit theory to guide the item writing for the expectations and desires dimensions, focusing on assessing a lack of complementary and needs-supplies fit. When writing these items, I referenced previous measures of underload (i.e., Caplan et al., 1975), as some of the items from these scales reflect the concept of perceptions of quantitative underload. I also consulted with other individuals within the field of industrial/organizational psychology in order

to write new items that reflect the definition of underload. DeVellis (2012) suggests beginning with a pool of items that is three to four times larger than the number of items to be included in the final scale. Given that the scale is three-dimensional and that each dimension should contain at least three items in order to calculate reliability estimates, my initial item pool should have contained at least 9 items per dimension (Raykov & Marcoulides, 2011). Input from the initial sample of SME participants assisted in the process of choosing the best possible items for the final scale.

In the process of writing the items, I followed DeVellis's (2012) guidelines regarding the characteristics of good and bad items. I avoided writing items that were too long or too difficult in terms of reading level. I also ensured that the items were not double-barreled and did not include ambiguous statements. Finally, the items were consistently positively worded to reduce confusion for the respondents (Cordery & Sevastos, 1993; Greenberger, Chen, Dmitrieva, & Farruggia, 2003).

Determine the response format. In the next step, I chose a response format for the scale items. DeVellis (2012) discusses many different options for response formats including Thurston scaling, Guttman scaling, and Likert scaling. Hinkin (1995) suggests that 5 or 7 point Likert scales should be adequate to use for most measures. Additionally, Dawes (2008) found that in terms of mean scores and variance, data obtained using 5-point scales and those obtained using 7-point scales are not significantly different, and the distinction between response choices is clearer for 5-point scales when compared to 7-point scales. Likert scales that allow participants to indicate the extent to which they agree with a statement are often used to measure opinions, beliefs, and attitudes; thus, I used a 5-point Likert scale for the expectations and desires factors

(Floyd & Widaman, 1995). The response options are as follows: 1- *strongly disagree*, 2- *disagree*, 3- *neither agree nor disagree*, 4- *agree*, and 5- *strongly agree*.

For the measure of quantitative underload, it is important to consider the time frame in which participants should be thinking about their experiences with underload. Thus, instead of asking them to agree or disagree with a statement, the response format asks them to indicate how often certain events related to underload occur at work. I used a 5-point Likert scale with the following options: 1- *less than once per month or never*, 2- *once or twice per month*, 3- *once or twice per week*, 4- *once or twice per day*, and 5- *several times per day*. This response format has been used in other scales to assess workload such as the Quantitative Workload Inventory (Spector & Jex, 1998). This should be an appropriate way to distinguish between different levels of perceived quantitative underload. Additionally, the instructions for the items specify that participants should think about their experiences at work over the past month. This will limit the timeframe in which participants are reflecting on, reducing potential errors associated with attempting to remember less salient events (Eisenhower, Mathiowetz, & Morganstein, 2011).

Experts review the initial item pool. Schwab (1980) suggests that a group of experts should review the initial item pool of the measure. As detailed in previous sections, a sample of eleven Industrial/Organizational Psychology graduate students reviewed the initial items for the scale. They received a survey in which the definitions of each of the three dimensions were presented, followed by each item within the specified dimension. The SMEs rated the relevance of each item to the respective dimension of the construct based on the definition provided. SMEs rated relevance on a response scale ranging from 1 (*Not at all Relevant*) to 5 (*Extremely Relevant*). They were also given the opportunity to provide qualitative input for each item regarding the quality the items including the wording, clarity, and length of items. Finally, the

experts were able to suggest any additional items that could be added to the scale in order to fully represent the construct as defined as defined, and they were given the opportunity to provide additional comments regarding the items or measure in general. This is a common practice in the scale development process (Polit & Beck, 2006).

Study 1- Results

I began by assessing inter-rater agreement of the subject matter experts regarding the relevance of each item to the specified dimension for each of the 30 underload scale items. I assessed inter-rater agreement using James, Demaree, and Wolf's (1984) single-item r_{wg} index. This is an appropriate method for assessing inter-rater agreement when multiple raters provide ratings for a single variable using an interval scale (LeBreton & Senter, 2008). The r_{wg} value is calculated by dividing the observed variance on the variable by the variance expected when there is complete lack of agreement among the judges and subtracting this number from one. The mean relevance rating and r_{wg} values for each original item are shown in Tables 1-3. Each item received at least a mean rating of 3, indicating that they were moderately relevant. After reviewing the relevance ratings and qualitative comments from the subject matter experts, I made several changes to the original items. Some items did not change, some were reworded to increase clarity, some were removed completely, and two new items were added to the desires dimension. The final list of items used in the MTurk survey are displayed in Appendix B.

Study 2: Scale Evaluation

The purpose of Study 2 was to test the items developed in Study 1 with a development sample in order to assess the psychometric properties of the measure and gather evidence of the reliability and validity of the scale.

Study 2- Method

Participants. Working U.S. adults were targeted for the development samples of the study using Amazon's Mechanical Turk (MTurk). MTurk is an online labor market in which workers complete surveys and human intelligence tasks (HITs) for compensation by requesters. MTurk is increasingly being used for data collection in social sciences because it is a fast, inexpensive, and generally reliable way to obtain data from human subjects (Buhrmester, Kwang, & Gosling, 2011). Numerous studies that have compared data obtained through MTurk to other data collection methods, such as in-person student samples, have found similar results in terms of participant commitment and attention as well as overall data quality (Casler, Bickel, & Hackett, 2013; Goodman, Cryder, & Cheema, 2013; Mason & Suri, 2012). Data from MTurk have also shown evidence of reliability estimates that are similar to those obtained through other sampling methods (Buhrmester et al., 2011). Regarding demographics, MTurk workers are generally representative of the overall U.S. population, but they have lower income and higher levels of education (Paolacci, Chandler, & Ipeirotis, 2010).

A total of $N = 1,210$ MTurk workers completed the initial screening survey containing questions about demographics and work experience. Of the pool of workers who were screened, 842 participants were eligible to take the full survey based on the specified criteria (i.e., at least 18 years of age, job tenure of at least one month, working at least 20 hours per week, not self-

employed). The full survey was made available to those 842 workers who were qualified to take it, and data was collected from 412 participants who chose to complete the survey. Four participants were removed from the sample due to incorrect responses on insufficient effort response items (i.e., “Please respond ‘strongly agree’ to this item). Additionally, although they indicated otherwise on the screening survey, 16 participants who took the full survey either identified as self-employed or indicated working less than 20 hours per week and were subsequently removed from the sample. Finally, five participants were eliminated because their response times on the survey were significantly lower than the average response time. Thus, a final sample of $N = 387$ participants was retained for analysis. See Figure 1 for a visual display of sample sizes for each phase of the study.

The final analysis sample contained 172 females (44%), and 215 males (56%) with an age range of 18 to 70 and a median age of 30 ($M = 33.5$, $SD = 10.5$). Within the sample, 14% had a master’s, doctoral, or professional degree, 52% had a baccalaureate or associate’s degree, and 34% did not have a degree. Participants reported a wide variety of job titles within several industries including service (16%), education (13%), information technology (10%), and medical (10%). The sample also included supervisors and higher-level employees (39%) as well as non-supervisors (61%).

Descriptive statistics for Sample 1 and Sample 2. The total sample was randomly split using random number generation in Microsoft Excel to form two separate samples—sample 1 ($N = 193$) and sample 2 ($N = 194$). The purpose of splitting the sample was to conduct an exploratory factor analysis on sample 1 and a confirmatory factor analysis on sample 2. Exploratory and confirmatory analyses should not be conducted using the same sample and

should ideally be performed using two independent samples, although it is common to conduct a cross-validation by randomly splitting a dataset in half (Hurley et al., 1997; Brown, 2006).

Sample 1 consisted of 85 females (44%) and 108 males (56%) ranging in age from 18 to 70 with a median age of 31 ($M = 33.52$, $SD = 10.34$). Within the sample 13% had a master's, doctoral, or professional degree, 54% had a baccalaureate or associate's degree, and 33% did not have a college degree. Participants' jobs were across industries including education (15%), service (15%), medical (12%), and information technology (11%). Job levels ranged from supervisors and higher-level employees (37%) to non-supervisors (61%).

Sample 2 included 87 females (45%) and 107 males (55%) ranging in age from 19 to 68 with a median age of 30 ($M = 33.52$, $SD = 10.75$). The demographics were similar to sample 1 regarding education, as 14% had a master's, doctoral, or professional degree, 50% had a baccalaureate or associate's degree, and 36% did not have a degree. Job levels and industries were also similar to sample 1—40% were supervisors or higher-level employees and 60% were non-supervisors, and participants worked in industries such as service (17%), education (12%), information technology (9%), and medical (8%).

I conducted a series of chi-square tests to check for significant differences in proportions of characteristics between sample 1 and sample 2. The tests revealed that the characteristics of the two samples were not significantly different based on gender, $\chi^2(1, N = 387) = .03$, $p = .87$; education, $\chi^2(7, N = 387) = .6.94$, $p = .44$; job level, $\chi^2(6, N = 387) = 3.91$, $p = .69$; or industry, $\chi^2(16, N = 387) = 10.88$, $p = .82$. I also conducted an independent samples t-test to assess differences in the mean age of the two samples, and the differences were not significant, $t(385) = .15$, $p = .88$.

Procedure. I continued to follow the steps for scale evaluation as discussed in DeVellis (2012):

Consider the inclusion of validation items. In order to assess the construct validity of the scale, I included measures of the constructs that I expect to be related to underload in the survey administered to the participants. These constructs are discussed in the *Nomological Net* section above (e.g., job boredom and underemployment). Examining the relationships between underload and these constructs and testing my hypotheses provided evidence that supports the construct validity of the scale (DeVellis, 2012). The specific scales used to measure these additional variables are discussed in the *Measures* section below.

Administer items to a development sample. In order to obtain a sample of working adults in the U.S., I administered an initial screening test via MTurk. This survey contained demographic questions and specifically asked participants how many years of work experience they have, how many hours a week they currently work for pay, how long they have been in their current job (outside of MTurk), and whether they identify themselves as being self-employed. I used the geographical tracking feature available through MTurk to ensure that the survey was only available to participants with an IP address within the United States. The target sample size was 400 participants based on the number of items in the final pool for the scale, as well as based on a power analysis assuming an alpha of .05 and a Pearson correlation coefficient (r) value of .20. Floyd and Widaman (1995) suggested that 10 participants should be recruited for each item in the scale. I obtained a sample that was large enough to randomly divide into two samples in order to cross-validate the factor structure using a holdout sample.

In order to take steps to ensure the quality of the data, I restricted participation to participants with an approval rate of at least 95%. The approval rate reflects the percentage of

previously completed HITs for which the worker's responses were accepted and approved by the requester of the task. Ninety-five percent is the default cutoff approval rating on MTurk, and this restriction has been used in previous studies to successfully increase the quality of the data collected (Goodman et al., 2013). Additionally, several items were included in the survey to check whether participants seem to be paying attention to the survey items. For example, one item asked participants to indicate "strongly agree" on the response scale. Participants who responded incorrectly to these items were screened out of the respondent pool. As a final step to ensure data quality, a small sample of individuals took the survey before it was administered on MTurk in order to gain a sense of the amount of time that the average participant should take to complete the survey. MTurk participants who took significantly less time to complete the survey were also eliminated.

Prior research has suggested that the average reservation wage for MTurk workers (i.e., the least amount of compensation that they must be offered in order to complete the task) is \$1.38 per hour of work (Chilton, Horton, Miller, & Azenkot, 2010). All survey respondents were compensated \$0.10 for completing the initial screening survey based on the amount of time required, but only those respondents in the U.S. who are over the age of 18, have been in their current jobs for at least one month, currently work 20 or more hours per week, and are not self-employed were qualified to complete the full survey. Once these targeted participants were identified, they received access to the full survey. The full survey contained the same demographic questions as the screening survey as well as all measures detailed in the *Measures* section below. Responses to demographic questions from the screening survey and final survey were compared, and respondents with inconsistent responses were eliminated from the dataset.

Evaluate the items. The next step in the scale development process as outlined in DeVellis (2012) is to evaluate the items. In order to complete this step, I performed several statistical analyses including basic item analysis, exploratory factor analysis, and confirmatory factor analysis. These steps are discussed in detail in the *Statistical Analyses for Scale Evaluation* section below.

Optimize scale length. DeVellis's (2012) final step in the scale development process involves determining the optimal length of the scale. While a scale that is too long is often too much of a burden on the respondents, a scale that is too short may lack reliability (DeVellis, 2012). Stanton, Sinar, Balzer, & Smith (2002) suggested several steps for reducing the length of a scale, beginning with the assessment of internal and external quality indices. The item-total correlations are indications of internal validity, and the average item-criterion correlations for each item are indications of external validity. Additionally, SMEs should rate the face validity of each item. Then, if the scale needs to be reduced, the items can be sorted by the quality indices in order of external indices, internal indices, and face validity indices. Finally, the top quality items can be chosen from the scale, and the validity from the full scale and reduced scale should be compared (Stanton et al., 2002). I describe each step taken to optimize the scale length in the *Scale Reduction* section below.

Measures. The following measures were included in the survey:

Underload. Underload was assessed using the scale developed in this study and included three dimensions—perceptions, expectations, and desires. High scores on the perceptions dimension indicate that the participant perceives a low workload, and low scores indicate a high workload. High scores on the expectations dimension indicate that the participant expected to have more work than they currently have, and high scores on the desires dimension indicate that

the participant wants to have more work than they currently have. The alpha coefficient is $\alpha = .97$ for the total scale, $\alpha = .94$ for the perceptions dimension, $\alpha = .97$ for the desires dimension, and $\alpha = .96$ for the expectations dimension.

Boredom. Boredom was measured using an adapted format of Lee's (1986) Job Boredom Scale ($\alpha = .93$). The Job Boredom Scale consists of 17 items with a Likert response scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). A sample item is "My job goes by too slowly." The full list of items is presented in Appendix C.

Underemployment. Underemployment was measured using the Scale of Perceived Overqualification (SPOQ; Maynard et al., 2006). The SPOQ is comprised of 9 items with a Likert scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). Overqualification is a dimension of underemployment, and is of particular interest for the purpose of the current study because it refers to a mismatch regarding an employee's knowledge, skills, or abilities (Luksyte et al., 2011). The SPOQ showed high internal consistency reliability evidence in this study ($\alpha = .93$) and has been endorsed as a theoretically sound method of measuring underemployment (McKee-Ryan & Harvey, 2011). A sample item is "I have job skills that are not required for this job." See Appendix D for the full scale.

Role ambiguity. Role ambiguity was assessed using an adapted version of the Breugh and Colihan (1994) Role Ambiguity Scale. The scale consists of 6 items with a Likert scale ranging from 1 (*disagree strongly*) to 7 (*agree strongly*). The two subscales of the adapted measure are work method ambiguity and performance criteria ambiguity. The internal consistency of the measure in this study was $\alpha = .92$. A sample item is "It is clear to me what is considered acceptable performance by my supervisor." See Appendix E for the full scale.

Job satisfaction. Job satisfaction was measured with a three item scale from Cammann, Fichman, Jenkins, and Klesh (1983; $\alpha = .91$). The measure contains a Likert response scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). A sample item is “All in all, I am satisfied with my job.” See Appendix F for the full scale.

Overall job stress. The Stress in General (SIG) scale was used to measure overall job stress (Stanton, Balzer, Smith, Parra, & Ironson, 2001). This measure assesses perceptions of stress broadly rather than tied to specific stressors or strains. The SIG consists of two factors: “pressure” ($\alpha = .88$) and “threat” ($\alpha = .91$). The scale asks participants “What is your job like most of the time?” and includes 15 items for which participants may indicate “No,” “?”, or “Yes.” A sample item from the “pressure” factor is “Demanding,” and a sample item from the “threaten” factor is “Overwhelming.” See Appendix G for the full scale.

Meaningfulness of work. Meaningfulness of work was measured using the Work and Meaning Inventory (WAMI; Steger, Dik, & Duffy, 2012). The WAMI consists of 10 items within three subscales: positive meaning, meaning-making through work, and greater good motivations. The response scale for the measure is a Likert scale ranging from 1 (*absolutely untrue*) to 5 (*absolutely true*). A sample item from the scale is “The work I do serves a greater purpose.” There is evidence of high internal consistency reliability for the WAMI in this study ($\alpha = .94$). See Appendix H for the full scale.

Role overload. Role overload was measured using 3 items ($\alpha = .77$) (Cammann et al., 1983). The scale consists of three items with a response scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). A sample item is “I never seem to have enough time to get everything done.” See Appendix I for the full scale.

Social desirability. Social desirability was measured using the Crowne-Marlowe Social Desirability Scale ($\alpha = .85$) (Crowne & Marlowe, 1960). The scale consists of 33 items with a True or False response scale. A sample item is “Before voting I thoroughly investigate the qualifications of all the candidates.” See Appendix J for the full scale.

Underload (previous measure). In order to collect evidence of the convergent validity of the newly developed scale, underload was also assessed using Caplan et al.’s (1975) scale ($\alpha = .76$). The scale consists of 11 items; for four of the items the response scale ranges from 1 (*rarely*) to 5 (*very often*), and for the other seven items the response scale ranges from 1 (*hardly any*) to 5 (*a great deal*). A sample item from the first part of the scale is “How often does your job require you to work very fast?” and a sample item from the second part of the scale is “How much slowdown in the workload do you experience?” See Appendix K for the full scale.

Study 2- Results

Statistical Analyses for Scale Evaluation. I completed the following steps in order to fully evaluate the items in the scale and test Hypothesis 1 (i.e., *Underload is comprised of three distinct dimensions: perceptions, desires, and expectations*). Using random number generation in Microsoft Excel, I split the total sample ($N = 387$) into two separate samples for the analyses ($N = 193$ and $N = 194$, respectively). I took this step in order to use the data from one sample to conduct an exploratory factor analysis and the data from the other sample to conduct a confirmatory factor analysis, as these analyses should not be conducted on the same sample (Brown, 2006).

Sample 1 Results. Sample 1 ($N = 193$) was used to conduct basic item analyses as well as an exploratory factor analysis.

Basic item analysis. The first step I conducted was a basic item analysis. I analyzed the 25 items in the underload scale using principles from Novick's (1966) classical test theory. The results of the basic item analysis for Sample 1 are displayed in Tables 4-6. First, I calculated the item difficulties by determining the mean and variances of the item responses. If the mean difficulty of the items falls toward the middle of the scale (i.e., 3 on the 5-point Likert scale) and the items have large variances, this indicates that the scale is measuring the full range of the underload construct. A low or high mean difficulty relative to the 5-point scale indicate that the scale is only measuring one extreme of the construct of underload. Additionally, low item variances indicate that the scale is not effectively distinguishing between individuals experiencing different levels of underload (DeVellis, 2012). Item difficulties for Sample 1 ranged from 1.98 to 2.66 with an average value of 2.20. This value was lower than the ideal midpoint of 3 which indicates that it was difficult for participants to endorse the underload items or to agree with the statements regarding the experience of underload. Thus, the items are better for measuring the lower end of the scale rather than assessing underload at all levels.

Next, I assessed the item discriminations by calculating the Pearson correlations for each item with the total score. I calculated the total score by adding the participants' responses across the full measure. Larger item-total correlations indicate that the specific item is particularly useful for distinguishing between individuals experiencing different levels of underload. In other words, highly discriminating items give us more information about a person's standing on the construct. Items with small correlations with the total score are more ambiguous as far as determining an individual's specific level of underload (DeVellis, 2012). Items that correlate less than .10 with the total score may not be effective and could eventually be removed from the scale (Varma, 2006). Item discriminations ranged from .69 to .86 with an average of .79. These higher

item-total correlations indicate that the items are discriminating between individuals who are experiencing different levels of underload. The correlation matrix containing all items revealed that there were not any negative inter-item correlations (see Table 7). This indicates that there are not any serious concerns regarding the internal consistency of the scale that need to be examined before continuing with the analyses.

Exploratory factor analysis. Exploratory factor analysis (EFA) is useful when there is not extensive literature on the theoretical structure of the construct (Floyd & Widaman, 1995; Hurley, Scandura, Brannick, & Vandenberg, 1997). Conducting an EFA allowed me to explore alternatives to my hypothesized factor structure of the underload scale. I conducted an EFA and used a Geomin oblique rotation in order to allow the factors to be correlated. This resembles the reality of most datasets rather than orthogonal rotations which do not allow the factors to be correlated. It is generally a better approach to begin with an oblique rotation because if the factors are truly uncorrelated, the analysis will default to an orthogonal rotation (Bandalos & Boehm-Kaufman, 2008). The results provide initial support for Hypothesis 1 by indicating a three-factor structure. All of the items had the highest factor loading on the intended dimension (Table 8).

The “eigenvalue greater than one” rule (Kaiser method) is a commonly used practice to extract factors. However, several researchers have cautioned against this method, as it can underestimate or overestimate the number of factors in a measure (Zwick & Velicer, 1986). A more reliable method of extracting factors involves a scree plot and parallel analysis (Cattell, 1966). A scree plot is a graph that displays the eigenvalues for each factor in order from greatest to least. The shape of the plot is vertical on the left (represented by the highest eigenvalues), and changes to horizontal on the right (represented by the lowest eigenvalues). By examining the

scree plot, one can determine the number of factors that should be extracted by determining the point at which the plot switches from vertical to horizontal. The factors that lie above this point of change should be retained (Cattell, 1966).

The point at which the shape of the scree plot changes is not always distinct; the change can rather be gradual. Thus, a less ambiguous method of factor extraction is a parallel analysis. In order to conduct a parallel analysis, I created a random matrix of data that was the same size as the actual obtained data. I performed the EFA on each dataset and compared the eigenvalues for the factors for each dataset. Based on this comparison, I retained the factors with eigenvalues that were larger in the actual data than in those in the artificial data (DeVellis, 2012). The scree plots for the data from the study as well as for a randomly generated dataset of the same size are displayed in Figure 2. The highest eigenvalue for the random data is 1.68, suggesting that the factors with eigenvalues higher than this value should be extracted for the actual data. Three factors in the sample 1 dataset had eigenvalues higher than 1.68 providing further evidence of a three-factor structure. The eigenvalues for both sets of data are displayed in Table 9.

Reliability. I took several steps in order to assess the reliability of the scale. Before calculating the reliability, I performed the preliminary step of ensuring that all of the items in the scale were positively correlated. Negative inter-item correlations can significantly reduce the reliability of the scale (Raykov & Marcoulides, 2011). In order to avoid this threat to reliability, the correlation matrix was examined, and none of the items caused negative correlations (see Table 7).

I assessed the internal consistency of the scale in order to evaluate the correlations between items and ensure that all items are consistently producing similar scores within the measure of underload. I calculated Cronbach's alpha using SPSS. The scale showed good

internal consistency as a whole ($\alpha = .97$) as well as for each dimension (expectations: $\alpha = .96$; desires: $\alpha = .97$; perceptions: $\alpha = .94$). The high alpha values for the expectations and desires dimensions suggest that there may be some redundancy among the items within each dimension. However, while the alpha coefficient assesses the relationships of the items with each other, it does not directly evaluate whether a single construct is being measured by the scale (Raykov & Marcoulides, 2011). The omega coefficient is another way to evaluate reliability because it indicates the homogeneity of the scale (McDonald, 1999). Thus, I also calculated the omega coefficient using *R* version 3.0.2. The results showed high reliability for the scale as a whole ($\omega = .99$) as well as for each dimension (expectation: $\omega = .99$; desires: $\omega = .99$; perceptions: $\omega = .99$). Correlations and reliabilities for all dimensions are displayed in Table 10.

Sample 2 Results. Sample 2 ($n=194$) was used to conduct additional item analyses as well as a confirmatory factor analysis.

Basic item analysis. I followed the same steps to conduct a classical item analysis for the 25 items using data from sample 2. The results are displayed in Tables 11-13. The item difficulties ranged from 1.95 to 2.60 with an average value of 2.25. This average was lower than the ideal mid-point of the scale indicating that the items measured the lower end of the underload spectrum. Discriminations ranged from .63 to .86 with an average of .77. All item-total correlations were at least moderate (i.e., greater than .30) which indicates that the items are discriminating between individuals experiencing different levels of underload. I evaluated a correlation matrix including all items, and there were no negative inter-item correlations, indicating that there are no threats to internal consistency reliability in this stage (Table 14).

Confirmatory factor analysis. Next, I performed a confirmatory factor analysis (CFA) to test my hypothesized three-factor structure based on the theoretical background related to

underload. I examined several fit statistics to determine how well the proposed model from the CFA fit the data. I also performed a chi-square test of model fit to determine whether the predicted data based on the specified model was significantly different than the actual data. Ideally, a non-significant chi-square test would be a good indication of fit, but the chi-square test is especially sensitive to sample size (Raykov & Marcoulides, 2011). The root mean square error of approximation (RMSEA) assesses the misfit per degree of freedom and is related to the chi-square value without being as sensitive to sample size. A model is generally considered to have good fit with an RMSEA value of less than .05 (Raykov & Maroculides, 2011). Finally, I examined the comparative fit index (CFI) and Tucker-Lewis index (TLI) as an indication of model fit. Values closer to 1 indicate that the model fits the data better than a baseline model (Raykov & Marcoulides, 2011). The results showed at least moderate fit based on several model fit statistics including the chi-square test ($\chi^2 = 599.26$, $df = 272$, $p < .001$), the root mean square error of approximation (RMSEA = .08, $p < .001$), the comparative fit index (CFI = .91), the Tucker-Lewis index (TLI = .90), and the standardized root mean square residual (SRMR = .05). I also compared the three-factor model to a one-factor model and a two-factor model using a chi-square difference test in order to test the hypothesized structure of underload. The data fit a three-factor model significantly better than a single-factor model based on a chi-square difference test ($\Delta \chi^2 = 836.50$, $\Delta df = 3$, $p < .001$) and significantly better than a two-factor model combining the expectations and desires dimensions into one factor ($\Delta \chi^2 = 560.28$, $\Delta df = 2$, $p < .001$), providing further evidence to support Hypothesis 1.

Next, I examined the factor loadings to determine whether the items loaded onto the hypothesized factors. Items should have loadings of at least .40 on to the proposed factor (Ford,

MacCallum, & Tait, 1986). All item loadings on their respective factors were higher than .40 with a range from .67 to .94 (see Table 15).

I calculated the discrepancy matrix for all items in the CFA by subtracting the observed item correlation matrix from the predicted item correlation matrix (see Table 16). Discrepancies above .10 may be problematic because this may indicate that certain items are causing a misfit in the scale (McDonald, 1999). Expectations items 3 and 7, desires items 6 and 8, and perceptions item 3 all had high discrepancy values, suggesting that they may not fit as well within the scale. I examined these items further and considered removing them from the final scale.

Reliability analysis. I calculated Cronbach's alpha and omega values to evaluate the reliability of the scale for the sample 2 data. The alpha and omega values provided evidence of high reliability for the scale as a whole ($\alpha = .97$; $\omega = .99$) as well as for each dimension (expectations: $\alpha = .96$, $\omega = .99$; desires: $\alpha = .97$, $\omega = .99$; perceptions: $\alpha = .93$, $\omega = .99$). Correlations and reliability coefficients for all dimensions are displayed in Table 17.

Scale Reduction. I reviewed the item characteristics for the total sample to identify potential items to remove from the scale. Following the guidelines for scale reduction provided by Stanton, Sinar, Balzer, and Smith (2002), several sources of evidence should be used to determine which items are most appropriate to remove from a scale. When reviewing the scale with the purpose of reducing the number of items, I considered criteria including factor loadings, item discriminations and difficulties, and the clarity and relevance of the items in relation to the guiding theoretical background. I did not have predetermined cutoff values for the factor loadings and item statistics; rather, I considered these components for each item in relation to the rest of the items in the scale and made decisions regarding eliminating items based on which ones had the best empirical and conceptual fit for the scale as a whole.

Item 1 in the perceptions dimension had the lowest factor loading, discrimination, and difficulty for the factor. I examined this item more closely and concluded that it did not fit the definition as well as I had initially thought. Although the item (“I do not have enough work to do”) seems to capture the idea of underload, it involves more than a subjective assessment of workload because individuals may take other factors into account when deciding how much work is “enough” work. For example, an individual might consider their expectations and desires as compared to their perceptions of their actual workload. Thus, I removed this item from the scale.

Next, I reexamined item 2 (“*Someone in my job position should have a higher workload*”) in the expectations dimension. This item had a relatively low factor loading within the dimension as well as a low item difficulty. Additionally, I concluded that the item may not capture the definition of expectations related to workload. Unlike the rest of the items in this dimension, this item does not focus on the individual but rather involves a reflection of expectations of other employees in similar job positions. I therefore concluded that it was not the best item to include in the scale, and I removed it.

Finally, I reviewed the discrepancy matrix for the initial items (Table 16). Large discrepancies with an absolute value greater than .10 can provide evidence that a particular item may not fit well within the scale because they indicate that items are correlating more or less than they should with other items (McDonald, 1999). Expectations items 3 and 7, desires items 6 and 8, and perceptions item 3 all had high discrepancy values, suggesting that they may not fit as well within the scale. In order to eliminate possible redundancy indicated by large discrepancies and misfit indicated by smaller discrepancies, I removed these 5 items from the scale. The final 18-item underload scale is shown in Appendix L.

Sample 2 re-analysis. I conducted a confirmatory factor analysis which confirmed that the remaining 18 items were a better fit to the data given a three-factor model based on improved fit statistics ($\chi^2 = 202.13$, $df = 132$, $p < .001$; RMSEA = .05, $p < .001$; CFI = .97; TLI = .97; SRMR = .03) and high factor loadings (see Table 18). A chi-square difference test showed that the 18-item scale had significantly better fit than the 25-item scale ($\Delta \chi^2 = 397.13$, $\Delta df = 140$, $p < .001$). The comparisons are shown in Table 19. Internal consistency reliability estimates remained high for the 18-item scale as a whole ($\alpha = .96$) as well as for each dimension (expectations: $\alpha = .95$; desires: $\alpha = .97$; perceptions: $\alpha = .92$). Correlations and reliability values for all dimensions are shown in Table 20.

Relations with Other Variables. Because the item characteristics for sample 1 and sample 2 were highly similar, I combined both samples when evaluating relations with other variables in order to increase statistical power. The descriptive statistics for all variables are reported in Table 21, and the correlation matrix is presented in Table 22. There were small, positive relationships between all dimensions of underload and underemployment (perceptions: $r = .24$, $p < .001$; desires: $r = .17$, $p < .001$; expectations: $r = .18$, $p < .001$) and small to moderate, positive relationships between all dimensions of underload and job boredom (perceptions: $r = .36$, $p < .001$; desires: $r = .23$, $p < .001$; expectations: $r = .20$, $p < .001$). Additionally, there were large, positive relationships between all dimensions of my new measure of underload and Caplan et al.'s (1975) measure of underload (perceptions: $r = .67$, $p < .001$; desires: $r = .56$, $p < .001$; expectations: $r = .61$, $p < .001$). This provided evidence of convergent validity as well as partial support for Hypothesis 2a (i.e., *All three dimensions of are moderately but not strongly positively related to boredom*) and full support for Hypothesis 2c (i.e., *Previous measures of underload are strongly positively related to all three dimensions of the new measure of underload*), but

Hypothesis 2b (i.e., *All three dimensions of underload are moderately but not strongly positively related to underemployment*) was not supported.

There were moderate, negative relationships between all dimensions of underload and role overload (perceptions: $r = -.38, p < .001$; desires: $r = -.36, p < .001$; expectations: $r = -.36, p < .001$). Thus, Hypothesis 3a (i.e., *All three dimensions of underload are strongly negatively related to overload*) was not supported. Additionally, there were nonsignificant relationships between the perceptions and expectations dimensions of underload and social desirability (perceptions: $r = -.03, p = .51$; expectations: $r = .03, p = .52$), providing evidence of discriminant validity. However, there was a small, positive correlation between the desires dimension of underload and social desirability ($r = .13, p = .01$) providing only partial support for Hypothesis 3b (i.e., *All three dimensions of underload are not related to social desirability*).

The perceptions dimension of underload was slightly negatively correlated with the positive meaning subscale of meaningfulness of work ($r = -.17, p = .001$) and with job satisfaction ($r = -.12, p = .02$). However, there were no significant relationships between the desires or expectations dimensions of underload and meaningfulness of work or job satisfaction. This provided partial support for Hypotheses 4a (i.e., *All three dimensions of underload are negatively related to job satisfaction*) and 4b (i.e., *All three dimensions of underload are negatively related to meaningfulness of work*). Finally, there were small, positive relationships between role ambiguity and the desires dimension of underload ($r = .11, p = .03$) as well as the expectations dimension of underload ($r = .11, p = .03$) providing partial support for Hypothesis 4c (i.e., *All three dimensions of underload are positively related to work role ambiguity*). I also found small, negative relationships between all three dimensions of underload and overall job stress (perceptions: $r = -.22, p < .001$; desires: $r = -.25, p < .001$; expectations: $r = -.27, p < .001$).

This finding contradicted Hypothesis 4d (i.e., *All three dimensions of underload are positively related to perceived work stress*). I will elaborate on the possible reasons for this in the discussion section below.

Discussion

The purpose of this study was to provide a clear definition of the construct of work-related underload based on previous empirical and theoretical research and to create and evaluate a scale to measure this construct. I defined underload as a condition in which an individual perceives that he or she does not have enough to do at work. I developed items for a three-dimensional scale to assess subjective perceptions of workload, expectations regarding workload, and desires regarding workload. Finally, I collected validity and reliability evidence to support the psychometric properties of my measure. Taken together, the results provide evidence of the reliability and validity of a three-dimensional scale to assess work-related underload.

The results of my analyses showed that the data fit a three-factor model with distinct dimensions for perceptions, expectations, and desires, providing support for Hypothesis 1. High Cronbach's alpha values and omega values for the scale as a whole as well as for each dimension provided evidence of the internal consistency reliability of the measure. Additionally, all three dimensions of underload were positively correlated with underemployment, job boredom, and a previous measure of underload, providing partial support for Hypothesis 2a, and full support for Hypothesis 2c. Convergent validity evidence was supported, as all of the dimensions of the scale were consistently positively related to theoretically similar constructs (i.e., previous measure of underload, underemployment, job boredom).

Some of the findings provided evidence of the discriminant validity of the scale, although the evidence was not as strong as anticipated. All dimensions of the underload scale were negatively related to overload as hypothesized, but the relationships were only moderate in magnitude. Additionally, the perceptions and expectations dimensions of the scale were not

significantly related to social desirability, adding to the evidence of discriminant validity.

However, the desires dimension was slightly positively correlated with social desirability. Social desirability reflects the tendency for individuals to respond to test items in a way that presents themselves favorably (Edwards, 1953). This suggests that participants may have perceived that it was socially desirable to indicate that they desired more work rather than indicating that they did not prefer a high workload. Participants may have been hesitant to report that they did not want to perform additional work, as this may be associated with a connotation of being lazy or unmotivated as an employee.

The concurrent criterion-related validity of the scale was partially supported, as only the perceptions dimension of underload was negatively related to the positive meaning subscale of meaningfulness of work and job satisfaction. The desires and expectations dimensions were not significantly related to meaningfulness of work or job satisfaction. Further, the desires and expectations dimensions of underload were positively related to role ambiguity. These findings suggest that individuals who want or expected more work may perceive a lack of clarity regarding their responsibilities at work.

Finally, I hypothesized that underload would be positively related to subjective perceptions of stress. However, I found that the two constructs were negatively related across all dimensions of underload. This was surprising to me at first, as the theoretical background and empirical evidence discussed in the introduction suggested that underloaded employees should experience more stress at work. However, a closer look at the way in which stress was measured in the study made it clear as to why a negative relationship exists. The Stress in General scale asks respondents to indicate whether or not their job can be described by a list of adjectives. The adjectives in the scale (e.g., demanding, pushed, hectic) reflect characteristics of the job that

might lead to *distress* in the workplace. An employee who is underloaded is not likely to perceive that his or her job is demanding, pushed, or hectic because it would be difficult to experience those characteristics when there is not enough work to do. A scale that included items to assess eustress, such as the Valencia Eustress-Distress Appraisal Scale (VEDAS; Rodriguez, Kozusznik, & Pero, 2013) may have been a better choice because it would have captured a lack of challenge stressors and positive stress that is associated with underload.

Collectively, the results point to the conclusion that my new underload scale provides valuable information above and beyond that provided by previously developed scales. The inclusion of more items in my scale allows one to more fully assess the extent to which an employee perceives underload. This is an improvement from single-item measures developed for previous studies. Additionally, the content of the items in my scale were designed to effectively distinguish between individuals experiencing different levels of underload, whereas Caplan et al.'s (1975) scale included several items that would not apply to those with very low workloads (e.g., "How many lulls between heavy workload periods do you have?"). My scale also had higher estimates of internal consistency as compared to Caplan et al.'s (1975) previously developed scale (new scale: $\alpha = .96$; Caplan et al. (1975) scale: $\alpha = .76$). Finally, the most important contribution of my scale is the consideration of expectations and desires related to workload. This adds a component to the construct of underload that reflects theoretical perspectives including the cognitive appraisal process.

The results and theoretical background provide evidence to support the significance of three distinct dimensions within the context of quantitative underload. Perceptions, desires, and expectations are all essential components of the experience of work-related underload. However, conceptually it is not logical to combine scores on each of these dimensions into one aggregate

score of underload. When assessing underload, all three dimensions should be considered in relation to each other, but simply combining them into one score does not allow us to connect underload to meaningful outcomes. For example, an employee who scores high on the perceptions dimension and low on the expectations and desires dimensions perceive that they do not have enough work to fill their time but do not want or expect to do more work. On the other hand, employees who score low on the perceptions and expectations dimensions and high on the desires dimension perceive that they have enough work to fill their time and did not expect more work but want to do more work. These two groups of employees would receive identical aggregate scores because of their dimension scores (i.e., high on one dimension, low on the other two dimensions) but they are experiencing very different situations. The first group of employees have workloads that match their desires and expectations whereas the second group of employees would prefer to do more work than they perceive they have. Thus, there would not be significant relations between the aggregate underload scale score and hypothesized outcome variables.

Instead of combining the dimensions into a single composite score, differences between dimension scores should be assessed and considered in relation to outcomes. When perceptions are in agreement with desires and expectations, person-environment fit theory posits that there should be good fit between the employee and the job and that job satisfaction should be higher (Kristof-Brown et al., 2005). In contrast, larger discrepancies between perceptions and desires and expectations should be related to more negative outcomes for the employee and for the organization. Future research can explore the extent to which agreement or discrepancy between dimensions is associated with individual and organizational outcomes using methods such as polynomial regression and response surface methodology (Edwards, 2001).

Theoretical Contribution

The findings of my study make a significant theoretical and empirical contribution to the work-related stress literature. First, there is not one definition of underload that has been used consistently within the literature or to guide measurement of the construct. I developed a definition that was based on theories of cognitive appraisal, person-environment fit, and psychological contracts. Additionally, my study furthers the underload literature by taking into consideration the role of employee perceptions, expectations, and desires in the experience of work-related underload. The lack of clarity regarding a definition of underload may have contributed to the deficiency of empirical studies that examine underload at work. This comprehensive definition of the construct may be the first step in continuing research on the topic.

The limited body of existing literature also does not include a validated measure of underload. The results of this study provide evidence for the reliability and validity of the three-dimensional underload scale. Further, this was the first study to conceptualize and systematically examine expectations and desires regarding workload in relation to underload. The results provide evidence that these dimensions are important to consider when assessing related constructs. Thus, this measure can be used in future studies to further explore the outcomes and antecedents associated with underload. Given the growth of automation in the workplace and the increased potential for underload at work, additional research can use the measure to contribute to the literature within the context of work-related stress and eustress.

Practical Implications

The findings of my study also have important practical implications for organizations especially when it comes to job redesign, recruitment, and selection. Ideally, organizations can

address employees' perceptions of underload by reorganizing job roles and responsibilities so that all employees have enough work to keep them busy during the work day. Leaders who have an increased awareness of the potential negative consequences of underload may be more perceptive regarding situations in which employees are underloaded and may actively change the circumstances to avoid perceptions of underload. Alternatively, leaders might encourage employees to engage in job crafting which involves enhancing one's own meaningfulness and personal identity in work (Wrzesniewski & Dutton, 2001). Specifically, task crafting involves changing the boundaries associated with specific tasks and can include adding tasks that make work more meaningful for the employee (Berg, Grant, & Johnson, 2010).

In many situations, jobs that involve low workloads or fluctuating workloads such as various times with low workload may be unavoidable. In this case, it is important for organizations to manage workers' expectations regarding workload during the recruitment process. This can be achieved through the use of realistic job previews as part of the recruitment and selection process. Realistic job previews are designed to give potential employees accurate information about the job position that they are applying for in order to allow them to make informed decisions about whether or not to accept the job (Wanous, 1973). Expectations are an important dimension in the experience of work-related underload; thus, a realistic job preview for a job involving a low workload should clearly communicate this aspect of the job to the candidate. Realistic expectations regarding workload can reduce the chances that an employee will perceive a violation of the psychological contract and that he or she will experience underload.

Finally, for jobs that involve low workload, organizations should select employees who prefer to do less work and do not desire a fast-paced job. Theories of person-environment fit

posit that the characteristics of an individual should match the characteristics of the work environment (Kristof-Brown, Zimmerman, & Johnson, 2005). Thus, a person who desires less work will be a better fit for a job with low workload and will be less likely to perceive underload. Future research should be conducted to explore possible personality traits (e.g., conscientiousness) that may be associated with an increased desire for more work. Selection tests can then be used to identify candidates who would be the best fit for the job.

Limitations

There are several limitations to this study. First, samples from a single source (MTurk) were used in this study. Ideally, data should be collected from several independent samples so that the results can be cross-validated (Floyd & Widaman, 1995). A strength of my study, however, is that I collected several samples at different points in time from MTurk rather than collecting all of the data at one point in time. I randomly split the data in half and treated the data as two separate samples in order to conduct an EFA on one sample and a CFA on the other sample. This is an acceptable practice for cross-validation, but additional samples from different sources would provide more confirmation of the proposed factor structure (Hurley et al., 1997). Future research should target different worker populations to gain more validity evidence and increase the generalizability of the findings.

Another limitation stems from the fact that common method bias may occur as a result of the self-reported data collected at a single point in time (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). This could ultimately be a source of measurement error and may threaten the validity of the proposed relationships between measures. Additionally, collecting the data at a single point in time does not provide any information about the experience of underload over time. For example, employees who experience underload during a brief period of time may not

face the same negative outcomes as employees who consistently experience underload in their jobs. Future studies should aim to collect longitudinal data which may be better for reflecting the negative effects associated with the experience of consistent underload rather than short periods of underload. Social desirability bias is another potential limitation given the positive correlation between measures of social desirability and the desires dimension. This is a common limitation of self-report measures, and a possible solution is to control for social desirability when conducting further analyses using the desires dimension of underload (Podsakoff et al., 2003)

Finally, all of the subject matter experts in Study 1 were Industrial/Organizational Psychology doctoral students. Although all SMEs had general knowledge of Industrial/Organizational Psychology and psychological measurement, not all of them were familiar with the concepts of underload or related constructs. Researchers in the field who have studied topics related to underload may provide more valuable feedback regarding the scale items in the future.

Future Research

Now that I have created this measure of underload and obtained initial validity evidence, there are many additional research questions that warrant explanation related to the experience of underload at work. Future research should explore the antecedents and consequences of underload. Who is more susceptible to experiencing underload? One possibility is that certain leadership styles make it more likely that subordinates will experience underload. Who is more susceptible to experiencing negative outcomes as a consequence of underload? Perhaps individuals who are high in conscientiousness are more likely to experience strain and job dissatisfaction when they do not have enough to do at work. What other negative outcomes are associated with underload? Research on boredom and counterproductive work behaviors has suggest that underloaded employees may engage in more nonwork activities during working

hours (Fisher, 1987). Future research can begin to examine these relationships and build a model to represent the process of underload. This research can then be used to inform organizations about the potential consequences of underload and help leaders identify and prevent this from occurring.

I also plan to further investigate the relationships between the three dimensions of the underload scale (i.e., perceptions of workload, expectations related to workload, and desires for workload) in relation to outcomes at work. The agreement or discrepancy between these dimensions may be important when it comes to certain outcomes. For example, an employee who scores high on the perceptions factor and low on the desires factor perceives that they do not have enough work to do but does not express any desire to do more work. Alternatively, an employee who scores high on both the perceptions and desires factors does not have enough work to do and wants to do additional work. These employees may vary when it comes to outcomes such as job satisfaction and stress. This research question can be addressed using polynomial regression and response surface methodology.

As indicated earlier, future research should also examine the experience of underload over time. Data collected at a single point in time may introduce common method bias that can influence the results and conclusions. The implications associated with the results could be strengthened if a measure of underload was administered at several points in time. Additionally, this would allow me to evaluate the test-retest reliability of the measure, adding to the reliability evidence. Finally, assessing the outcomes associated with underload over time could lead to additional implications. It is likely that prolonged periods of perceived work-related underload would lead to more negative outcomes than brief periods of underload.

Finally, I plan to take steps to explore the potential of reducing the length of the scale in order to make it more practical to use within surveys. I will follow the steps outlined in Stanton et al. (2002) to determine the best set of items for a shorter version of the scale.

Conclusion

Although underload has received limited attention in the organizational and occupational health psychology literature, the existing research and theoretical background suggest that the experience of underload can significantly impact employees and organizations. Confusion about the definition and measurement of the construct in the past has prevented researchers from moving forward with studies that explore the antecedents and consequences related to underload. In this study, I provided a methodologically and theoretically sound definition of underload and created a scale with evidence of validity and reliability in order to effectively measure the construct. This step will set the stage for future research to continue to examine the nature of underload as an important organizational issue.

Table 1.*Initial item pool and SME ratings for perceptions dimension from Study 1.*

Perceptions items:	<i>M</i>	<i>SD</i>	<i>r_{wg}</i>
1. I do not have enough work to do.	4.91	0.30	0.98
2. I find myself with nothing to do.	4.09	1.04	0.75
3. I have more than enough time to complete my job duties.	3.45	0.93	0.80
4. After I complete all of my work, there is still time left in my work day.	3.91	1.22	0.66
5. I am capable of completing more work in the time that I am required to be at work.	3.18	0.87	0.83
6. The work that I do does not fill my entire work day.	4.09	0.94	0.80
7. In order to work at my full capacity, I would need more work to do	4.45	0.93	0.80
8. I have too much time on my hands.	3.82	0.87	0.83
9. I do not have enough work to do to fill my entire work day.	4.36	0.67	0.90
10. I do not have enough to focus on.	3.09	1.14	0.71

Note: $N=11$. The reported mean is the mean rated representativeness of the item. r_{wg} = index of rater agreement. Relevance was rated on a 5-point scale where 1 = Not at All Relevant, 2 = Minimally Relevant, 3 = Moderately Relevant, 4 = Substantially Relevant, 5 = Extremely Relevant.

Table 2.*Initial item pool and SME ratings for desires dimension from Study 1.*

Desires items:	<i>M</i>	<i>SD</i>	<i>r_{wg}</i>
1. I wish that I had more to do.	4.91	0.30	0.98
2. Having more work to do would help the time pass by more quickly.	3.54	1.13	0.71
3. I wish my manager/supervisor would give me more work to / complete.	4.27	0.90	0.81
4. I would prefer to be busier.	4.40	0.84	0.84
5. I want more work to do.	4.82	0.40	0.96
6. The amount of work that I have is less than what I would like.	4.45	0.52	0.94
7. Having more work to do would better satisfy my preferences.	4.09	0.70	0.89
8. Having more work to do would be ideal.	4.45	0.52	0.94
9. I would like to do additional work.	4.09	0.83	0.84
10. I would be more satisfied with a higher workload.	4.64	0.67	0.90

Note: $N=11$. The reported mean is the mean rated representativeness of the item. r_{wg} = index of rater agreement. Relevance was rated on a 5-point scale where 1 = Not at All Relevant, 2 = Minimally Relevant, 3 = Moderately Relevant, 4 = Substantially Relevant, 5 = Extremely Relevant.

Table 3.*Initial item pool and SME ratings for expectations dimension from Study 1.*

Expectations items:	<i>M</i>	<i>SD</i>	<i>r_{wg}</i>
1. When I accepted this job, I thought that it would involve more work.	4.91	0.30	0.98
2. Someone in my job should have a higher workload.	4.27	0.79	0.86
3. There should be more for me to do when I am at work.	3.82	0.87	0.83
4. The amount of work that I have to do is less than what I expected when I was hired.	4.73	0.47	0.95
5. I thought that I would have more to do in this job.	4.45	0.52	0.94
6. I expected to be busier in this job.	4.55	0.52	0.94
7. When I started my job, I had the impression that I would have more work to do.	4.55	0.69	0.89
8. The information about my job that I had before I was hired led me to believe that I would have more work to do.	3.91	0.94	0.80
9. I expected to have less free time during my work day.	4.18	0.87	0.83
10. When I accepted this job, I thought that my work would take up more time in my work day.	4.27	0.65	0.90

Note: $N = 11$. The reported mean is the mean rated representativeness of the item. r_{wg} = index of rater agreement. Relevance was rated on a 5-point scale where 1 = Not at All Relevant, 2 = Minimally Relevant, 3 = Moderately Relevant, 4 = Substantially Relevant, 5 = Extremely Relevant.

Table 4.*Perceptions items included in MTurk survey—difficulties and discriminations for Sample 1.*

Perceptions items:	Item difficulty	Item discrimination
1. I do not have enough work to do.	2.00	0.73
2. I find myself with nothing to do.	1.98	0.74
3. I have more time than I need to do my work.	2.66	0.69
4. After I complete all of my work, there is still time left in my work day.	2.45	0.75
5. In order to work at my full capacity, I would need more work to do.	2.18	0.79
6. I have too much time to complete my work.	2.11	0.75
7. I do not have enough work to do to fill my entire work day.	2.12	0.74
8. I could be more productive if I had more work to do.	2.14	0.78

Note: $N = 193$. All items begin with the stem “When I am at work...”

Table 5.*Desires items included in MTurk survey—difficulties and discriminations for Sample 1.*

Desires items:	Item difficulty	Item discrimination
1. I wish that I had more to do.	2.00	0.82
2. I wish that more of my time was filled up.	2.13	0.86
3. I wish that there were not as many lulls in my work day.	2.05	0.81
4. I would prefer to be busier.	2.17	0.83
5. I want more work to do.	2.16	0.85
6. The amount of work that I have is less than what I would like.	2.08	0.84
7. I would be more satisfied if I had more work to do.	2.07	0.86
8. I would like to do additional work.	2.15	0.80
9. I would enjoy having a higher workload	2.11	0.81

Note: $N = 193$. All items begin with the stem “When I am at work...”

Table 6.*Expectations items included in MTurk survey—difficulties and discriminations for Sample 1.*

Item:	Item difficulty	Item discrimination
1. When I accepted this job, I thought that it would involve more work.	2.25	0.70
2. Someone in my job position should have a higher workload.	2.25	0.77
3. There should be more for me to do when I am at work.	2.22	0.85
4. When I started my job, I had the impression that I would have more work to do.	2.30	0.82
5. I thought that I would have more work to do in this job.	2.18	0.85
6. I expected to be busier in this job.	2.27	0.85
7. I expected to have less free time during my work day.	2.43	0.70
8. When I accepted this job, I thought that my work would take up more time in my work day.	2.41	0.77

Note: N = 193.

Table 7.*Sample 1 correlation table for all 25 items included in MTurk survey.*

	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12	13
SWP1	2.01	1.26	1.00												
SWP2	1.98	1.31	0.75	1.00											
SWP3	2.66	1.41	0.55	0.55	1.00										
SWP4	2.45	1.37	0.66	0.68	0.63	1.00									
SWP5	2.18	1.25	0.67	0.68	0.58	0.66	1.00								
SWP6	2.11	1.27	0.67	0.73	0.67	0.74	0.73	1.00							
SWP7	2.12	1.26	0.60	0.65	0.59	0.62	0.72	0.65	1.00						
SWP8	2.14	1.29	0.63	0.71	0.60	0.63	0.77	0.70	0.75	1.00					
DES1	2.00	1.10	0.51	0.49	0.48	0.51	0.55	0.51	0.47	0.53	1.00				
DES2	2.13	1.18	0.55	0.56	0.53	0.50	0.63	0.51	0.57	0.61	0.83	1.00			
DES3	2.05	1.09	0.51	0.57	0.51	0.55	0.57	0.55	0.49	0.53	0.72	0.77	1.00		
DES4	2.17	1.18	0.53	0.53	0.49	0.60	0.58	0.50	0.52	0.55	0.77	0.81	0.77	1.00	
DES5	2.16	1.16	0.51	0.56	0.52	0.58	0.60	0.53	0.52	0.62	0.79	0.83	0.76	0.86	1.00
DES6	2.09	1.11	0.56	0.59	0.56	0.56	0.57	0.55	0.55	0.60	0.78	0.82	0.78	0.78	0.86
DES7	2.07	1.15	0.53	0.55	0.55	0.57	0.63	0.58	0.53	0.61	0.82	0.84	0.79	0.82	0.86
DES8	2.15	1.14	0.46	0.50	0.48	0.48	0.52	0.49	0.50	0.56	0.78	0.77	0.73	0.77	0.83
DES9	2.11	1.13	0.46	0.52	0.47	0.53	0.56	0.53	0.49	0.56	0.77	0.76	0.76	0.81	0.88
EXP1	2.25	1.09	0.49	0.45	0.41	0.49	0.51	0.51	0.42	0.44	0.52	0.54	0.51	0.51	0.45
EXP2	2.25	1.17	0.56	0.49	0.48	0.48	0.58	0.46	0.56	0.56	0.61	0.64	0.57	0.51	0.58
EXP3	2.22	1.20	0.59	0.55	0.53	0.52	0.62	0.52	0.61	0.62	0.70	0.75	0.67	0.66	0.69
EXP4	2.30	1.25	0.54	0.47	0.49	0.57	0.59	0.51	0.53	0.52	0.62	0.65	0.58	0.62	0.62
EXP5	2.18	1.19	0.52	0.51	0.50	0.56	0.61	0.55	0.56	0.59	0.67	0.70	0.63	0.64	0.65
EXP6	2.27	1.24	0.57	0.52	0.48	0.59	0.62	0.54	0.56	0.61	0.69	0.70	0.62	0.66	0.66
EXP7	2.44	1.24	0.45	0.41	0.37	0.45	0.44	0.38	0.47	0.45	0.55	0.54	0.50	0.52	0.50
EXP8	2.41	1.23	0.48	0.43	0.40	0.50	0.54	0.44	0.49	0.51	0.57	0.62	0.58	0.62	0.56

Note: N = 193. SWP = perceptions items, DES = desires items, EXP = expectations items.

Table 7 (continued).*Sample 1 correlation table for all 25 items included in MTurk survey.*

	14	15	16	17	18	19	20	21	22	23	24	25
DES6	1.00											
DES7	0.81	1.00										
DES8	0.78	0.83	1.00									
DES9	0.78	0.84	0.88	1.00								
EXP1	0.46	0.52	0.44	0.46	1.00							
EXP2	0.60	0.60	0.57	0.51	0.63	1.00						
EXP3	0.71	0.71	0.67	0.67	0.58	0.79	1.00					
EXP4	0.59	0.65	0.58	0.60	0.73	0.68	0.76	1.00				
EXP5	0.65	0.69	0.62	0.60	0.73	0.73	0.81	0.86	1.00			
EXP6	0.64	0.68	0.63	0.61	0.73	0.69	0.76	0.87	0.88	1.00		
EXP7	0.52	0.54	0.48	0.50	0.64	0.66	0.67	0.70	0.69	0.70	1.00	
EXP8	0.52	0.60	0.52	0.58	0.72	0.62	0.72	0.82	0.75	0.77	0.80	1.00

Note: $N = 193$. SWP = perceptions items, DES = desires items, EXP = expectations items.

Table 8.*Sample 1 EFA rotated factor matrix.*

Item	Factor 1	Factor 2	Factor 3
SWP 1	0.75*	-0.05	0.11
SWP 2	0.88*	0.05	-0.10
SWP 3	0.65*	0.11	-0.00
SWP 4	0.73*	0.02	0.08
SWP 5	0.76*	0.03	0.12
SWP 6	0.91*	-0.03	-0.05
SWP 7	0.75*	-0.04	0.11
SWP 8	0.77*	0.08	0.02
DES 1	-0.07	0.78*	0.18
DES 2	0.02	0.74*	0.19
DES 3	0.09	0.72*	0.08
DES 4	-0.01	0.83*	0.08
DES 5	0.02	0.96*	-0.04
DES 6	0.11	0.82*	-0.01
DES 7	0.04	0.84*	0.08
DES 8	-0.06	0.94*	0.00
DES 9	-0.04	0.96*	-0.03
EXP 1	0.05	-0.17	0.88*
EXP 2	0.11	0.07	0.65*
EXP 3	0.06	0.27	0.61*
EXP 4	-0.07	0.00	0.98*
EXP 5	-0.02	0.06	0.90*
EXP 6	0.00	0.06	0.88*
EXP 7	-0.12	-0.01	0.87*
EXP 8	-0.11	0.00	0.94*

Note: Geomin oblique rotation. * indicates highest loading for item.

Table 9.*Sample 1 eigenvalues from EFA for actual and random data.*

Eigenvalues for actual data		Eigenvalues for random data	
Factor	Eigenvalue	Factor	Eigenvalue
1	15.66	1	1.68
2	1.92	2	1.61
3	1.77	3	1.51
4	0.65	4	1.48
5	0.53	5	1.37
6	0.51	6	1.26
7	0.44	7	1.22
8	0.37	8	1.19
9	0.33	9	1.16
10	0.30	10	1.14
11	0.30	11	1.11
12	0.28	12	1.04
13	0.26	13	0.96
14	0.24	14	0.96
15	0.20	15	0.91
16	0.19	16	0.84
17	0.18	17	0.79
18	0.15	18	0.75
19	0.14	19	0.71
20	0.13	20	0.66
21	0.12	21	0.61
22	0.10	22	0.58
23	0.09	23	0.53
24	0.08	24	0.50
25	0.07	25	0.43

Note. $N = 193$.

Table 10.*Sample 1 correlation table and reliabilities for underload dimensions—25-item scale.*

Dimension	<i>M</i> (SD)	1	2	3	4
Perceptions	2.21 (1.09)	(.94, .99)			
Desires	2.10 (1.03)	.71	(.97, .99)		
Expectations	2.29 (1.05)	.70	.75	(.96, .99)	
Overall Underload	2.20 (.95)	.89	.91	.90	(.97, .99)

Note. $p < .001$ for all correlations. For reliabilities, alpha is reported first, followed by omega.

Table 11.*Perceptions items—difficulties and discriminations for Sample 2.*

Perceptions items:	Item difficulty	Item discrimination
1. I do not have enough work to do.	1.95	0.63
2. I find myself with nothing to do.	2.07	0.76
3. I have more time than I need to do my work.	2.46	0.67
4. After I complete all of my work, there is still time left in my work day.	2.44	0.68
5. In order to work at my full capacity, I would need more work to do.	2.17	0.78
6. I have too much time to complete my work.	2.07	0.79
7. I do not have enough work to do to fill my entire work day.	2.06	0.66
8. I could be more productive if I had more work to do.	2.19	0.79

Note: $N = 194$. All items begin with the stem “When I am at work...”

Table 12.*Desires items—difficulties and discriminations for Sample 2.*

Desires items:	Item difficulty	Item discrimination
1. I wish that I had more to do.	2.04	0.82
2. I wish that more of my time was filled up.	2.03	0.86
3. I wish that there were not as many lulls in my work day.	2.11	0.78
4. I would prefer to be busier.	2.21	0.81
5. I want more work to do.	2.18	0.81
6. The amount of work that I have is less than what I would like.	2.06	0.85
7. I would be more satisfied if I had more work to do.	2.19	0.83
8. I would like to do additional work.	2.30	0.74
9. I would enjoy having a higher workload	2.20	0.76

Note: $N = 194$. All items begin with the stem “When I am at work...”

Table 13.*Expectations items—difficulties and discriminations for Sample 2.*

Expectations items:	Item difficulty	Item discrimination
1. When I accepted this job, I thought that it would involve more work.	2.52	0.69
2. Someone in my job position should have a higher workload.	2.29	0.75
3. There should be more for me to do when I am at work.	2.34	0.82
4. When I started my job, I had the impression that I would have more work to do.	2.42	0.83
5. I thought that I would have more work to do in this job.	2.45	0.82
6. I expected to be busier in this job.	2.41	0.80
7. I expected to have less free time during my work day.	2.60	0.72
8. When I accepted this job, I thought that my work would take up more time in my work day.	2.56	0.77

Note: N = 194.

Table 14.*Sample 2 correlation table for all 25 items included in MTurk survey.*

	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12	13
SWP1	1.95	1.18	1.00												
SWP2	2.07	1.32	0.59	1.00											
SWP3	2.46	1.33	0.48	0.63	1.00										
SWP4	2.44	1.37	0.49	0.62	0.67	1.00									
SWP5	2.17	1.30	0.56	0.66	0.60	0.59	1.00								
SWP6	2.07	1.27	0.62	0.77	0.67	0.71	0.78	1.00							
SWP7	2.06	1.22	0.43	0.65	0.59	0.54	0.59	0.61	1.00						
SWP8	2.19	1.28	0.57	0.72	0.56	0.62	0.82	0.78	0.66	1.00					
DES1	2.04	1.19	0.48	0.58	0.41	0.40	0.55	0.56	0.44	0.58	1.00				
DES2	2.03	1.18	0.48	0.62	0.47	0.48	0.59	0.63	0.46	0.63	0.90	1.00			
DES3	2.11	1.22	0.47	0.54	0.43	0.49	0.55	0.57	0.41	0.56	0.77	0.79	1.00		
DES4	2.21	1.27	0.49	0.51	0.43	0.38	0.55	0.53	0.44	0.53	0.86	0.83	0.77	1.00	
DES5	2.18	1.25	0.43	0.54	0.41	0.41	0.54	0.53	0.39	0.52	0.84	0.82	0.76	0.86	1.00
DES6	2.06	1.18	0.44	0.57	0.47	0.46	0.64	0.58	0.50	0.62	0.77	0.81	0.72	0.82	0.82
DES7	2.19	1.23	0.43	0.51	0.43	0.42	0.56	0.52	0.44	0.57	0.82	0.83	0.75	0.84	0.84
DES8	2.30	1.30	0.33	0.42	0.43	0.40	0.46	0.41	0.38	0.45	0.73	0.72	0.66	0.78	0.80
DES9	2.20	1.29	0.39	0.49	0.46	0.44	0.49	0.50	0.36	0.47	0.70	0.73	0.63	0.72	0.81
EXP1	2.52	1.24	0.34	0.40	0.39	0.42	0.46	0.46	0.41	0.43	0.47	0.51	0.40	0.44	0.45
EXP2	2.29	1.14	0.40	0.48	0.41	0.50	0.53	0.51	0.49	0.54	0.51	0.60	0.46	0.51	0.49
EXP3	2.34	1.19	0.42	0.52	0.48	0.52	0.61	0.54	0.50	0.59	0.66	0.70	0.64	0.68	0.64
EXP4	2.42	1.30	0.47	0.57	0.50	0.50	0.60	0.62	0.47	0.62	0.60	0.64	0.57	0.57	0.58
EXP5	2.45	1.32	0.49	0.57	0.50	0.52	0.59	0.61	0.47	0.63	0.54	0.61	0.55	0.56	0.56
EXP6	2.41	1.27	0.48	0.54	0.46	0.48	0.58	0.60	0.45	0.61	0.55	0.59	0.50	0.54	0.54
EXP7	2.60	1.30	0.37	0.46	0.42	0.46	0.55	0.50	0.48	0.55	0.50	0.51	0.47	0.46	0.47
EXP8	2.56	1.31	0.43	0.51	0.44	0.50	0.54	0.55	0.47	0.55	0.54	0.54	0.50	0.48	0.53

Note: N = 194. SWP = perceptions items, DES = desires items, EXP = expectations items.

Table 14 (continued).*Sample 2 correlation table for all 25 items included in MTurk survey.*

	14	15	16	17	18	19	20	21	22	23	24	25
DES6	1.00											
DES7	0.87	1.00										
DES8	0.73	0.81	1.00									
DES9	0.77	0.77	0.86	1.00								
EXP1	0.51	0.45	0.37	0.44	1.00							
EXP2	0.61	0.55	0.47	0.51	0.73	1.00						
EXP3	0.75	0.73	0.62	0.59	0.57	0.76	1.00					
EXP4	0.59	0.61	0.50	0.51	0.74	0.68	0.67	1.00				
EXP5	0.60	0.59	0.48	0.54	0.78	0.69	0.70	0.89	1.00			
EXP6	0.59	0.57	0.44	0.48	0.69	0.70	0.67	0.87	0.85	1.00		
EXP7	0.50	0.49	0.43	0.41	0.65	0.68	0.61	0.70	0.67	0.71	1.00	
EXP8	0.55	0.54	0.45	0.49	0.74	0.71	0.62	0.81	0.76	0.76	0.80	1.00

Note: $N = 194$. SWP = perceptions items, DES = desires items, EXP = expectations items.

Table 15.*Sample 2 CFA factor loadings for 3-factor 25-item scale.*

Item	Factor 1	Factor 2	Factor 3
SWP 1	0.67		
SWP 2	0.83		
SWP 3	0.73		
SWP 4	0.75		
SWP 5	0.86		
SWP 6	0.90		
SWP 7	0.72		
SWP 8	0.88		
DES 1		0.91	
DES 2		0.91	
DES 3		0.83	
DES 4		0.92	
DES 5		0.92	
DES 6		0.90	
DES 7		0.92	
DES 8		0.85	
DES 9		0.83	
EXP 1			0.81
EXP 2			0.79
EXP 3			0.76
EXP 4			0.94
EXP 5			0.93
EXP 6			0.91
EXP 7			0.78
EXP 8			0.86

Note. $N = 194$. SWP = perceptions items, DES = desires items, EXP = expectations items.

Table 16.*Sample 2 discrepancy matrix for all 25 items included in MTurk survey.*

	1	2	3	4	5	6	7	8	9	10	11	12	13
SWP1	0.00												
SWP2	0.04	0.00											
SWP3	-0.01	0.02	0.00										
SWP4	-0.01	-0.01	0.12	0.00									
SWP5	-0.01	-0.06	-0.03	-0.05	0.00								
SWP6	0.02	0.01	0.02	0.03	0.00	0.00							
SWP7	-0.05	0.05	0.06	0.01	-0.02	-0.04	0.00						
SWP8	-0.02	-0.01	-0.08	-0.03	0.07	-0.01	0.03	0.00					
DES1	0.06	0.05	-0.05	-0.07	0.01	-0.01	-0.01	0.03	0.00				
DES2	0.06	0.10	0.01	0.01	0.05	0.06	0.01	0.07	0.07	0.00			
DES3	0.09	0.06	0.02	0.07	0.06	0.05	0.00	0.06	0.02	0.03	0.00		
DES4	0.06	-0.01	-0.03	-0.09	0.01	-0.04	-0.02	-0.02	0.02	-0.01	0.02	0.00	
DES5	0.00	0.01	-0.06	-0.07	-0.01	-0.04	-0.07	-0.04	0.00	-0.03	0.00	0.01	0.00
DES6	0.02	0.05	0.02	0.00	0.11	0.02	0.05	0.08	-0.04	-0.01	-0.02	0.00	0.00
DES7	0.00	-0.02	-0.04	-0.06	0.02	-0.06	-0.02	0.01	-0.02	-0.02	-0.01	-0.01	-0.01
DES8	-0.06	-0.06	0.01	-0.04	-0.04	-0.11	-0.05	-0.06	-0.04	-0.05	-0.03	0.00	0.02
DES9	0.00	0.01	0.04	0.01	0.00	-0.02	-0.05	-0.03	-0.05	-0.03	-0.06	-0.04	0.05
EXP1	-0.06	-0.10	-0.05	-0.03	-0.06	-0.08	-0.03	-0.10	-0.05	0.00	-0.07	-0.07	-0.07
EXP2	0.01	0.00	-0.01	0.06	0.03	-0.01	0.07	0.03	0.01	0.09	0.00	0.01	-0.02
EXP3	0.04	0.06	0.07	0.10	0.13	0.03	0.10	0.09	0.18	0.22	0.20	0.20	0.15
EXP4	0.01	-0.01	0.00	-0.02	0.01	-0.01	-0.03	0.02	0.00	0.04	0.04	-0.02	-0.02
EXP5	0.04	0.00	0.00	0.00	0.01	-0.01	-0.02	0.03	-0.05	0.02	0.02	-0.03	-0.04
EXP6	0.03	-0.01	-0.02	-0.02	0.01	-0.01	-0.03	0.02	-0.03	0.02	-0.02	-0.04	-0.05
EXP7	-0.02	-0.02	-0.01	0.02	0.05	-0.02	0.06	0.04	0.01	0.01	0.02	-0.04	-0.03
EXP8	0.00	-0.02	-0.03	0.03	-0.01	-0.03	0.02	-0.01	-0.01	0.00	0.01	-0.07	-0.02

Note: $N = 194$. Large discrepancy values are bold. SWP = perceptions items, DES = desires items, EXP = expectations items.

Table 16 (continued).*Sample 2 discrepancy matrix for all 25 items included in MTurk survey.*

	14	15	16	17	18	19	20	21	22	23	24	25
DES6	0.00											
DES7	0.05	0.00										
DES8	-0.02	0.03	0.00									
DES9	0.02	0.01	0.16	0.00								
EXP1	0.00	-0.07	-0.11	-0.03	0.00							
EXP2	0.12	0.04	0.00	0.05	0.09	0.00						
EXP3	0.28	0.24	0.17	0.15	-0.05	0.15	0.00					
EXP4	0.01	0.01	-0.05	-0.03	-0.01	-0.06	-0.04	0.00				
EXP5	0.02	-0.01	-0.06	0.00	0.03	-0.04	-0.01	0.03	0.00			
EXP6	0.02	-0.01	-0.09	-0.04	-0.04	-0.02	-0.02	0.03	0.02	0.00		
EXP7	0.01	-0.01	-0.03	-0.05	0.02	0.06	0.01	-0.04	-0.06	0.00	0.00	
EXP8	0.01	-0.01	-0.05	-0.01	0.05	0.03	-0.03	0.00	-0.04	-0.02	0.13	0.00

Note: $N = 194$. Large discrepancy values are bold. SWP = perceptions items, DES = desires Items, EXP = expectations items.

Table 17.*Sample 2 correlation table and reliabilities for underload dimensions—25-item scale.*

Dimension	<i>M</i> (SD)	1	2	3	4
Perceptions	2.18 (1.06)	(.93, .99)			
Desires	2.15 (1.11)	.66	(.97, .99)		
Expectations	2.45 (1.10)	.70	.68	(.96, .99)	
Overall Underload	2.26 (.96)	.88	.88	.90	(.97, .99)

Note. $N = 194$. $p < .001$ for all correlations. For reliabilities, alpha is reported first, followed by omega.

Table 18.*Sample 2 CFA factor loadings for 3-factor 18-item scale.*

Item	Factor 1	Factor 2	Factor 3
SWP 2	0.83		
SWP 4	0.73		
SWP 5	0.86		
SWP 6	0.90		
SWP 7	0.72		
SWP 8	0.90		
DES 1		0.93	
DES 2		0.93	
DES 3		0.84	
DES 4		0.92	
DES 5		0.92	
DES 7		0.90	
DES 9		0.81	
EXP 1			0.80
EXP 4			0.96
EXP 5			0.94
EXP 6			0.91
EXP 8			0.84

Note. $N = 194$. SWP = perceptions items, DES = desires items, EXP = expectations items.

Table 19.*Sample 2: Comparing CFA fit statistics for alternative models.*

Model	χ^2	<i>df</i>	CFI	TLI	RMSEA	SRMR
25 Items Single Factor Model	1435.76	275	.67	.64	.15	.10
25 Items 2-Factor Model	1135.80	274	.76	.73	.13	.10
25 Items 3-Factor Model	599.26	272	.91	.90	.08	.05
18 Items Single Factor Model	882.99	135	.68	.63	.17	.11
18 Items 2-Factor Model	643.23	134	.78	.75	.14	.10
18 Items 3-Factor Model	202.13	132	.97	.97	.05	.03

Note. CFI = comparative fit index; TLI = Tucker–Lewis Index. RMSEA = Root Mean Square Error of Approximation. SRMR = Standardized Root Mean Square Residual

Table 20.*Sample 2 correlation table and reliabilities for underload dimensions—18-item scale.*

Dimension	<i>M</i> (SD)	1	2	3	4
Perceptions	2.17 (1.10)	(.92, .99)			
Desires	2.14 (1.12)	.67	(.97, .99)		
Expectations	2.47 (1.18)	.67	.64	(.95, .99)	
Overall Underload	2.26 (.99)	.88	.87	.88	(.96, .99)

Note. $p < .001$ for all correlations. For reliabilities, alpha is reported first, followed by omega.

Table 21.*Means, standard deviations, skewness, and kurtosis for all variables.*

	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
WAMI - Positive Meaning	3.45	1.03	-0.58	-0.41
WAMI - Making Meaning	3.31	0.97	-0.43	-0.28
Overall stress (sum)	23.02	14.62	-0.09	-1.28
Job Satisfaction	3.71	0.99	-1.00	0.40
Underemployment	4.47	1.55	-0.16	-0.93
Role Ambiguity	1.98	0.93	1.54	3.71
Job Boredom	2.79	0.90	0.15	-0.53
Role Overload	3.37	1.48	0.44	-0.63
Caplan's Underload	2.58	0.75	0.25	0.09
Social Desirability (sum)	49.41	6.39	0.02	-0.39
Underload-Perceptions	2.16	1.11	0.80	-0.34
Underload- Desires	2.12	1.08	0.68	-0.60
Underload-Expectations	2.38	1.14	0.45	-0.97
Underload- Total	2.22	0.98	0.59	-0.49

Note. $N = 387$. All scale scores except stress and social desirability reflect means of items. Overall stress and social desirability scale scores are sums of items.

Table 22.*Correlation matrix for all variables.*

	1	2	3	4	5	6	7
WAMI-PM	(0.94)						
WAMI-MM	0.83**	(0.82)					
SIG	-0.26**	-0.23**	(0.93)				
JS	0.71**	0.63**	-0.40**	(0.91)			
SPQ	-0.46**	-0.40**	-0.02	-0.31**	(0.93)		
RA	-0.18**	-0.09	0.23**	-0.30**	-0.09	(0.92)	
JBS	-0.64**	-0.56**	0.25**	-0.63**	0.44**	0.19**	(0.93)
RO	-0.13*	-0.10*	0.60**	-0.31**	-0.13*	0.30**	0.15**
CUL	-0.07	-0.02	-0.55**	0.02	0.17**	0.00	0.17**
SD	0.13*	0.14**	-0.20**	0.15**	-0.03	-0.10*	-0.19**
SWP	-0.17**	-0.07	-0.22**	-0.12*	0.24**	0.08	0.36**
DES	0.00	0.10	-0.25**	0.01	0.17**	0.11*	0.23**
EXP	-0.02	0.04	-0.27**	-0.03	0.18**	0.11*	0.20**
UNDER	-0.07	0.02	-0.28**	-0.05	0.22**	0.11*	0.30**

Note. N = 387. ** Correlation is significant at the 0.01 level. * Correlation is significant at the 0.05 level. WAMI- PM = Working and Meaning Inventory (positive meaning subscale), WAMI- MM = Work and Meaning Inventory (meaning making subscale), SIG = Stress in General scale, JS = Job Satisfaction, SPQ = Scale of Perceived Overqualification, RA = Role Ambiguity, JBS = Job Boredom Scale, RO = Role Overload, CUL = Caplan et al. (1975) Underload Scale, SD = Social Desirability, SWP = Underload Scale (perceptions dimension), DES = Underload Scale (desires dimension), EXP = Underload Scale (expectations dimension), UNDER = Underload Scale (total score).

Table 22 (continued).*Correlation matrix for all variables.*

	8	9	10	11	12	13	14
WAMI- PM							
WAMI- MM							
SIG							
JS							
SPQ							
RA							
JBS							
RO	(0.77)						
CUL	-0.60**	(0.76)					
SD	-0.12*	0.03	(0.85)				
SWP	-0.38**	0.67**	-0.03	(0.95)			
DES	-0.36**	0.56**	0.13**	0.68**	(0.96)		
EXP	-0.36**	0.61**	0.03	0.67**	0.68**	(0.95)	
UNDER	-0.41**	0.69**	0.05	0.88**	0.88**	0.89**	(0.96)

Note. N = 387. ** Correlation is significant at the 0.01 level. * Correlation is significant at the 0.05 level. WAMI- PM = Working and Meaning Inventory (positive meaning subscale), WAMI- MM = Work and Meaning Inventory (meaning making subscale), SIG = Stress in General scale, JS = Job Satisfaction, SPQ = Scale of Perceived Overqualification, RA = Role Ambiguity, JBS = Job Boredom Scale, RO = Role Overload, CUL = Caplan et al. (1975) Underload Scale, SD = Social Desirability, SWP = Underload Scale (perceptions dimension), DES = Underload Scale (desires dimension), EXP = Underload Scale (expectations dimension), UNDER = Underload Scale (total score).

Study 1: Content Validity
 $N = 11$ psychology graduate student SMEs

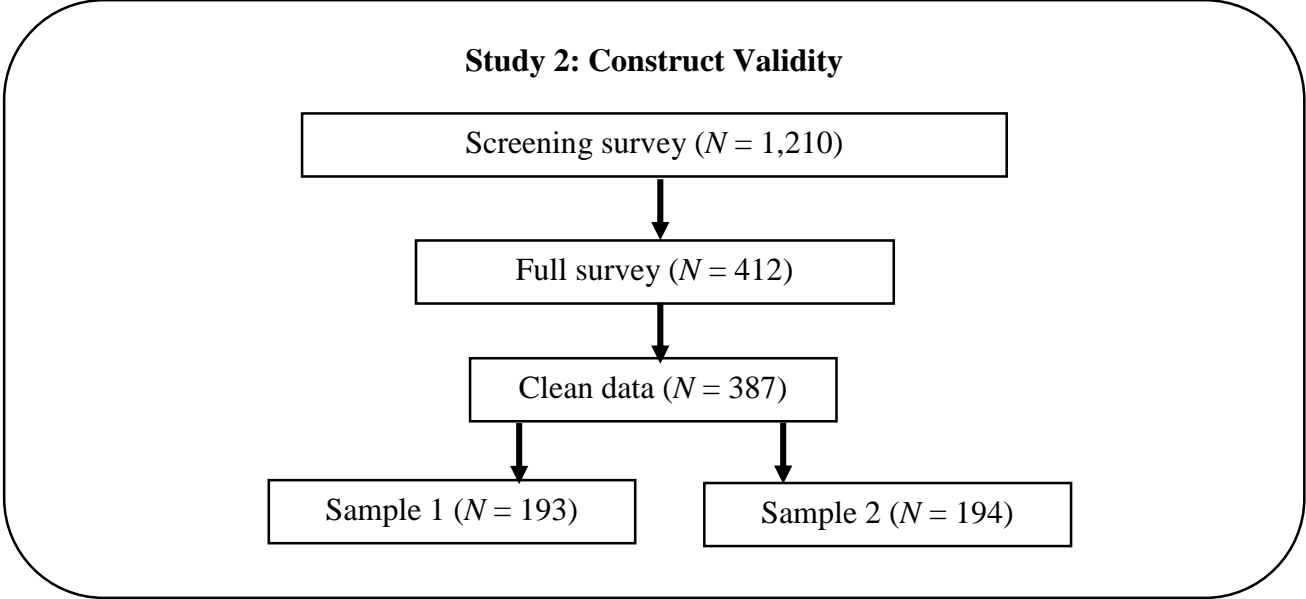


Figure 2. Sample sizes for each phase of the study.

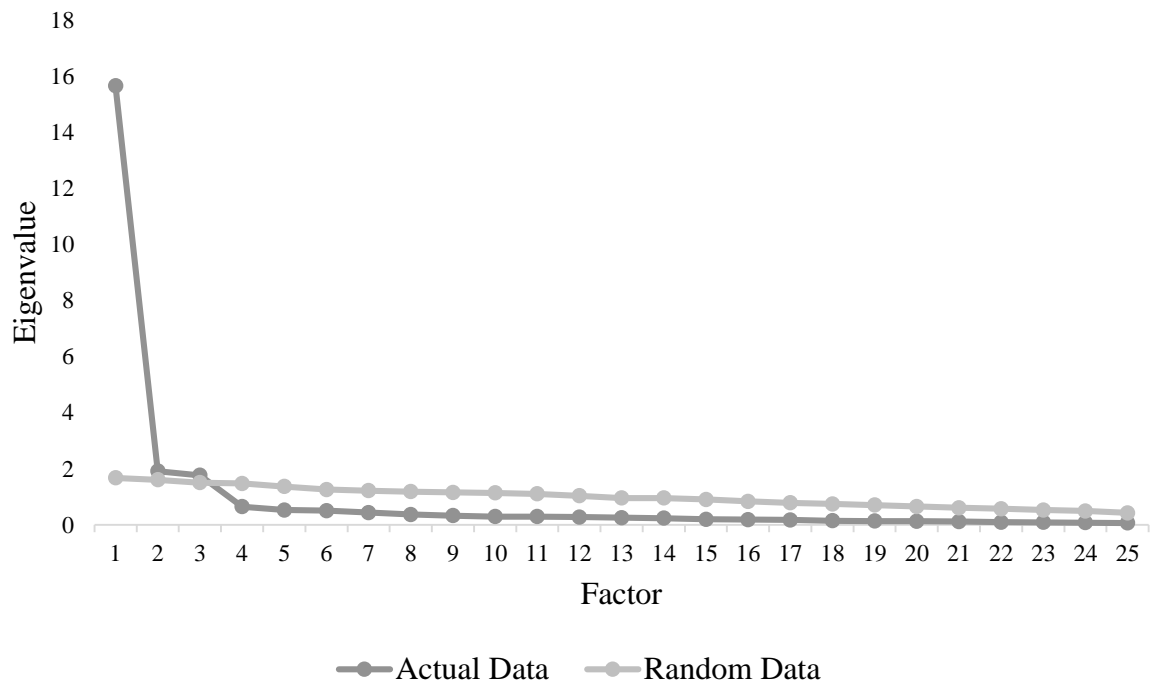


Figure 2. Scree plots for actual data and randomly generated data.

References

- Anastasi, A. & Urbina, S. (1997). *Psychological testing*. Upper Saddle River, NJ: Prentice Hall.
- Autor, D. H. (2015). Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives*, 29(3), 3-30.
- Bandalos, D. L. & Boehm-Kaufman, M. R. (2008). Four common misconceptions in exploratory factor analysis. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrines, verity, and fable in the organizational and social sciences* (pp. 61-88). New York: Taylor and Francis Group.
- Berg, J. M., Grant, A. M., & Johnson, V. (2010). When callings are calling: Crafting work and leisure in pursuit of unanswered occupational callings. *Organization Science*, 21(5), 973-994.
- Biddle, E. J. & Thomas, B. J. (1966). *Role theory: Concepts and research*. New York: Wiley.
- Breaugh, J. A. & Colihan, J. P. (1994). Measuring facets of job ambiguity: Construct validity evidence. *Journal of Applied Psychology*, 79, 191-202.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data? *Perspectives on Psychological Science*, 6(1), 3-5.
- Cammann, C., Fichman, M., Jenkins, G. D., & Klesh, J. (1983). Michigan Organizational Assessment Questionnaire. In S. E. Seashore, E. E. Lawler, P. H. Mirvis, & C. Cammann

- (Eds.), *Assessing organizational change: A guide to methods, measures, and practices* (pp. 71-138). New York: Wiley-Interscience.
- Caplan, R. D., Cobb, S., French, J. R. P., Harrison, R. V., & Pinneau, S. R. (1975). *Job Demands and Worker Health*. Washington D.C.: H.E.W. Publication No. NIOSH, 75-160.
- Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, 29, 2156-2160.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 10, 165-179.
- Cavanaugh, M. A., Boswell, W. R., Roehling, M. V., & Boudreau, J. W. (2000). An empirical examination of self-reported work stress among U.S. managers. *Journal of Applied Psychology*, 85(1), 65-74.
- Chilton, L. B., Horton, J. J., Miller, R. C., & Azenkot, S. (2010). *Task search in a human computation market*. In *Proceedings of the ACM SIGKDD Workshop on Human Computation* (pp. 1-9). New York: ACM.
- Cohen, J. (1988) *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cook, T. D. & Campbell, D. T. (1976). The design and conduct of quasi-experiments and true experiments in field settings. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 223-326). Chicago: Rand McNally.

- Cordery, J. L. & Sevastos, P. P. (1993). Responses to the original and revised Job Diagnostic Survey: Is education a factor in responses to negatively worded items? *Journal of Applied Psychology, 78*(1), 141-143.
- Crowne, D. P. & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Counseling Psychology, 14*(4), 349-354.
- Dawes, J. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point, and 10-point scales. *International Journal of Market Research, 50*(1), 61-77.
- DeVellis, R. F. (2012). *Scale development*. United States: SAGE Publications, Inc.
- Edwards, A. L. (1953). The relationship between the judged desirability of a trait and the probability that it will be endorsed. *Journal of Applied Psychology, 37*, 90-93.
- Edwards, J. R. (1991). Person-job fit: A conceptual integration, literature review, and methodological critique. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (Vol. 6, pp. 283-357). New York: Wiley.
- Edwards, J. R. (2002). Alternatives to difference scores: Polynomial regression analysis and response surface methodology. In F. Drasgow & N. W. Schmitt (Eds.), *Advances in measurement and data analysis* (pp. 350-400). San Francisco: Jossey-Bass.
- Eisenhower, D., Mathiowetz, N. A., & Morganstein, D. (2011). Recall error: Sources and Bias reduction techniques. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman, (Eds.). *Measurement errors in surveys* (pp. 127-144). Hoboken, NJ: John Wiley & Sons, Inc.

- Erdogan, B., Bauer, T., Peiro, J. M., Truxillo, D. M. (2011). Overqualification theory, research, and practice: Things that matter. *Industrial and Organizational Psychology*, 4, 260-267.
- Fisher, C. D. (1987). Boredom: Construct, causes, and consequences. Technical Report ONR-9, Texas A&M University.
- Fisher, C. D. (1991). Boredom at work: A neglected concept. *Bond University School of Business Discussion Papers*, Paper 19.
- Floyd, F. J. & Widaman, K. F. (1995). Factor analysis in the development of refinement of clinical assessment instruments. *Psychological Assessment*, 7(3), 286-299.
- Folkman, S., Lazarus, R. S., Dunkel-Schetter, C., DeLongis, A., & Gruen, R. J. (1986). Dynamics of a stressful encounter: Cognitive appraisal, coping, and encounter outcomes. *Journal of Personality and Social Psychology*, 50(5), 992-1003.
- Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology*, 39, 291-314.
- Frankenhaeuser, M. & Gardell, B. (1976). Underload and overload in working life: Outline of a multidisciplinary approach. *Journal of Human Stress*, 2(3), 35-46.
- Froggat, K. I. & Cotton, J. L. (1984). Effects of sex and type A behavior pattern on overload- and underload-induced stress: A laboratory investigation. *Academy of Management Proceedings*, 1984(1), 207-211.
- Gallup (2014, March 6). *U.S. Payroll to Population Rate 43.1% in February*. Retrieved from <http://www.gallup.com/poll/167750/payroll-population-rate-february.aspx>.

- Ganster, D. C., Fusilier, M. R., & Mayes, B. T. (1986). Role of social support in the experience of stress at work. *Journal of Applied Psychology, 71*(1), 102-110.
- Ganster, D. C. & Rosen, C. C. (2013). Work stress and employee health: A multidisciplinary review. *Journal of Management, 39*(5), 1085-1122.
- Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making, 26*, 213-224.
- Greenberger, E., Chen, C., Dmitrieva, J., & Farruggia, S. P. (2003). Item-wording and the dimensionality of the Rosenberg Self-Esteem Scale: do they matter? *Personality and Individual Differences, 35*(6), 1241-1254.
- Guion, R. M. (1998). *Assessment, measurement, and prediction for personnel decisions*. Mahwah, NJ: Erlbaum.
- Hackman, J. R. & Oldham, G. R. (1975). Development of the Job Diagnostic Survey. *Journal of Applied Psychology, 60*, 159-170.
- Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *Journal of Management, 21*(5), 967-988.
- James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology, 69*(1), 85-98.
- Kahn, R. L., Wolfe, D. M., Quinn, R. P., Snoek, J. D., & Rosenthal, R. A. (1964). *Organizational stress: Studies in role conflict and ambiguity*. New York: Wiley.

- Kahn, W. A. (1990). Psychological conditions of personal engagement and disengagement at work. *Academy of Management Journal*, 33(4), 692-724.
- Keenan, A. & Newton, T. J. (1984). Frustration in organizations: Relationships to role stress, climate, and psychological strain. *Journal of Occupational Psychology*, 57, 57-65.
- Kristof-Brown, A. L. & Guay, R. P. (2011). Person-environment fit. In S. Zedeck (ed.), *APA Handbook of Industrial and Organizational Psychology* (Vol. 3, pp. 3-50). Washington, D.C.: American Psychological Association.
- Kristof-Brown, A. L., Zimmerman, R. D., & Johnson, E. C. (2005). Consequences of individuals' fit at work: A meta-analysis of person-job, person-organization, person-group, and person-supervisor fit. *Personnel Psychology*, 58(2), 281-342.
- Lazarus, R. S. (1966). *Psychological stress and the coping process*. New York: McGraw-Hill.
- Lazarus, R. S. & Folkman, S. (1984). *Stress, appraisal, and coping*. New York: Springer.
- LeBreton, J. M. & Senter, J. L. (2008). Answers to 20 Questions about Interrater Reliability and Interrater Agreement. *Organizational Research Methods*, 11(4), 815-852.
- Lee, T. W. (1986). Toward the development and validation of a measure of job boredom. *Manhattan College Journal of Business*, 15, 22-28.
- Le Fevre, M., Matheny, J., & Kolt, G. S. (2003). Eustress, distress, and interpretation in occupational stress. *Journal of Managerial Psychology*, 18(7), 726-744.
- Levinson, H., Price, C., Munden, K., Mandl, H., & Solley, C. (1962). *Men, management, and mental health*. Cambridge, MA: Harvard University Press.

- Lindell, M. K., Brandt, C. J., & Whitney, D. J. (1999). A revised index of interrater agreement for multi-item ratings of a single target. *Applied Psychological Measurement, 23*(2), 127-135.
- Luksyte, A., Spitzmueller, C., & Maynard, D. C. (2011). Why do overqualified incumbents deviate? Examining multiple mediators. *Journal of Occupational Health Psychology, 16*(3), 279-296.
- Mason, W. & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods, 44*(1), 1-23.
- Maynard, D. C., Joseph, T. A., & Maynard, A. M. (2006). Underemployment, job attitudes, and turnover intentions. *Journal of Organizational Behaviour, 27*, 509-536.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Psychology Press.
- McKee-Ryan, F. M. & Harvey, J. (2011). "I have a job, but...": A review of underemployment. *Journal of Management, 37*(4), 962-996.
- Melamed, S., Ben-Avi, I., Luz, J., & Green, M. S. (1995a). Objective and subjective work monotony: Effects on job satisfaction, psychological distress, and absenteeism in blue-collar workers. *Journal of Applied Psychology, 80*(1), 29-42.
- Melamed, S., Ben-Avi, I., Luz, J., & Green, M. S. (1995b). Repetitive work, work underload, and coronary heart disease risk factors among blue-collar workers—the CORDIS study. *Journal of Psychosomatic Research, 39*(1), 19-29.

- Morrison, E. W. & Robinson, S. L. (1997). When employees feel betrayed: A model of how psychological contract violation develops. *Academy of Management Review*, 22(1), 236-256.
- Muchinsky, P. M. & Monahan, C. J. (1987). What is person-environment congruence? Supplementary versus complementary models of fit. *Journal of Vocational Behavior*, 31, 268-277.
- Murphy, K. R., & Davidshofer, C. O. (1991). *Psychological testing*. Prentice-Hall.
- Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, 3(1), 1-18.
- Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.
- Paolacci, G., Chandler, J. & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5(5), 411-419.
- Parasuraman, R. & Hancock, P. A. (2001). Adaptive control of mental workload. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, workload, and fatigue* (pp. 305-320). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879-903.
- Polit, D. F., & Beck, C. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health*, 29(5), 489-497.
doi:10.1002/nur.20147

- Raykov, T. & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. New York: Taylor and Francis Group.
- Robinson, W. P. (1975). Boredom at school. *British Journal of Educational Psychology*, 45, 141-152.
- Rodriguez, I., Kozusznik, M. W., & Peiro, J. M. (2013). Development and validation of the Valencia Eustress-Distress Appraisal Scale. *International Journal of Stress Management*, 20(4), 279-308.
- Rousseau, D. M. (1989). Psychological and implied contracts in organizations. *Employee Responsibilities and Rights Journal*, 2, 322-337.
- Sales, S. M. (1970). Some effects of role overload and role underload. *Organizational Behavior and Human Performance*, 5, 592-608.
- Schwab, D. P. (1980). Construct validity in organization behavior. In B. M. Staw & L. L. Cummings (Eds.), *Research in organizational behavior*, Vol. 2 (pp. 3-43). Greenwich, CT: JAI Press.
- Selye, H. (1987). *Stress without distress*. London: Corgi.
- Shaw, J. B. & Weekley, J. A. (1985). The effects of objective workload variations of psychological strain and post-workload performance. *Journal of Management*, 11(1), 87-98.
- Shultz, K. S., Wang, M., & Olson, D. A. (2010). Role overload and underload in relation to occupational stress and health. *Stress and Health*, 26, 99-111.

- Spector, P. E. & Jex, S. M. (1998). Development of four self-report measures of job stressors and strain: Interpersonal Conflict at Work Scale, Organizational Constraints Scale, Quantitative Workload Inventory, and Physical Symptoms Inventory. *Journal of Occupational Health Psychology, 3*(4), 356-367.
- Stanton, J. M., Balzer, W. K., Smith, P. C., Parra, L. F., & Ironson, G. (2001). A general measure of work stress: The Stress in General Scale. *Educational and Psychological Measurement, 61*(5), 866-888.
- Stanton, J. M., Sinar, E. F., Balzer, W. K., & Smith, P. C. (2002). Issues and strategies for reducing the length of self-report scales. *Personnel Psychology, 55*, 167-194.
- Steger, M. F., Dik, B. J., & Duffy, R. D. (2012). Measuring meaningful work: The Work and Meaning Inventory (WAMI). *Journal of Career Assessment, 20*(3), 322-337.
- Wanous, J. P. (1973). Effects of a realistic job preview on job acceptance, job attitudes, and job survival. *Journal of Applied Psychology, 58*(3), 327-332.
- Widmer, P. S., Semmer, N. K., Kalin, W., Jacobshagen, N., & Meier, L. L. (2012). The ambivalence of challenge stressors: Time pressure associated with both negative and positive well-being. *Journal of Vocational Behavior, 80*, 422-433.
- Wrzesniewski, A., & Dutton, J. E. (2001). Crafting a job: Revisioning employees as active crafters of their work. *Academy of Management Review, 26*(2), 179-201.
- Zwick, W. R. & Velicer, W. F. (1986). A comparison of five rules for determining the number of components in data sets. *Psychological Bulletin, 99*, 432-442.

Appendix A: Previous Measures of Underload

Authors	Items
Sales (1970)	1. How much work do you feel you had to do during the work period?
Caplan et al. (1975)	<ol style="list-style-type: none"> 1. How often does your job require you to work very fast? 2. How often does your job require you to work very hard? 3. How often does your job leave you with little time to get things done? 4. How often is there a great deal to be done? 5. How much slowdown in the work load do you experience? 6. How much time do you have to think and contemplate? 7. How much work load do you have? 8. What quantity of work do others expect you to do? 9. How much time do you have to do all your work? 10. How many projects, assignments, or tasks do you have? 11. How many lulls between heavy work load periods do you have?
Froggat & Cotton (1984)	<ol style="list-style-type: none"> 1. How much work did you have to do during the work period? 2. To what extent do you think the rate at which the anagrams were administered should be decreased? 3. To what extent do you think the rate at which the anagrams were administered should be increased?
Shaw & Weekley (1985)	1. I could have easily solved many more anagrams in the allotted time.
Shultz et al. (2010)	1. How well do you think your skills match the demands imposed on you by your job?

Appendix B: 25-Item Underload Scale for Study 2

Perceptions dimension:

1. I do not have enough work to do.
2. I find myself with nothing to do.
3. I have more time than I need to do my work.
4. After I complete all of my work, there is still time left in my work day.
5. In order to work at my full capacity, I would need more work to do.
6. I have too much time to complete my work.
7. I do not have enough work to do to fill my entire work day.
8. I could be more productive if I had more work to do.

Desires dimension:

1. I wish that I had more to do.
2. I wish that more of my time was filled up.
3. I wish that there were not as many lulls in my work day.
4. I would prefer to be busier.
5. I want more work to do.
6. The amount of work that I have is less than what I would like.
7. I would be more satisfied if I had more work to do.
8. I would like to do additional work.
9. I would enjoy having a higher workload

Expectations dimension:

1. When I accepted this job, I thought that it would involve more work.
2. Someone in my job position should have a higher workload.
3. There should be more for me to do when I am at work.
4. When I started my job, I had the impression that I would have more work to do.
5. I thought that I would have more work to do in this job.

6. I expected to be busier in this job.
7. I expected to have less free time during my work day.
8. When I accepted this job, I thought that my work would take up more time in my work day.

Appendix C: Job Boredom Scale (Lee, 1986)

1. I often get bored with my work.
2. My work is monotonous.
3. I would like to change from one type of work to another from time to time if the pay were the same.
4. I like the work I do.
5. I often get tired on the job.
6. I find my job dull.
7. My work day goes by too slowly.
8. I become irritable on the job.
9. I get mentally sluggish during the day.
10. I get drowsy on the job.
11. Time seems to go by slowly on the job.
12. There are long periods of boredom on the job.
13. My job seems repetitive.
14. During the work day I think about doing other tasks.
15. Monotony describes my job.
16. My work is pretty much the same day after day.
17. I get apathetic on the job.

Note. Items were rated on a 1 to 5 scale where 1 = Strongly Disagree, 2 = Disagree, 3 = Neither Agree nor Disagree, 4 = Agree, 5 = Strongly Agree.

Appendix D: Scale of Perceived Overqualification (Maynard et al, 2006)

1. My job requires less education than I have.
2. The work experience that I have is not necessary to be successful on this job.
3. I have job skills that are not required for this job.
4. Someone with less education than myself could perform well on my job.
5. My previous training is not being fully utilized on this job.
6. I have a lot of knowledge that I do not need in order to do my job.
7. My education level is above the education level required by my job.
8. Someone with less work experience than myself could perform well on my job.
9. I have more abilities than I need in order to do my job.

Note. Items were rated on a 1 to 7 scale where 1 = Strongly Disagree, 2 = Disagree, 3 = Somewhat Disagree, 4 = Neither Agree nor Disagree, 5 = Somewhat Agree, 6 = Agree, 7 = Strongly Agree.

Appendix E: Role Ambiguity Scale (Breugh & Colihan, 1994)

1. I am certain how to go about getting my job done (the methods to use).
2. I know what is the best way (approach) to go about getting my work done.
3. I know how to get my work done (what procedures to use).
4. I know what my supervisor considers satisfactory work performance.
5. It is clear to me what is considered acceptable performance by my supervisor.
6. I know what level of performance is considered acceptable by my supervisor.

Note. Items were rated on a 1 to 7 scale where 1 = Disagree Strongly, 2 = Disagree, 3 = Disagree Slightly, 4 = Neutral, 5 = Agree Slightly, 6 = Agree, 7 = Agree Strongly. Items 1-3 belong in the work method ambiguity dimension; items 4-6 belong in the performance criteria ambiguity dimension.

Appendix F: Job Satisfaction Scale (Cammann et al., 1979)

1. All in all, I am satisfied with my job.
2. In general, I don't like my job.
3. In general, I like working here.

Note. Items were rated on a 1 to 5 scale where 1 = Strongly Disagree, 2 = Disagree, 3 = Neither Agree nor Disagree, 4 = Agree, 5 = Strongly Agree.

Appendix G: Stress in General Scale (Stanton et al., 2001)

What is your job like most of the time?

1. Demanding
2. Pressured
3. Hectic
4. Calm
5. Relaxed
6. Many things stressful
7. Pushed
8. Irritating
9. Under Control
10. Nerve-wracking
11. Hassled
12. Comfortable
13. More stressful than I'd like
14. Smooth running
15. Overwhelming

Note. Items were rated on 3-point scale consisting of “No,” “?”, and “Yes”. Items 1-7 belong to the pressured dimension; items 8-15 belong to the threatened dimension.

Appendix H: Work and Meaning Inventory (Steger et al., 2012)

1. I have found a meaningful career.
2. I view my work as contributing to my personal growth.
3. My work really makes no difference to the world.
4. I understand how my work contributes to my life's meaning.
5. I have a good sense of what makes my job meaningful.
6. I know my work makes a positive difference in the world.
7. My work helps me better understand myself.
8. I have discovered work that has a satisfying purpose.
9. My work helps me make sense of the world around me.
10. The work I do serves a greater purpose.

Note. Items were rated on a 1 to 5 scale where 1 = Absolutely Untrue, 2 = Mostly Untrue, 3 = Neither True nor Untrue, 4 = Mostly True, 5 = Absolutely True. Items 1, 4, 5, and 8 belong to the positive meaning dimension; items 2, 7, and 9 belong to the meaning-making from work dimension; items 3, 6, and 10 belong to the greater good motivations dimension.

Appendix I: Role Overload Scale (Cammann et al., 1983)

1. I have too much work to do to do everything well.
2. The amount of work I am asked to do is fair.
3. I never seem to have enough time to get everything done.

Note. Items were rated on a 1 to 7 scale where 1 = Strongly Disagree, 2 = Disagree, 3 = Slightly Disagree, 4 = Neither Agree nor Disagree, 5 = Slightly Agree, 6 = Agree, 7 = Strongly Agree.

Appendix J: Crowne-Marlowe Social Desirability Scale (Crowne & Marlowe, 1960)

1. Before voting I thoroughly investigate the qualifications of all the candidates.
2. I never hesitate to go out of my way to help someone in trouble.
3. It is sometimes hard for me to go on with my work if I am not encouraged.
4. I have never intensely disliked anyone.
5. On occasion I have had doubts about my ability to succeed in life.
6. I sometimes feel resentful when I don't get my way.
7. I am always careful about my manner of dress.
8. My table manners at home are as good as when I eat out in a restaurant.
9. If I could get into a movie without paying and be sure I was not seen I would probably do it.
10. On a few occasions, I have given up doing something because I thought too little of my ability.
11. I like to gossip at times.
12. There have been times when I felt like rebelling against people in authority even though I knew they were right.
13. No matter who I'm talking to, I'm always a good listener.
14. I can remember "playing sick" to get out of something.
15. There have been occasions when I took advantage of someone.
16. I'm always willing to admit it when I make a mistake.
17. I always try to practice when I preach.
18. I don't find it particularly difficult to get along with loud mouthed, obnoxious people.
19. I sometimes try to get even rather than forgive and forget.
20. When I don't know something I don't at all mind admitting it.
21. I am always courteous, even to people who are disagreeable.
22. At times I have really insisted on having things my own way.
23. There have been occasions when I felt like smashing things.
24. I would never think of letting someone else be punished for my wrongdoings.
25. I never resent being asked to return a favor.
26. I have never been irked when people expressed ideas very different from my own.

27. I never make a long trip without checking the safety of my car.
28. There have been times when I was quite jealous of the good fortune of others.
29. I have almost never felt the urge to tell someone off.
30. I am sometimes irritated by people who ask favors of me.
31. I have never felt that I was punished without cause.
32. I sometimes think when people have a misfortune they only got what they deserved.
33. I have never deliberately said something that hurt someone's feelings.

Note. Items were rated on a True or False response scale.

Appendix K: Underload Scale (Caplan et al., 1975)

1. How often does your job require you to work very fast?
2. How often does your job require you to work very hard?
3. How often does your job leave you with little time to get things done?
4. How often is there a great deal to be done?
5. How much slowdown in the work load do you experience?
6. How much time do you have to think and contemplate?
7. How much work load do you have?
8. What quantity of work do others expect you to do?
9. How much time do you have to do all your work?
10. How many projects, assignments, or tasks do you have?
11. How many lulls between heavy work load periods do you have?

Note. Items 1-4 were rated on a 1 to 5 scale where 1 = Rarely, 2 = Occasionally, 3 = Sometimes, 4 = Fairly Often, 5 = Very Often. Items 5-11 were rated on a 1-5 scale where 1 = Hardly Any, 2 = A Little, 3 = Some, 4 = A Lot, 5 = A Great Deal.

Appendix L: Final 18-Item Underload Scale

Perceptions items:

1. I find myself with nothing to do.
 2. After I complete all of my work, there is still time left in my work day.
 3. In order to work at my full capacity, I would need more work to do.
 4. I have too much time to complete my work.
 5. I do not have enough work to do to fill my entire work day.
 6. I could be more productive if I had more work to do.
-

Desires items:

1. I wish that I had more to do.
 2. I wish that more of my time was filled up.
 3. I wish that there were not as many lulls in my work day.
 4. I would prefer to be busier.
 5. I want more work to do.
 6. I would be more satisfied if I had more work to do.
 7. I would enjoy having a higher workload
-

Expectations items:

1. When I accepted this job, I thought that it would involve more work.
2. When I started my job, I had the impression that I would have more work to do.
3. I thought that I would have more work to do in this job.
4. I expected to be busier in this job.
5. When I accepted this job, I thought that my work would take up more time in my work day.